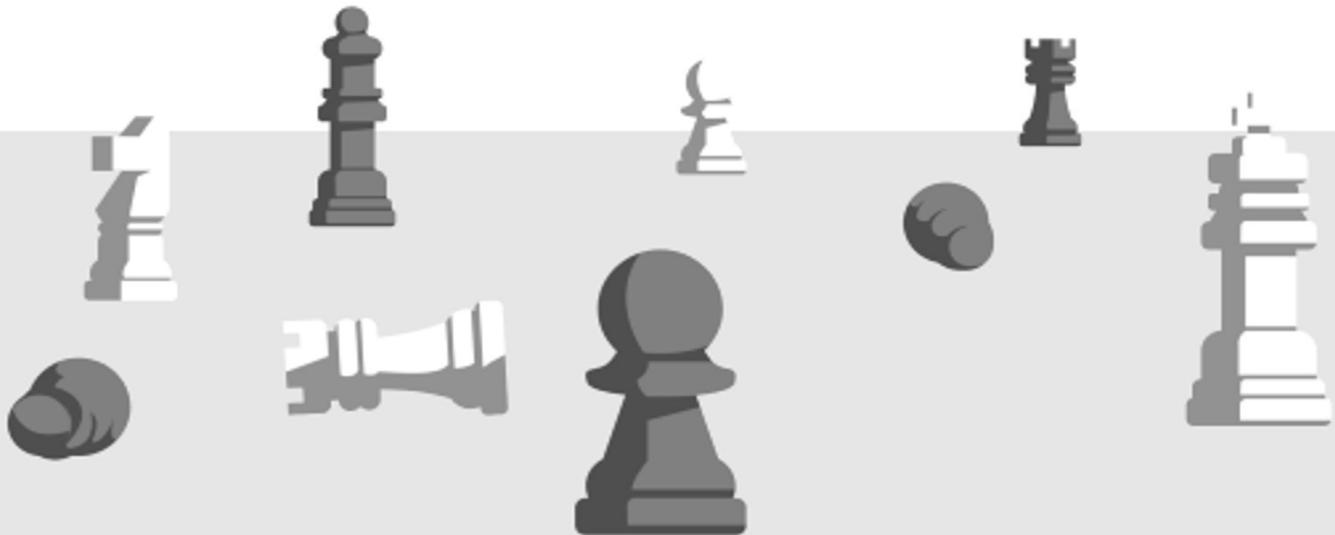


SECOND EDITION

# YOU DON'T KNOW JS YET



Frontend **Masters**

Kyle Simpson

---

# Table of Contents

Introduction	1.1
Get Started	1.2
Chapter 1: What Is JavaScript?	1.2.1
Chapter 2: Surveying JS	1.2.2
Chapter 3: Digging to the Roots of JS	1.2.3
Chapter 4: The Bigger Picture	1.2.4
Appendix A: Exploring Further	1.2.5
Appendix B: Practice, Practice, Practice!	1.2.6
Scope & Closures	1.3
Chapter 1: How Is Scope Determined?	1.3.1
Chapter 2: Understanding Lexical Scope	1.3.2
Chapter 3: Working With Scope	1.3.3
Chapter 4: Function vs. Block Scope	1.3.4
Chapter 5: Closures	1.3.5
Chapter 6: Module Pattern	1.3.6
Appendix A: Exploring Further	1.3.7
Appendix B: Practice	1.3.8
Objects & Classes	1.4
Chapter 1: this Or That?	1.4.1
Chapter 2: this All Makes Sense Now!	1.4.2
Chapter 3: Objects	1.4.3
Chapter 4: Mixing (objects-classes/Up) "Class" Objects	1.4.4
Chapter 5: Prototypes	1.4.5
Chapter 6: Behavior Delegation	1.4.6
Appendix A: ES6 class	1.4.7
Types & Grammar	1.5
Chapter 1: Types	1.5.1
Chapter 2: Values	1.5.2
Chapter 3: Natives	1.5.3
Chapter 4: Coercion	1.5.4
Chapter 5: Grammar	1.5.5
Appendix A: Mixed Environment JavaScript	1.5.6
Sync & Async	1.6
Chapter 1: Asynchrony: Now & Later	1.6.1
Chapter 2: Callbacks	1.6.2
Chapter 3: Promises	1.6.3
Chapter 4: Generators	1.6.4

---

<a href="#">Chapter 5: Program Performance</a>	1.6.5
<a href="#">Chapter 6: Benchmarking &amp; Tuning</a>	1.6.6
<a href="#">Appendix A: Library: asynquence</a>	1.6.7
<a href="#">Appendix B: Advanced Async Patterns</a>	1.6.8
<a href="#">ES.Next &amp; Beyond</a>	1.7
<a href="#">Chapter 1: ES? Now &amp; Future</a>	1.7.1
<a href="#">Chapter 2: Syntax</a>	1.7.2
<a href="#">Chapter 3: Organization</a>	1.7.3
<a href="#">Chapter 4: Async Flow Control</a>	1.7.4
<a href="#">Chapter 5: Collections</a>	1.7.5
<a href="#">Chapter 6: API Additions</a>	1.7.6
<a href="#">Chapter 7: Meta Programming</a>	1.7.7
<a href="#">Chapter 8: Beyond ES6</a>	1.7.8
<a href="#">Appendix A: TODO</a>	1.7.9

---

## You Don't Know JS Yet (book series) - 2nd Edition

This is a series of books diving deep into the core mechanisms of the JavaScript language.

### License & Copyright

The materials herein are all © 2019-2020 Kyle Simpson.



This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivs 4.0 Unported License](#).

## You Don't Know JS Yet: Get Started - 2nd Edition



# GET STARTED



---

[Purchase ebook/PDF from Leanpub](#)

---

### Table of Contents

- [Foreword \(by Brian Holt\)](#)
- [Preface](#)
- [Chapter 1: What Is JavaScript?](#)
- [Chapter 2: Surveying JS](#)
- [Chapter 3: Digging to the Roots of JS](#)
- [Chapter 4: The Bigger Picture](#)
- [Appendix A: Exploring Further](#)
- [Appendix B: Practice, Practice, Practice!](#)

# You Don't Know JS Yet: Get Started - 2nd Edition

## Chapter 1: What Is JavaScript?

You don't know JS, yet. Neither do I, not fully anyway. None of us do. But we can all start getting to know JS better.

In this first chapter of the first book of the *You Don't Know JS Yet* (YDKJSY) series, we will take some time to build a foundation to move forward on. We need to start by covering a variety of important background housekeeping details, clearing up some myths and misconceptions about what the language really is (and isn't!).

This is valuable insight into the identity and process of how JS is organized and maintained; all JS developers should understand it. If you want to get to know JS, this is how to *get started* taking the first steps in that journey.

## About This Book

I emphasize the word *journey* because *knowing JS* is not a destination, it's a direction. No matter how much time you spend with the language, you will always be able to find something else to learn and understand a little better. So don't look at this book as something to rush through for a quick achievement. Instead, patience and persistence are best as you take these first few steps.

Following this background chapter, the rest of the book lays out a high-level map of what you will find as you dig into and study JS with the YDKJSY books.

In particular, Chapter 4 identifies three main pillars around which the JS language is organized: scope/closures, prototypes/objects, and types/coercion. JS is a broad and sophisticated language, with many features and capabilities. But all of JS is founded on these three foundational pillars.

Keep in mind that even though this book is titled "Get Started," it's **not intended as a beginner/intro book**. This book's main job is to get you ready for studying JS deeply throughout the rest of the series; it's written assuming you already have familiarity with JS over at least several months experience before moving on in YDKJSY. So to get the most out *Get Started*, make sure you spend plenty of time writing JS code to build up your experience.

Even if you've already written a lot of JS before, this book should not be skimmed over or skipped; take your time to fully process the material here. **A good start always depends on a solid first step.**

## What's With That Name?

The name JavaScript is probably the most mistaken and misunderstood programming language name.

Is this language related to Java? Is it only the script form for Java? Is it only for writing scripts and not real programs?

The truth is, the name JavaScript is an artifact of marketing shenanigans. When Brendan Eich first conceived of the language, he code-named it Mocha. Internally at Netscape, the brand LiveScript was used. But when it came time to publicly name the language, "JavaScript" won the vote.

Why? Because this language was originally designed to appeal to an audience of mostly Java programmers, and because the word "script" was popular at the time to refer to lightweight programs. These lightweight "scripts" would be the first ones to embed inside of pages on this new thing called the web!

In other words, JavaScript was a marketing ploy to try to position this language as a palatable alternative to writing the heavier and more well-known Java of the day. It could just as easily have been called "WebJava," for that matter.

There are some superficial resemblances between JavaScript's code and Java code. Those similarities don't particularly come from shared development, but from both languages targeting developers with assumed syntax expectations from C (and to an extent, C++).

For example, we use the `{` to begin a block of code and the `}` to end that block of code, just like C/C++ and Java. We also use the `;` to punctuate the end of a statement.

In some ways, legal relationships run even deeper than the syntax. Oracle (via Sun), the company that still owns and runs Java, also owns the official trademark for the name "JavaScript" (via Netscape). This trademark is almost never enforced, and likely couldn't be at this point.

For these reasons, some have suggested we use JS instead of JavaScript. That is a very common shorthand, if not a good candidate for an official language branding itself. Indeed, these books use JS almost exclusively to refer to the language.

Further distancing the language from the Oracle-owned trademark, the official name of the language specified by TC39 and formalized by the ECMA standards body is **ECMAScript**. And indeed, since 2016, the official language name has also been suffixed by the revision year; as of this writing, that's ECMAScript 2019, or otherwise abbreviated ES2019.

In other words, the JavaScript/JS that runs in your browser or in Node.js, is *an* implementation of the ES2019 standard.

<b>NOTE:</b>
Don't use terms like "JS6" or "ES8" to refer to the language. Some do, but those terms only serve to perpetuate confusion. "ES20xx" or just "JS" are what you should stick to.

Don't use terms like "JS6" or "ES8" to refer to the language. Some do, but those terms only serve to perpetuate confusion. "ES20xx" or just "JS" are what you should stick to.

Whether you call it JavaScript, JS, ECMAScript, or ES2019, it's most definitely not a variant of the Java language!

"Java is to JavaScript as ham is to hamster." --Jeremy Keith, 2009

## Language Specification

I mentioned TC39, the technical steering committee that manages JS. Their primary task is managing the official specification for the language. They meet regularly to vote on any agreed changes, which they then submit to ECMA, the standards organization.

JS's syntax and behavior are defined in the ES specification.

ES2019 happens to be the 10th major numbered specification/revision since JS's inception in 1995, so in the specification's official URL as hosted by ECMA, you'll find "10.0":

<https://www.ecma-international.org/ecma-262/10.0/>

The TC39 committee is comprised of between 50 and about 100 different people from a broad section of web-invested companies, such as browser makers (Mozilla, Google, Apple) and device makers (Samsung, etc). All members of the committee are volunteers, though many of them are employees of these companies and so may receive compensation in part for their duties on the committee.

TC39 meets generally about every other month, usually for about three days, to review work done by members since the last meeting, discuss issues, and vote on proposals. Meeting locations rotate among member companies willing to host.

All TC39 proposals progress through a five-stage process—of course, since we're programmers, it's 0-based!—Stage 0 through Stage 4. You can read more about the Stage process here: <https://tc39.es/process-document/>

Stage 0 means roughly, someone on TC39 thinks it's a worthy idea and plans to champion and work on it. That means lots of ideas that non-TC39 members "propose," through informal means such as social media or blog posts, are really "pre-stage 0." You have to get a TC39 member to champion a proposal for it to be considered "Stage 0" officially.

Once a proposal reaches "Stage 4" status, it is eligible to be included in the next yearly revision of the language. It can take anywhere from several months to a few years for a proposal to work its way through these stages.

All proposals are managed in the open, on TC39's Github repository: <https://github.com/tc39/proposals>

Anyone, whether on TC39 or not, is welcome to participate in these public discussions and the processes for working on the proposals. However, only TC39 members can attend meetings and vote on the proposals and changes. So in effect, the voice of a TC39 member carries a lot of weight in where JS will go.

Contrary to some established and frustratingly perpetuated myth, there are *not* multiple versions of JavaScript in the wild. There's just **one JS**, the official standard as maintained by TC39 and ECMA.

Back in the early 2000s, when Microsoft maintained a forked and reverse-engineered (and not entirely compatible) version of JS called "JScript," there were legitimately "multiple versions" of JS. But those days are long gone. It's outdated and inaccurate to make such claims about JS today.

All major browsers and device makers have committed to keeping their JS implementations compliant with this one central specification. Of course, engines implement features at different times. But it should never be the case that the v8 engine (Chrome's JS engine) implements a specified feature differently or incompatibly as compared to the SpiderMonkey engine (Mozilla's JS engine).

That means you can learn **one JS**, and rely on that same JS everywhere.

## The Web Rules Everything About (JS)

While the array of environments that run JS is constantly expanding (from browsers, to servers (Node.js), to robots, to lightbulbs, to...), the one environment that rules JS is the web. In other words, how JS is implemented for web browsers is, in all practicality, the only reality that matters.

For the most part, the JS defined in the specification and the JS that runs in browser-based JS engines is the same. But there are some differences that must be considered.

Sometimes the JS specification will dictate some new or refined behavior, and yet that won't exactly match with how it works in browser-based JS engines. Such a mismatch is historical: JS engines have had 20+ years of observable behaviors around corner cases of features that have come to be relied on by web content. As such, sometimes the JS engines will refuse to conform to a specification-dictated change because it would break that web content.

In these cases, often TC39 will backtrack and simply choose to conform the specification to the reality of the web. For example, TC39 planned to add a `contains()` method for Arrays, but it was found that this name conflicted with old JS frameworks still in use on some sites, so they changed the name to a non-conflicting `includes()`. The same happened with a comedic/tragic JS *community crisis* dubbed "smooshgate," where the planned `flatten()` method was eventually renamed `flat()`.

But occasionally, TC39 will decide the specification should stick firm on some point even though it is unlikely that browser-based JS engines will ever conform.

The solution? Appendix B, "Additional ECMAScript Features for Web Browsers".<sup>specApB</sup> The JS specification includes this appendix to detail out any known mismatches between the official JS specification and the reality of JS on the web. In other words, these are exceptions that are allowed *only* for web JS; other JS environments must stick to the letter of the law.

Section B.1 and B.2 cover *additions* to JS (syntax and APIs) that web JS includes, again for historical reasons, but which TC39 does not plan to formally specify in the core of JS. Examples include `0`-prefixed octal literals, the global `escape(..)` / `unescape(..)` utilities, String "helpers" like `anchor()` and `blink()`, and the RegExp `compile(..)` method.

Section B.3 includes some conflicts where code may run in both web and non-web JS engines, but where the behavior *could* be observably different, resulting in different outcomes. Most of the listed changes involve situations that are labeled as early errors when code is running in strict mode.

Appendix B *gotchas* aren't encountered very often, but it's still a good idea to avoid these constructs to be future safe. Whenever possible, adhere to the JS specification and don't rely on behavior that's only applicable in certain JS engine environments.

## Not All (Web) JS...

Is this code a JS program?

```
alert("Hello, JS!");
```

Depends on how you look at things. The `alert(..)` function shown here is not included in the JS specification, but it *is* in all web JS environments. Yet, you won't find it in Appendix B, so what gives?

Various JS environments (like browser JS engines, Node.js, etc.) add APIs into the global scope of your JS programs that give you environment-specific capabilities, like being able to pop an alert-style box in the user's browser.

In fact, a wide range of JS-looking APIs, like `fetch(..)`, `getCurrentLocation(..)`, and `getUserMedia(..)`, are all web APIs that look like JS. In Node.js, we can access hundreds of API methods from various built-in modules, like `fs.write(..)`.

Another common example is `console.log(..)` (and all the other `console.*` methods!). These are not specified in JS, but because of their universal utility are defined by pretty much every JS environment, according to a roughly agreed consensus.

So `alert(..)` and `console.log(..)` are not defined by JS. But they *look* like JS. They are functions and object methods and they obey JS syntax rules. The behaviors behind them are controlled by the environment running the JS engine, but on the surface they definitely have to abide by JS to be able to play in the JS playground.

Most of the cross-browser differences people complain about with "JS is so inconsistent!" claims are actually due to differences in how those environment behaviors work, not in how the JS itself works.

So an `alert(..)` call *is* JS, but `alert` itself is really just a guest, not part of the official JS specification.

## It's Not Always JS

Using the console/REPL (Read-Evaluate-Print-Loop) in your browser's Developer Tools (or Node) feels like a pretty straightforward JS environment at first glance. But it's not, really.

Developer Tools are... tools for developers. Their primary purpose is to make life easier for developers. They prioritize DX (Developer Experience). It is *not* a goal of such tools to accurately and purely reflect all nuances of strict-spec JS behavior. As such, there's many quirks that may act as "gotchas" if you're treating the console as a *pure* JS environment.

This convenience is a good thing, by the way! I'm glad Developer Tools make developers' lives easier! I'm glad we have nice UX charms like auto-complete of variables/properties, etc. I'm just pointing out that we can't and shouldn't expect such tools to *always* adhere strictly to the way JS programs are handled, because that's not the purpose of these tools.

Since such tools vary in behavior from browser to browser, and since they change (sometimes rather frequently), I'm not going to "hardcode" any of the specific details into this text, thereby ensuring this book text is outdated quickly.

But I'll just hint at some examples of quirks that have been true at various points in different JS console environments, to reinforce my point about not assuming native JS behavior while using them:

- Whether a `var` or `function` declaration in the top-level "global scope" of the console actually creates a real global variable (and mirrored `window` property, and vice versa!).
- What happens with multiple `let` and `const` declarations in the top-level "global scope".
- Whether `"use strict";` on one line-entry (pressing `<enter>` after) enables strict mode for the rest of that console session, the way it would on the first line of a `.js` file, as well as whether you can use `"use strict";` beyond the "first line" and still get strict mode turned on for that session.
- How non-strict mode `this` default-binding works for function calls, and whether the "global object" used will contain expected global variables.

- How hoisting (see Book 2, *Scope & Closures*) works across multiple line entries.
- ...several others

The developer console is not trying to pretend to be a JS compiler that handles your entered code exactly the same way the JS engine handles a .js file. It's trying to make it easy for you to quickly enter a few lines of code and see the results immediately. These are entirely different use cases, and as such, it's unreasonable to expect one tool to handle both equally.

Don't trust what behavior you see in a developer console as representing *exact* to-the-letter JS semantics; for that, read the specification. Instead, think of the console as a "JS-friendly" environment. That's useful in its own right.

## Many Faces

The term "paradigm" in programming language context refers to a broad (almost universal) mindset and approach to structuring code. Within a paradigm, there are myriad variations of style and form that distinguish programs, including countless different libraries and frameworks that leave their unique signature on any given code.

But no matter what a program's individual style may be, the big picture divisions around paradigms are almost always evident at first glance of any program.

Typical paradigm-level code categories include procedural, object-oriented (OO/classes), and functional (FP):

- Procedural style organizes code in a top-down, linear progression through a pre-determined set of operations, usually collected together in related units called procedures.
- OO style organizes code by collecting logic and data together into units called classes.
- FP style organizes code into functions (pure computations as opposed to procedures), and the adaptations of those functions as values.

Paradigms are neither right nor wrong. They're orientations that guide and mold how programmers approach problems and solutions, how they structure and maintain their code.

Some languages are heavily slanted toward one paradigm—C is procedural, Java/C++ are almost entirely class oriented, and Haskell is FP through and through.

But many languages also support code patterns that can come from, and even mix and match from, different paradigms. So called "multi-paradigm languages" offer ultimate flexibility. In some cases, a single program can even have two or more expressions of these paradigms sitting side by side.

JavaScript is most definitely a multi-paradigm language. You can write procedural, class-oriented, or FP-style code, and you can make those decisions on a line-by-line basis instead of being forced into an all-or-nothing choice.

## Backwards & Forwards

One of the most foundational principles that guides JavaScript is preservation of *backwards compatibility*. Many are confused by the implications of this term, and often confuse it with a related but different term: *forwards compatibility*.

Let's set the record straight.

Backwards compatibility means that once something is accepted as valid JS, there will not be a future change to the language that causes that code to become invalid JS. Code written in 1995—however primitive or limited it may have been!—should still work today. As TC39 members often proclaim, "we don't break the web!"

The idea is that JS developers can write code with confidence that their code won't stop working unpredictably because a browser update is released. This makes the decision to choose JS for a program a more wise and safe investment, for years into the future.

That "guarantee" is no small thing. Maintaining backwards compatibility, stretched out across almost 25 years of the language's history, creates an enormous burden and a whole slew of unique challenges. You'd be hard pressed to find many other examples in computing of such a commitment to backwards compatibility.

The costs of sticking to this principle should not be casually dismissed. It necessarily creates a very high bar to including changing or extending the language; any decision becomes effectively permanent, mistakes and all. Once it's in JS, it can't be taken out because it might break programs, even if we'd really, really like to remove it!

There are some small exceptions to this rule. JS has had some backwards-incompatible changes, but TC39 is extremely cautious in doing so. They study existing code on the web (via browser data gathering) to estimate the impact of such breakage, and browsers ultimately decide and vote on whether they're willing to take the heat from users for a very small-scale breakage weighed against the benefits of fixing or improving some aspect of the language for many more sites (and users).

These kinds of changes are rare, and are almost always in corner cases of usage that are unlikely to be observably breaking in many sites.

Compare *backwards compatibility* to its counterpart, *forwards compatibility*. Being forwards-compatible means that including a new addition to the language in a program would not cause that program to break if it were run in an older JS engine. **JS is not forwards-compatible**, despite many wishing such, and even incorrectly believing the myth that it is.

HTML and CSS, by contrast, are forwards-compatible but not backwards-compatible. If you dug up some HTML or CSS written back in 1995, it's entirely possible it would not work (or work the same) today. But, if you use a new feature from 2019 in a browser from 2010, the page isn't "broken" -- the unrecognized CSS/HTML is skipped over, while the rest of the CSS/HTML would be processed accordingly.

It may seem desirable for forwards-compatibility to be included in programming language design, but it's generally impractical to do so. Markup (HTML) or styling (CSS) are declarative in nature, so it's much easier to "skip over" unrecognized declarations with minimal impact to other recognized declarations.

But chaos and non-determinism would ensue if a programming language engine selectively skipped statements (or even expressions!) that it didn't understand, as it's impossible to ensure that a subsequent part of the program wasn't expecting the skipped-over part to have been processed.

Though JS isn't, and can't be, forwards-compatible, it's critical to recognize JS's backwards compatibility, including the enduring benefits to the web and the constraints and difficulties it places on JS as a result.

## Jumping the Gaps

Since JS is not forwards-compatible, it means that there is always the potential for a gap between code that you can write that's valid JS, and the oldest engine that your site or application needs to support. If you run a program that uses an ES2019 feature in an engine from 2016, you're very likely to see the program break and crash.

If the feature is a new syntax, the program will in general completely fail to compile and run, usually throwing a syntax error. If the feature is an API (such as ES6's `object.is(...)`), the program may run up to a point but then throw a runtime exception and stop once it encounters the reference to the unknown API.

Does this mean JS developers should always lag behind the pace of progress, using only code that is on the trailing edge of the oldest JS engine environments they need to support? No!

But it does mean that JS developers need to take special care to address this gap.

For new and incompatible syntax, the solution is transpiling. Transpiling is a contrived and community-invented term to describe using a tool to convert the source code of a program from one form to another (but still as textual source code). Typically, forwards-compatibility problems related to syntax are solved by using a transpiler (the most common one being Babel (<https://babeljs.io>)) to convert from that newer JS syntax version to an equivalent older syntax.

For example, a developer may write a snippet of code like:

```

if (something) {
  let x = 3;
  console.log(x);
}
else {
  let x = 4;
  console.log(x);
}

```

This is how the code would look in the source code tree for that application. But when producing the file(s) to deploy to the public website, the Babel transpiler might convert that code to look like this:

```

var x$0, x$1;
if (something) {
  x$0 = 3;
  console.log(x$0);
}
else {
  x$1 = 4;
  console.log(x$1);
}

```

The original snippet relied on `let` to create block-scoped `x` variables in both the `if` and `else` clauses which did not interfere with each other. An equivalent program (with minimal re-working) that Babel can produce just chooses to name two different variables with unique names, producing the same non-interference outcome.

#### NOTE:

The `let` keyword was added in ES6 (in 2015). The preceding example of transpiling would only need to apply if an application needed to run in a pre-ES6 supporting JS environment. The example here is just for simplicity of illustration. When ES6 was new, the need for such a transpilation was quite prevalent, but in 2020 it's much less common to need to support pre-ES6 environments. The "target" used for transpilation is thus a sliding window that shifts upward only as decisions are made for a site/application to stop supporting some old browser/engine.

You may wonder: why go to the trouble of using a tool to convert from a newer syntax version to an older one? Couldn't we just write the two variables and skip using the `let` keyword? The reason is, it's strongly recommended that developers use the latest version of JS so that their code is clean and communicates its ideas most effectively.

Developers should focus on writing the clean, new syntax forms, and let the tools take care of producing a forwards-compatible version of that code that is suitable to deploy and run on the oldest-supported JS engine environments.

## Filling the Gaps

If the forwards-compatibility issue is not related to new syntax, but rather to a missing API method that was only recently added, the most common solution is to provide a definition for that missing API method that stands in and acts as if the older environment had already had it natively defined. This pattern is called a polyfill (aka "shim").

Consider this code:

```

// getSomeRecords() returns us a promise for some
// data it will fetch
var pr = getSomeRecords();

// show the UI spinner while we get the data
startSpinner();

pr
  .then(renderRecords) // render if successful
  .catch(showError) // show an error if not
  .finally(hideSpinner) // always hide the spinner

```

This code uses an ES2019 feature, the `finally(..)` method on the promise prototype. If this code were used in a pre-ES2019 environment, the `finally(..)` method would not exist, and an error would occur.

A polyfill for `finally(..)` in pre-ES2019 environments could look like this:

```
if (!Promise.prototype.finally) {
  Promise.prototype.finally = function fn(fn){
    return this.then(
      function t(v){
        return Promise.resolve( fn() )
          .then(function t(){
            return v;
          });
      },
      function c(e){
        return Promise.resolve( fn() )
          .then(function t(){
            throw e;
          });
      }
    );
  };
}
```

#### WARNING:

This is only a simple illustration of a basic (not entirely spec-compliant) polyfill for `finally(..)`. Don't use this polyfill in your code; always use a robust, official polyfill wherever possible, such as the collection of polyfills/shims in ES-Shim.

The `if` statement protects the polyfill definition by preventing it from running in any environment where the JS engine has already defined that method. In older environments, the polyfill is defined, but in newer environments the `if` statement is quietly skipped.

Transpilers like Babel typically detect which polyfills your code needs and provide them automatically for you. But occasionally you may need to include/define them explicitly, which works similar to the snippet we just looked at.

Always write code using the most appropriate features to communicate its ideas and intent effectively. In general, this means using the most recent stable JS version. Avoid negatively impacting the code's readability by trying to manually adjust for the syntax/API gaps. That's what tools are for!

Transpilation and polyfilling are two highly effective techniques for addressing that gap between code that uses the latest stable features in the language and the old environments a site or application needs to still support. Since JS isn't going to stop improving, the gap will never go away. Both techniques should be embraced as a standard part of every JS project's production chain going forward.

## What's in an Interpretation?

A long-debated question for code written in JS: is it an interpreted script or a compiled program? The majority opinion seems to be that JS is an interpreted (scripting) language. But the truth is more complicated than that.

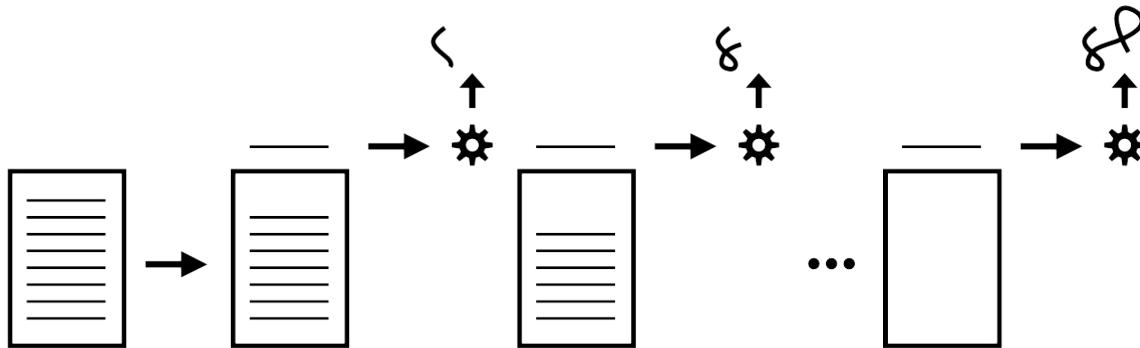
For much of the history of programming languages, "interpreted" languages and "scripting" languages have been looked down on as inferior compared to their compiled counterparts. The reasons for this acrimony are numerous, including the perception that there is a lack of performance optimization, as well as dislike of certain language characteristics, such as scripting languages generally using dynamic typing instead of the "more mature" statically typed languages.

Languages regarded as "compiled" usually produce a portable (binary) representation of the program that is distributed for execution later. Since we don't really observe that kind of model with JS (we distribute the source code, not the binary form), many claim that disqualifies JS from the category. In reality, the distribution model for a program's "executable" form has become

drastically more varied and also less relevant over the last few decades; to the question at hand, it doesn't really matter so much anymore what form of a program gets passed around.

These misinformed claims and criticisms should be set aside. The real reason it matters to have a clear picture on whether JS is interpreted or compiled relates to the nature of how errors are handled.

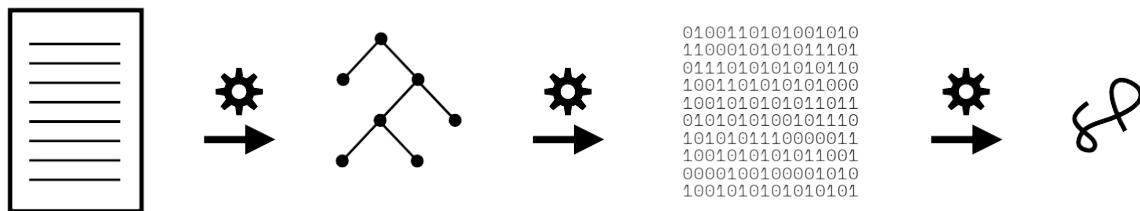
Historically, scripted or interpreted languages were executed in generally a top-down and line-by-line fashion; there's typically not an initial pass through the program to process it before execution begins (see Figure 1).



*Fig. 1: Interpreted/Scripted Execution*

In scripted or interpreted languages, an error on line 5 of a program won't be discovered until lines 1 through 4 have already executed. Notably, the error on line 5 might be due to a runtime condition, such as some variable or value having an unsuitable value for an operation, or it may be due to a malformed statement/command on that line. Depending on context, deferring error handling to the line the error occurs on may be a desirable or undesirable effect.

Compare that to languages which do go through a processing step (typically, called parsing) before any execution occurs, as illustrated in Figure 2:



*Fig. 2: Parsing + Compilation + Execution*

In this processing model, an invalid command (such as broken syntax) on line 5 would be caught during the parsing phase, before any execution has begun, and none of the program would run. For catching syntax (or otherwise "static") errors, generally it's preferred to know about them ahead of any doomed partial execution.

So what do "parsed" languages have in common with "compiled" languages? First, all compiled languages are parsed. So a parsed language is quite a ways down the road toward being compiled already. In classic compilation theory, the last remaining step after parsing is code generation: producing an executable form.

Once any source program has been fully parsed, it's very common that its subsequent execution will, in some form or fashion, include a translation from the parsed form of the program—usually called an Abstract Syntax Tree (AST)—to that executable form.

In other words, parsed languages usually also perform code generation before execution, so it's not that much of a stretch to say that, in spirit, they're compiled languages.

JS source code is parsed before it is executed. The specification requires as much, because it calls for "early errors"—statically determined errors in code, such as a duplicate parameter name—to be reported before the code starts executing. Those errors cannot be recognized without the code having been parsed.

So **JS is a parsed language**, but is it *compiled*?

The answer is closer to yes than no. The parsed JS is converted to an optimized (binary) form, and that "code" is subsequently executed (Figure 2); the engine does not commonly switch back into line-by-line execution (like Figure 1) mode after it has finished all the hard work of parsing—most languages/engines wouldn't, because that would be highly inefficient.

To be specific, this "compilation" produces a binary byte code (of sorts), which is then handed to the "JS virtual machine" to execute. Some like to say this VM is "interpreting" the byte code. But then that means Java, and a dozen other JVM-driven languages, for that matter, are interpreted rather than compiled. Of course, that contradicts the typical assertion that Java/etc are compiled languages.

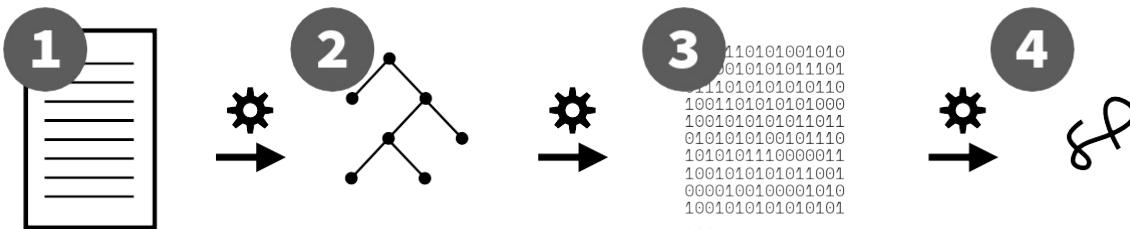
Interestingly, while Java and JavaScript are very different languages, the question of interpreted/compiled is pretty closely related between them!

Another wrinkle is that JS engines can employ multiple passes of JIT (Just-In-Time) processing/optimization on the generated code (post parsing), which again could reasonably be labeled either "compilation" or "interpretation" depending on perspective. It's actually a fantastically complex situation under the hood of a JS engine.

So what do these nitty-gritty details boil down to? Step back and consider the entire flow of a JS source program:

1. After a program leaves a developer's editor, it gets transpiled by Babel, then packed by Webpack (and perhaps half a dozen other build processes), then it gets delivered in that very different form to a JS engine.
2. The JS engine parses the code to an AST.
3. Then the engine converts that AST to a kind-of byte code, a binary intermediate representation (IR), which is then refined/converted even further by the optimizing JIT compiler.
4. Finally, the JS VM executes the program.

To visualize those steps, again:



*Fig. 3: Parsing, Compiling, and Executing JS*

Is JS handled more like an interpreted, line-by-line script, as in Figure 1, or is it handled more like a compiled language that's processed in one-to-several passes first, before execution (as in Figures 2 and 3)?

I think it's clear that in spirit, if not in practice, **JS is a compiled language**.

And again, the reason that matters is, since JS is compiled, we are informed of static errors (such as malformed syntax) before our code is executed. That is a substantively different interaction model than we get with traditional "scripting" programs, and arguably more helpful!

## Web Assembly (WASM)

One dominating concern that has driven a significant amount of JS's evolution is performance, both how quickly JS can be parsed/compiled and how quickly that compiled code can be executed.

In 2013, engineers from Mozilla Firefox demonstrated a port of the Unreal 3 game engine from C to JS. The ability for this code to run in a browser JS engine at full 60fps performance was predicated on a set of optimizations that the JS engine could perform specifically because the JS version of the Unreal engine's code used a style of code that favored a subset of the JS language, named "ASM.js".

This subset is valid JS written in ways that are somewhat uncommon in normal coding, but which signal certain important typing information to the engine that allow it to make key optimizations. ASM.js was introduced as one way of addressing the pressures on the runtime performance of JS.

But it's important to note that ASM.js was never intended to be code that was authored by developers, but rather a representation of a program having been transpiled from another language (such as C), where these typing "annotations" were inserted automatically by the tooling.

Several years after ASM.js demonstrated the validity of tooling-created versions of programs that can be processed more efficiently by the JS engine, another group of engineers (also, initially, from Mozilla) released Web Assembly (WASM).

WASM is similar to ASM.js in that its original intent was to provide a path for non-JS programs (C, etc.) to be converted to a form that could run in the JS engine. Unlike ASM.js, WASM chose to additionally get around some of the inherent delays in JS parsing/compilation before a program can execute, by representing the program in a form that is entirely unlike JS.

WASM is a representation format more akin to Assembly (hence, its name) that can be processed by a JS engine by skipping the parsing/compilation that the JS engine normally does. The parsing/compilation of a WASM-targeted program happen ahead of time (AOT); what's distributed is a binary-packed program ready for the JS engine to execute with very minimal processing.

An initial motivation for WASM was clearly the potential performance improvements. While that continues to be a focus, WASM is additionally motivated by the desire to bring more parity for non-JS languages to the web platform. For example, if a language like Go supports threaded programming, but JS (the language) does not, WASM offers the potential for such a Go program to be converted to a form the JS engine can understand, without needing a threads feature in the JS language itself.

In other words, WASM relieves the pressure to add features to JS that are mostly/exclusively intended to be used by transpiled programs from other languages. That means JS feature development can be judged (by TC39) without being skewed by interests/demands in other language ecosystems, while still letting those languages have a viable path onto the web.

Another perspective on WASM that's emerging is, interestingly, not even directly related to the web (W). WASM is evolving to become a cross-platform virtual machine (VM) of sorts, where programs can be compiled once and run in a variety of different system environments.

So, WASM isn't only for the web, and WASM also isn't JS. Ironically, even though WASM runs in the JS engine, the JS language is one of the least suitable languages to source WASM programs with, because WASM relies heavily on static typing information. Even TypeScript (TS)—ostensibly, JS + static types—is not quite suitable (as it stands) to transpile to WASM, though language variants like AssemblyScript are attempting to bridge the gap between JS/TS and WASM.

This book isn't about WASM, so I won't spend much more time discussing it, except to make one final point. Some folks have suggested WASM points to a future where JS is excised from, or minimized in, the web. These folks often harbor ill feelings about JS, and want some other language—any other language!—to replace it. Since WASM lets other languages run in the JS engine, on its face this isn't an entirely fanciful fairytale.

But let me just state simply: WASM will not replace JS. WASM significantly augments what the web (including JS) can accomplish. That's a great thing, entirely orthogonal to whether some people will use it as an escape hatch from having to write JS.

## Strictly Speaking

---

Back in 2009 with the release of ES5, JS added *strict mode* as an opt-in mechanism for encouraging better JS programs.

The benefits of strict mode far outweigh the costs, but old habits die hard and the inertia of existing (aka "legacy") code bases is really hard to shift. So sadly, more than 10 years later, strict mode's *optionality* means that it's still not necessarily the default for JS programmers.

Why strict mode? Strict mode shouldn't be thought of as a restriction on what you can't do, but rather as a guide to the best way to do things so that the JS engine has the best chance of optimizing and efficiently running the code. Most JS code is worked on by teams of developers, so the *strict-ness* of strict mode (along with tooling like linters!) often helps collaboration on code by avoiding some of the more problematic mistakes that slip by in non-strict mode.

Most strict mode controls are in the form of *early errors*, meaning errors that aren't strictly syntax errors but are still thrown at compile time (before the code is run). For example, strict mode disallows naming two function parameters the same, and results in an early error. Some other strict mode controls are only observable at runtime, such as how `this` defaults to `undefined` instead of the global object.

Rather than fighting and arguing with strict mode, like a kid who just wants to defy whatever their parents tell them not to do, the best mindset is that strict mode is like a linter reminding you how JS *should* be written to have the highest quality and best chance at performance. If you find yourself feeling handcuffed, trying to work around strict mode, that should be a blaring red warning flag that you need to back up and rethink the whole approach.

Strict mode is switched on per file with a special pragma (nothing allowed before it except comments/whitespace):

```
// only whitespace and comments are allowed
// before the use-strict pragma
"use strict";
// the rest of the file runs in strict mode
```

#### WARNING:

Something to be aware of is that even a stray `;` all by itself appearing before the strict mode pragma will render the pragma useless; no errors are thrown because it's valid JS to have a string literal expression in a statement position, but it also will silently *not* turn on strict mode!

Strict mode can alternatively be turned on per-function scope, with exactly the same rules about its surroundings:

```
function someOperations() {
    // whitespace and comments are fine here
    "use strict";

    // all this code will run in strict mode
}
```

Interestingly, if a file has strict mode turned on, the function-level strict mode pragmas are disallowed. So you have to pick one or the other.

The **only** valid reason to use a per-function approach to strict mode is when you are converting an existing non-strict mode program file and need to make the changes little by little over time. Otherwise, it's vastly better to simply turn strict mode on for the entire file/program.

Many have wondered if there would ever be a time when JS made strict mode the default? The answer is, almost certainly not. As we discussed earlier around backwards compatibility, if a JS engine update started assuming code was strict mode even if it's not marked as such, it's possible that this code would break as a result of strict mode's controls.

However, there are a few factors that reduce the future impact of this non-default "obscurity" of strict mode.

For one, virtually all transpiled code ends up in strict mode even if the original source code isn't written as such. Most JS code in production has been transpiled, so that means most JS is already adhering to strict mode. It's possible to undo that assumption, but you really have to go out of your way to do so, so it's highly unlikely.

Moreover, a wide shift is happening toward more/most new JS code being written using the ES6 module format. ES6 modules assume strict mode, so all code in such files is automatically defaulted to strict mode.

Taken together, strict mode is largely the de facto default even though technically it's not actually the default.

## Defined

JS is an implementation of the ECMAScript standard (version ES2019 as of this writing), which is guided by the TC39 committee and hosted by ECMA. It runs in browsers and other JS environments such as Node.js.

JS is a multi-paradigm language, meaning the syntax and capabilities allow a developer to mix and match (and bend and reshape!) concepts from various major paradigms, such as procedural, object-oriented (OO/classes), and functional (FP).

JS is a compiled language, meaning the tools (including the JS engine) process and verify a program (reporting any errors!) before it executes.

With our language now *defined*, let's start getting to know its ins and outs.

specApB<sup>1</sup>. ECMAScript 2019 Language Specification, Appendix B: Additional ECMAScript Features for Web Browsers,  
<https://www.ecma-international.org/ecma-262/10.0/#sec-additional-ecmascript-features-for-web-browsers> (latest as of time  
of this writing in January 2020) ↪

# You Don't Know JS Yet: Get Started - 2nd Edition

## Chapter 2: Surveying JS

The best way to learn JS is to start writing JS.

To do that, you need to know how the language works, and that's what we'll focus on here. Even if you've programmed in other languages before, take your time getting comfortable with JS, and make sure to practice each piece.

This chapter is not an exhaustive reference on every bit of syntax of the JS language. It's also not intended to be a complete "intro to JS" primer.

Instead, we're just going to survey some of the major topic areas of the language. Our goal is to get a better *feel* for it, so that we can move forward writing our own programs with more confidence. We'll revisit many of these topics in successively more detail as you go through the rest of this book, and the rest of the series.

Please don't expect this chapter to be a quick read. It's long and there's plenty of detail to chew on. Take your time.

**TIP:**

If you're still getting familiar with JS, I suggest you reserve plenty of extra time to work through this chapter. Take each section and ponder and explore the topic for awhile. Look through existing JS programs and compare what you see in them to the code and explanations (and opinions!) presented here. You will get a lot more out of the rest of the book and series with a solid foundation of JS's *nature*.

## Each File is a Program

Almost every website (web application) you use is comprised of many different JS files (typically with the .js file extension). It's tempting to think of the whole thing (the application) as one program. But JS sees it differently.

In JS, each standalone file is its own separate program.

The reason this matters is primarily around error handling. Since JS treats files as programs, one file may fail (during parse/compile or execution) and that will not necessarily prevent the next file from being processed. Obviously, if your application depends on five .js files, and one of them fails, the overall application will probably only partially operate, at best. It's important to ensure that each file works properly, and that to whatever extent possible, they handle failure in other files as gracefully as possible.

It may surprise you to consider separate .js files as separate JS programs. From the perspective of your usage of an application, it sure seems like one big program. That's because the execution of the application allows these individual *programs* to cooperate and act as one program.

**NOTE:**

Many projects use build process tools that end up combining separate files from the project into a single file to be delivered to a web page. When this happens, JS treats this single combined file as the entire program.

The only way multiple standalone .js files act as a single program is by sharing their state (and access to their public functionality) via the "global scope." They mix together in this global scope namespace, so at runtime they act as a whole.

Since ES6, JS has also supported a module format in addition to the typical standalone JS program format. Modules are also file-based. If a file is loaded via module-loading mechanism such as an `import` statement or a `<script type="module">` tag, all its code is treated as a single module.

Though you wouldn't typically think about a module—a collection of state and publicly exposed methods to operate on that state

Though you wouldn't typically think about a module—a collection of state and publicly exposed methods to operate on that state—as a standalone program, JS does in fact still treat each module separately. Similar to how "global scope" allows standalone files to mix together at runtime, importing a module into another allows runtime interoperation between them.

Regardless of which code organization pattern (and loading mechanism) is used for a file (standalone or module), you should still think of each file as its own (mini) program, which may then cooperate with other (mini) programs to perform the functions of your overall application.

## Values

The most fundamental unit of information in a program is a value. Values are data. They're how the program maintains state. Values come in two forms in JS: **primitive** and **object**.

Values are embedded in programs using *literals*:

```
greeting("My name is Kyle.");
```

In this program, the value `"My name is Kyle."` is a primitive string literal; strings are ordered collections of characters, usually used to represent words and sentences.

I used the double-quote `"` character to *delimit* (surround, separate, define) the string value. But I could have used the single-quote `'` character as well. The choice of which quote character is entirely stylistic. The important thing, for code readability and maintainability sake, is to pick one and to use it consistently throughout the program.

Another option to delimit a string literal is to use the back-tick ``` character. However, this choice is not merely stylistic; there's a behavioral difference as well. Consider:

```
console.log("My name is ${ firstName }.");
// My name is ${ firstName }.

console.log('My name is ${ firstName }.');
// My name is ${ firstName }.

console.log(`My name is ${ firstName }.`);
// My name is Kyle.
```

Assuming this program has already defined a variable `firstName` with the string value `"Kyle"`, the ```-delimited string then resolves the variable expression (indicated with ``${ ... }``) to its current value. This is called **interpolation**.

The back-tick ```-delimited string can be used without including interpolated expressions, but that defeats the whole purpose of that alternate string literal syntax:

```
console.log(
  `Am I confusing you by omitting interpolation?`
);
// Am I confusing you by omitting interpolation?
```

The better approach is to use `"` or `'` (again, pick one and stick to it!) for strings *unless you need* interpolation; reserve ``` only for strings that will include interpolated expressions.

Other than strings, JS programs often contain other primitive literal values such as booleans and numbers:

```
while (false) {
  console.log(3.141592);
}
```

`while` represents a loop type, a way to repeat operations *while* its condition is true.

In this case, the loop will never run (and nothing will be printed), because we used the `false` boolean value as the loop conditional. `true` would have resulted in a loop that keeps going forever, so be careful!

The number `3.141592` is, as you may know, an approximation of mathematical PI to the first six digits. Rather than embed such a value, however, you would typically use the predefined `Math.PI` value for that purpose. Another variation on numbers is the `bigint` (big-integer) primitive type, which is used for storing arbitrarily large numbers.

Numbers are most often used in programs for counting steps, such as loop iterations, and accessing information in numeric positions (i.e., an array index). We'll cover arrays/objects in a little bit, but as an example, if there was an array called `names`, we could access the element in its second position like this:

```
console.log(`My name is ${ names[1] }`);  
// My name is Kyle.
```

We used `1` for the element in the second position, instead of `2`, because like in most programming languages, JS array indices are 0-based (`0` is the first position).

In addition to strings, numbers, and booleans, two other *primitive* values in JS programs are `null` and `undefined`. While there are differences between them (some historic and some contemporary), for the most part both values serve the purpose of indicating *emptiness* (or absence) of a value.

Many developers prefer to treat them both consistently in this fashion, which is to say that the values are assumed to be indistinguishable. If care is taken, this is often possible. However, it's safest and best to use only `undefined` as the single empty value, even though `null` seems attractive in that it's shorter to type!

```
while (value != undefined) {  
    console.log("Still got something!");  
}
```

The final primitive value to be aware of is a symbol, which is a special-purpose value that behaves as a hidden unguessable value. Symbols are almost exclusively used as special keys on objects:

```
hitchhikersGuide[ Symbol("meaning of life") ];  
// 42
```

You won't encounter direct usage of symbols very often in typical JS programs. They're mostly used in low-level code such as in libraries and frameworks.

## Arrays And Objects

Besides primitives, the other value type in JS is an object value.

As mentioned earlier, arrays are a special type of object that's comprised of an ordered and numerically indexed list of data:

```
var names = [ "Frank", "Kyle", "Peter", "Susan" ];  
  
names.length;  
// 4  
  
names[0];  
// Frank  
  
names[1];  
// Kyle
```

JS arrays can hold any value type, either primitive or object (including other arrays). As we'll see toward the end of Chapter 3, even functions are values that can be held in arrays or objects.

**NOTE:**

Functions, like arrays, are a special kind (aka, sub-type) of object. We'll cover functions in more detail in a bit.

Objects are more general: an unordered, keyed collection of any various values. In other words, you access the element by a string location name (aka "key" or "property") rather than by its numeric position (as with arrays). For example:

```
var me = {
  first: "Kyle",
  last: "Simpson",
  age: 39,
  specialties: [ "JS", "Table Tennis" ]
};

console.log(`My name is ${me.first}.`);
```

Here, `me` represents an object, and `first` represents the name of a location of information in that object (value collection).

Another syntax option that accesses information in an object by its property/key uses the square-brackets `[ ]`, such as

`me["first"]`.

## Value Type Determination

For distinguishing values, the `typeof` operator tells you its built-in type, if primitive, or "object" otherwise:

```
typeof 42;           // "number"
typeof "abc";        // "string"
typeof true;         // "boolean"
typeof undefined;   // "undefined"
typeof null;         // "object" -- oops, bug!
typeof { "a": 1 };   // "object"
typeof [1,2,3];     // "object"
typeof function hello(){}; // "function"
```

**WARNING:**

`typeof null` unfortunately returns "object" instead of the expected "null". Also, `typeof` returns the specific "function" for functions, but not the expected "array" for arrays.

Converting from one value type to another, such as from string to number, is referred to in JS as "coercion." We'll cover this in more detail later in this chapter.

Primitive values and object values behave differently when they're assigned or passed around. We'll cover these details in Appendix A, "Values vs References."

## Declaring and Using Variables

To be explicit about something that may not have been obvious in the previous section: in JS programs, values can either appear as literal values (as many of the preceding examples illustrate), or they can be held in variables; think of variables as just containers for values.

Variables have to be declared (created) to be used. There are various syntax forms that declare variables (aka, "identifiers"), and each form has different implied behaviors.

For example, consider the `var` statement:

```
var myName = "Kyle";
var age;
```

The `var` keyword declares a variable to be used in that part of the program, and optionally allows an initial assignment of a value.

Another similar keyword is `let` :

```
let myName = "Kyle";
let age;
```

The `let` keyword has some differences to `var`, with the most obvious being that `let` allows a more limited access to the variable than `var`. This is called "block scoping" as opposed to regular or function scoping.

Consider:

```
var adult = true;

if (adult) {
  var myName = "Kyle";
  let age = 39;
  console.log("Shhh, this is a secret!");
}

console.log(myName);
// Kyle

console.log(age);
// Error!
```

The attempt to access `age` outside of the `if` statement results in an error, because `age` was block-scoped to the `if`, whereas `myName` was not.

Block-scoping is very useful for limiting how widespread variable declarations are in our programs, which helps prevent accidental overlap of their names.

But `var` is still useful in that it communicates "this variable will be seen by a wider scope (of the whole function)". Both declaration forms can be appropriate in any given part of a program, depending on the circumstances.

#### NOTE:

It's very common to suggest that `var` should be avoided in favor of `let` (or `const`!), generally because of perceived confusion over how the scoping behavior of `var` has worked since the beginning of JS. I believe this to be overly restrictive advice and ultimately unhelpful. It's assuming you are unable to learn and use a feature properly in combination with other features. I believe you *can* and *should* learn any features available, and use them where appropriate!

A third declaration form is `const`. It's like `let` but has an additional limitation that it must be given a value at the moment it's declared, and cannot be re-assigned a different value later.

Consider:

```
const myBirthday = true;
let age = 39;

if (myBirthday) {
  age = age + 1;    // OK!
  myBirthday = false; // Error!
}
```

The `myBirthday` constant is not allowed to be re-assigned.

`const` declared variables are not "unchangeable", they just cannot be re-assigned. It's ill-advised to use `const` with object values, because those values can still be changed even though the variable can't be re-assigned. This leads to potential confusion down the line, so I think it's wise to avoid situations like:

```
const actors = [
  "Morgan Freeman", "Jennifer Aniston"
];

actors[2] = "Tom Cruise"; // OK :(
actors = []; // Error!
```

The best semantic use of a `const` is when you have a simple primitive value that you want to give a useful name to, such as using `myBirthday` instead of `true`. This makes programs easier to read.

**TIP:**

If you stick to using `const` only with primitive values, you avoid any confusion of re-assignment (not allowed) vs. mutation (allowed)! That's the safest and best way to use `const`.

Besides `var` / `let` / `const`, there are other syntactic forms that declare identifiers (variables) in various scopes. For example:

```
function hello(myName) {
  console.log(`Hello, ${myName}`);
}

hello("Kyle");
// Hello, Kyle.
```

The identifier `hello` is created in the outer scope, and it's also automatically associated so that it references the function. But the named parameter `myName` is created only inside the function, and thus is only accessible inside that function's scope. `hello` and `myName` generally behave as `var`-declared.

Another syntax that declares a variable is a `catch` clause:

```
try {
  someError();
}
catch (err) {
  console.log(err);
}
```

The `err` is a block-scoped variable that exists only inside the `catch` clause, as if it had been declared with `let`.

## Functions

The word "function" has a variety of meanings in programming. For example, in the world of Functional Programming, "function" has a precise mathematical definition and implies a strict set of rules to abide by.

In JS, we should consider "function" to take the broader meaning of another related term: "procedure." A procedure is a collection of statements that can be invoked one or more times, may be provided some inputs, and may give back one or more outputs.

From the early days of JS, function definition looked like:

```
function awesomeFunction(coolThings) {
  // ..
  return amazingStuff;
}
```

This is called a function declaration because it appears as a statement by itself, not as an expression in another statement. The association between the identifier `awesomeFunction` and the function value happens during the compile phase of the code, before that code is executed.

In contrast to a function declaration statement, a function expression can be defined and assigned like this:

```
// let awesomeFunction = ...
// const awesomeFunction = ...
var awesomeFunction = function(coolThings) {
  // ...
  return amazingStuff;
};
```

This function is an expression that is assigned to the variable `awesomeFunction`. Different from the function declaration form, a function expression is not associated with its identifier until that statement during runtime.

It's extremely important to note that in JS, functions are values that can be assigned (as shown in this snippet) and passed around. In fact, JS functions are a special type of the object value type. Not all languages treat functions as values, but it's essential for a language to support the functional programming pattern, as JS does.

JS functions can receive parameter input:

```
function greeting(myName) {
  console.log(`Hello, ${ myName }!`);
}

greeting("Kyle"); // Hello, Kyle!
```

In this snippet, `myName` is called a parameter, which acts as a local variable inside the function. Functions can be defined to receive any number of parameters, from none upward, as you see fit. Each parameter is assigned the argument value that you pass in that position (`"Kyle"`, here) of the call.

Functions also can return values using the `return` keyword:

```
function greeting(myName) {
  return `Hello, ${ myName }!`;
}

var msg = greeting("Kyle");

console.log(msg); // Hello, Kyle!
```

You can only `return` a single value, but if you have more values to return, you can wrap them up into a single object/array.

Since functions are values, they can be assigned as properties on objects:

```
var whatToSay = {
  greeting() {
    console.log("Hello!");
  },
  question() {
    console.log("What's your name?");
  },
  answer() {
    console.log("My name is Kyle.");
  }
};

whatToSay.greeting();
// Hello!
```

In this snippet, references to three functions (`greeting()`, `question()`, and `answer()`) are included in the object held by `whatToSay`. Each function can be called by accessing the property to retrieve the function reference value. Compare this straightforward style of defining functions on an object to the more sophisticated `class` syntax discussed later in this chapter.

There are many varied forms that `function`s take in JS. We dig into these variations in Appendix A, "So Many Function Forms."

## Comparisons

Making decisions in programs requires comparing values to determine their identity and relationship to each other. JS has several mechanisms to enable value comparison, so let's take a closer look at them.

### Equal...ish

The most common comparison in JS programs asks the question, "Is this X value *the same as* that Y value?" What exactly does "the same as" really mean to JS, though?

For ergonomic and historical reasons, the meaning is more complicated than the obvious *exact identity* sort of matching. Sometimes an equality comparison intends *exact* matching, but other times the desired comparison is a bit broader, allowing *closely similar* or *interchangeable* matching. In other words, we must be aware of the nuanced differences between an **equality** comparison and an **equivalence** comparison.

If you've spent any time working with and reading about JS, you've certainly seen the so-called "triple-equals" `==` operator, also described as the "strict equality" operator. That seems rather straightforward, right? Surely, "strict" means strict, as in narrow and *exact*.

Not *exactly*.

Yes, most values participating in an `==` equality comparison will fit with that *exact same* intuition. Consider some examples:

```
3 === 3.0;           // true
"yes" === "yes";    // true
null === null;      // true
false === false;    // true

42 === "42";         // false
"hello" === "Hello"; // false
true === 1;          // false
0 === null;          // false
"" === null;         // false
null === undefined; // false
```

#### NOTE:

Another way `==`'s equality comparison is often described is, "checking both the value and the type". In several of the examples we've looked at so far, like `42 === "42"`, the *type* of both values (number, string, etc.) does seem to be the distinguishing factor. There's more to it than that, though. All value comparisons in JS consider the type of the values being compared, not just the `==` operator. Specifically, `==` disallows any sort of type conversion (aka, "coercion") in its comparison, where other JS comparisons *do* allow coercion.

But the `==` operator does have some nuance to it, a fact many JS developers gloss over, to their detriment. The `==` operator is designed to *lie* in two cases of special values: `NaN` and `-0`. Consider:

```
NaN === NaN;        // false
0 === -0;           // true
```

In the case of `Nan`, the `==` operator *lies* and says that an occurrence of `Nan` is not equal to another `Nan`. In the case of `-0` (yes, this is a real, distinct value you can use intentionally in your programs!), the `==` operator *lies* and says it's equal to the regular `0` value.

Since the *lying* about such comparisons can be bothersome, it's best to avoid using `==` for them. For `Nan` comparisons, use the `Number.isNaN(...)` utility, which does not *lie*. For `-0` comparison, use the `Object.is(...)` utility, which also does not *lie*. `Object.is(...)` can also be used for non-*lying* `Nan` checks, if you prefer. Humorously, you could think of `Object.is(...)` as the "quadruple-equals" `=====`, the really-really-strict comparison!

There are deeper historical and technical reasons for these *lies*, but that doesn't change the fact that `==` is not actually *strictly equal* comparison, in the *strictest* sense.

The story gets even more complicated when we consider comparisons of object values (non-primitives). Consider:

```
[ 1, 2, 3 ] === [ 1, 2, 3 ];      // false
{ a: 42 } === { a: 42 };          // false
(x => x * 2) === (x => x * 2)   // false
```

What's going on here?

It may seem reasonable to assume that an equality check considers the *nature* or *contents* of the value; after all, `42 === 42` considers the actual `42` value and compares it. But when it comes to objects, a content-aware comparison is generally referred to as "structural equality."

JS does not define `==` as *structural equality* for object values. Instead, `==` uses *identity equality* for object values.

In JS, all object values are held by reference (see "Values vs References" in Appendix A), are assigned and passed by reference-copy, **and** to our current discussion, are compared by reference (identity) equality. Consider:

```
var x = [ 1, 2, 3 ];

// assignment is by reference-copy, so
// y references the *same* array as x,
// not another copy of it.
var y = x;

y === x;                  // true
y === [ 1, 2, 3 ];        // false
x === [ 1, 2, 3 ];        // false
```

In this snippet, `y === x` is true because both variables hold a reference to the same initial array. But the `== [1,2,3]` comparisons both fail because `y` and `x`, respectively, are being compared to new *different* arrays `[1,2,3]`. The array structure and contents don't matter in this comparison, only the **reference identity**.

JS does not provide a mechanism for structural equality comparison of object values, only reference identity comparison. To do structural equality comparison, you'll need to implement the checks yourself.

But beware, it's more complicated than you'll assume. For example, how might you determine if two function references are "structurally equivalent"? Even stringifying to compare their source code text wouldn't take into account things like closure. JS doesn't provide structural equality comparison because it's almost intractable to handle all the corner cases!

## Coercive Comparisons

Coercion means a value of one type being converted to its respective representation in another type (to whatever extent possible). As we'll discuss in Chapter 4, coercion is a core pillar of the JS language, not some optional feature that can reasonably be avoided.

But where coercion meets comparison operators (like equality), confusion and frustration unfortunately crop up more often than not.

Few JS features draw more ire in the broader JS community than the `==` operator, generally referred to as the "loose equality" operator. The majority of all writing and public discourse on JS condemns this operator as poorly designed and dangerous/bug-ridden when used in JS programs. Even the creator of the language himself, Brendan Eich, has lamented how it was designed as a big mistake.

From what I can tell, most of this frustration comes from a pretty short list of confusing corner cases, but a deeper problem is the extremely widespread misconception that it performs its comparisons without considering the types of its compared values.

The `==` operator performs an equality comparison similarly to how the `==` performs it. In fact, both operators consider the type of the values being compared. And if the comparison is between the same value type, both `==` and `==` **do exactly the same thing, no difference whatsoever.**

If the value types being compared are different, the `==` differs from `==` in that it allows coercion before the comparison. In other words, they both want to compare values of like types, but `==` allows type conversions *first*, and once the types have been converted to be the same on both sides, then `==` does the same thing as `==`. Instead of "loose equality," the `==` operator should be described as "coercive equality."

Consider:

```
42 == "42";           // true
1 == true;            // true
```

In both comparisons, the value types are different, so the `==` causes the non-number values (`"42"` and `true`) to be converted to numbers (`42` and `1`, respectively) before the comparisons are made.

Just being aware of this nature of `==`—that it prefers primitive numeric comparisons—helps you avoid most of the troublesome corner cases, such as staying away from a gotchas like `"" == 0` or `0 == false`.

You may be thinking, "Oh, well, I will just always avoid any coercive equality comparison (using `==` instead) to avoid those corner cases!" Eh, sorry, that's not quite as likely as you would hope.

There's a pretty good chance that you'll use relational comparison operators like `<`, `>` (and even `<=` and `>=`).

Just like `==`, these operators will perform as if they're "strict" if the types being relationally compared already match, but they'll allow coercion first (generally, to numbers) if the types differ.

Consider:

```
var arr = [ "1", "10", "100", "1000" ];
for (let i = 0; i < arr.length && arr[i] < 500; i++) {
  // will run 3 times
}
```

The `i < arr.length` comparison is "safe" from coercion because `i` and `arr.length` are always numbers. The `arr[i] < 500` invokes coercion, though, because the `arr[i]` values are all strings. Those comparisons thus become `1 < 500`, `10 < 500`, `100 < 500`, and `1000 < 500`. Since that fourth one is false, the loop stops after its third iteration.

These relational operators typically use numeric comparisons, except in the case where **both** values being compared are already strings; in this case, they use alphabetical (dictionary-like) comparison of the strings:

```
var x = "10";
var y = "9";

x < y;      // true, watch out!
```

There's no way to get these relational operators to avoid coercion, other than to just never use mismatched types in the comparisons. That's perhaps admirable as a goal, but it's still pretty likely you're going to run into a case where the types *may* differ.

The wiser approach is not to avoid coercive comparisons, but to embrace and learn their ins and outs.

Coercive comparisons crop up in other places in JS, such as conditionals (`if`, etc.), which we'll revisit in Appendix A, "Coercive Conditional Comparison."

## How We Organize in JS

Two major patterns for organizing code (data and behavior) are used broadly across the JS ecosystem: classes and modules. These patterns are not mutually exclusive; many programs can and do use both. Other programs will stick with just one pattern, or even neither!

In some respects, these patterns are very different. But interestingly, in other ways, they're just different sides of the same coin. Being proficient in JS requires understanding both patterns and where they are appropriate (and not!).

### Classes

The terms "object-oriented," "class-oriented," and "classes" are all very loaded full of detail and nuance; they're not universal in definition.

We will use a common and somewhat traditional definition here, the one most likely familiar to those with backgrounds in "object-oriented" languages like C++ and Java.

A class in a program is a definition of a "type" of custom data structure that includes both data and behaviors that operate on that data. Classes define how such a data structure works, but classes are not themselves concrete values. To get a concrete value that you can use in the program, a class must be *instantiated* (with the `new` keyword) one or more times.

Consider:

```
class Page {
  constructor(text) {
    this.text = text;
  }

  print() {
    console.log(this.text);
  }
}

class Notebook {
  constructor() {
    this.pages = [];
  }

  addPage(text) {
    var page = new Page(text);
    this.pages.push(page);
  }

  print() {
    for (let page of this.pages) {
      page.print();
    }
  }
}

var mathNotes = new Notebook();
mathNotes.addPage("Arithmetic: + - * / ...");
```

```
mathNotes.addPage("Trigonometry: sin cos tan ...");

mathNotes.print();
// ..
```

In the `Page` class, the data is a string of text stored in a `this.text` member property. The behavior is `print()`, a method that dumps the text to the console.

For the `Notebook` class, the data is an array of `Page` instances. The behavior is `addPage(..)`, a method that instantiates new `Page` pages and adds them to the list, as well as `print()` (which prints out all the pages in the notebook).

The statement `mathNotes = new Notebook()` creates an instance of the `Notebook` class, and `page = new Page(text)` is where instances of the `Page` class are created.

Behavior (methods) can only be called on instances (not the classes themselves), such as `mathNotes.addPage(..)` and `page.print()`.

The `class` mechanism allows packaging data (`text` and `pages`) to be organized together with their behaviors (e.g., `addPage(..)` and `print()`). The same program could have been built without any `class` definitions, but it would likely have been much less organized, harder to read and reason about, and more susceptible to bugs and subpar maintenance.

## Class Inheritance

Another aspect inherent to traditional "class-oriented" design, though a bit less commonly used in JS, is "inheritance" (and "polymorphism"). Consider:

```
class Publication {
    constructor(title, author, pubDate) {
        this.title = title;
        this.author = author;
        this.pubDate = pubDate;
    }

    print() {
        console.log(`Title: ${this.title}
        By: ${this.author}
        ${this.pubDate}`);
    }
}
```

This `Publication` class defines a set of common behavior that any publication might need.

Now let's consider more specific types of publication, like `Book` and `BlogPost`:

```
class Book extends Publication {
    constructor(bookDetails) {
        super(
            bookDetails.title,
            bookDetails.author,
            bookDetails.publishedon
        );
        this.publisher = bookDetails.publisher;
        this.ISBN = bookDetails.ISBN;
    }

    print() {
        super.print();
        console.log(`Publisher: ${this.publisher}
        ISBN: ${this.ISBN}`);
    }
}
```

```

        `);
    }

    class BlogPost extends Publication {
        constructor(title,author,pubDate,URL) {
            super(title,author,pubDate);
            this.URL = URL;
        }

        print() {
            super.print();
            console.log(this.URL);
        }
    }
}

```

Both `Book` and `BlogPost` use the `extends` clause to *extend* the general definition of `Publication` to include additional behavior. The `super(..)` call in each constructor delegates to the parent `Publication` class's constructor for its initialization work, and then they do more specific things according to their respective publication type (aka, "sub-class" or "child class").

Now consider using these child classes:

```

var YDKJS = new Book({
    title: "You Don't Know JS",
    author: "Kyle Simpson",
    publishedOn: "June 2014",
    publisher: "O'Reilly",
    ISBN: "123456-789"
});

YDKJS.print();
// Title: You Don't Know JS
// By: Kyle Simpson
// June 2014
// Publisher: O'Reilly
// ISBN: 123456-789

var forAgainstLet = new BlogPost(
    "For and against let",
    "Kyle Simpson",
    "October 27, 2014",
    "https://davidwalsh.name/for-and-against-let"
);

forAgainstLet.print();
// Title: For and against let
// By: Kyle Simpson
// October 27, 2014
// https://davidwalsh.name/for-and-against-let

```

Notice that both child class instances have a `print()` method, which was an override of the *inherited* `print()` method from the parent `Publication` class. Each of those overridden child class `print()` methods call `super.print()` to invoke the inherited version of the `print()` method.

The fact that both the inherited and overridden methods can have the same name and co-exist is called *polymorphism*.

Inheritance is a powerful tool for organizing data/behavior in separate logical units (classes), but allowing the child class to cooperate with the parent by accessing/using its behavior and data.

## Modules

The module pattern has essentially the same goal as the class pattern, which is to group data and behavior together into logical units. Also like classes, modules can "include" or "access" the data and behaviors of other modules, for cooperation sake.

But modules have some important differences from classes. Most notably, the syntax is entirely different.

## Classic Modules

ES6 added a module syntax form to native JS syntax, which we'll look at in a moment. But from the early days of JS, modules was an important and common pattern that was leveraged in countless JS programs, even without a dedicated syntax.

The key hallmarks of a *classic module* are an outer function (that runs at least once), which returns an "instance" of the module with one or more functions exposed that can operate on the module instance's internal (hidden) data.

Because a module of this form is *just a function*, and calling it produces an "instance" of the module, another description for these functions is "module factories".

Consider the classic module form of the earlier `Publication`, `Book`, and `BlogPost` classes:

```
function Publication(title,author,pubDate) {
  var publicAPI = {
    print() {
      console.log(`Title: ${ title }
      By: ${ author }
      ${ pubDate }`);
    }
  };
  return publicAPI;
}

function Book(bookDetails) {
  var pub = Publication(
    bookDetails.title,
    bookDetails.author,
    bookDetails.publishedOn
  );
  var publicAPI = {
    print() {
      pub.print();
      console.log(`Publisher: ${ bookDetails.publisher }
      ISBN: ${ bookDetails.ISBN }`);
    }
  };
  return publicAPI;
}

function BlogPost(title,author,pubDate,URL) {
  var pub = Publication(title,author,pubDate);

  var publicAPI = {
    print() {
      pub.print();
      console.log(URL);
    }
  };
  return publicAPI;
}
```

Comparing these forms to the `class` forms, there are more similarities than differences.

The `class` form stores methods and data on an object instance, which must be accessed with the `this.` prefix. With modules, the methods and data are accessed as identifier variables in scope, without any `this.` prefix.

With `class`, the "API" of an instance is implicit in the class definition—also, all data and methods are public. With the module factory function, you explicitly create and return an object with any publicly exposed methods, and any data or other unreferenced methods remain private inside the factory function.

There are other variations to this factory function form that are quite common across JS, even in 2020; you may run across these forms in different JS programs: AMD (Asynchronous Module Definition), UMD (Universal Module Definition), and CommonJS (classic Node.js-style modules). The variations are minor (not quite compatible). However, all of these forms rely on the same basic principles.

Consider also the usage (aka, "instantiation") of these module factory functions:

```
var YDKJS = Book({
  title: "You Don't Know JS",
  author: "Kyle Simpson",
  publishedOn: "June 2014",
  publisher: "O'Reilly",
  ISBN: "123456-789"
});

YDKJS.print();
// Title: You Don't Know JS
// By: Kyle Simpson
// June 2014
// Publisher: O'Reilly
// ISBN: 123456-789

var forAgainstLet = BlogPost(
  "For and against let",
  "Kyle Simpson",
  "October 27, 2014",
  "https://davidwalsh.name/for-and-against-let"
);

forAgainstLet.print();
// Title: For and against let
// By: Kyle Simpson
// October 27, 2014
// https://davidwalsh.name/for-and-against-let
```

The only observable difference here is the lack of using `new`, calling the module factories as normal functions.

## ES Modules

ES modules (ESM), introduced to the JS language in ES6, are meant to serve much the same spirit and purpose as the existing *classic modules* just described, especially taking into account important variations and use cases from AMD, UMD, and CommonJS.

The implementation approach does, however, differ significantly.

First, there's no wrapping function to *define* a module. The wrapping context is a file. ESMs are always file-based; one file, one module.

Second, you don't interact with a module's "API" explicitly, but rather use the `export` keyword to add a variable or method to its public API definition. If something is defined in a module but not `export`ed, then it stays hidden (just as with *classic modules*).

Third, and maybe most noticeably different from previously discussed patterns, you don't "instantiate" an ES module, you just `import` it to use its single instance. ESMs are, in effect, "singletons," in that there's only one instance ever created, at first `import` in your program, and all other `import`s just receive a reference to that same single instance. If your module needs to

support multiple instantiations, you have to provide a *classic module-style* factory function on your ESM definition for that purpose.

In our running example, we do assume multiple-instantiation, so these following snippets will mix both ESM and *classic modules*.

Consider the file `publication.js` :

```
function printDetails(title,author,pubDate) {
    console.log(`Title: ${ title }
    By: ${ author }
    ${ pubDate }`);
}

export function create(title,author,pubDate) {
    var publicAPI = {
        print() {
            printDetails(title,author,pubDate);
        }
    };

    return publicAPI;
}
```

To import and use this module, from another ES module like `blogpost.js` :

```
import { create as createPub } from "publication.js";

function printDetails(pub,URL) {
    pub.print();
    console.log(URL);
}

export function create(title,author,pubDate,URL) {
    var pub = createPub(title,author,pubDate);

    var publicAPI = {
        print() {
            printDetails(pub,URL);
        }
    };

    return publicAPI;
}
```

And finally, to use this module, we import into another ES module like `main.js` :

```
import { create as newBlogPost } from "blogpost.js";

var forAgainstLet = newBlogPost(
    "For and against let",
    "Kyle Simpson",
    "October 27, 2014",
    "https://davidwalsh.name/for-and-against-let"
);

forAgainstLet.print();
// Title: For and against let
// By: Kyle Simpson
// October 27, 2014
// https://davidwalsh.name/for-and-against-let
```

**NOTE:**

The `as newBlogPost` clause in the `import` statement is optional; if omitted, a top-level function just named `create(..)` would be imported. In this case, I'm renaming it for readability sake; its more generic factory name of `create(..)` becomes more semantically descriptive of its purpose as `newBlogPost(..)`.

As shown, ES modules can use *classic modules* internally if they need to support multiple-instantiation. Alternatively, we could have exposed a `class` from our module instead of a `create(..)` factory function, with generally the same outcome. However, since you're already using ESM at that point, I'd recommend sticking with *classic modules* instead of `class`.

If your module only needs a single instance, you can skip the extra layers of complexity: `export` its public methods directly.

## The Rabbit Hole Deepens

As promised at the top of this chapter, we just glanced over a wide surface area of the main parts of the JS language. Your head may still be spinning, but that's entirely natural after such a firehose of information!

Even with just this "brief" survey of JS, we covered or hinted at a ton of details you should carefully consider and ensure you are comfortable with. I'm serious when I suggest: re-read this chapter, maybe several times.

In the next chapter, we're going to dig much deeper into some important aspects of how JS works at its core. But before you follow that rabbit hole deeper, make sure you've taken adequate time to fully digest what we've just covered here.

# You Don't Know JS Yet: Get Started - 2nd Edition

## Chapter 3: Digging to the Roots of JS

If you've read Chapters 1 and 2, and taken the time to digest and percolate, you're hopefully starting to *get JS* a little more. If you skipped/skimmed them (especially Chapter 2), I recommend going back to spend some more time with that material.

In Chapter 2, we surveyed syntax, patterns, and behaviors at a high level. In this chapter, our attention shifts to some of the lower-level root characteristics of JS that underpin virtually every line of code we write.

Be aware: this chapter digs much deeper than you're likely used to thinking about a programming language. My goal is to help you appreciate the core of how JS works, what makes it tick. This chapter should begin to answer some of the "Why?" questions that may be cropping up as you explore JS. However, this material is still not an exhaustive exposition of the language; that's what the rest of the book series is for! Our goal here is still just to *get started*, and become more comfortable with, the *feel* of JS, how it ebbs and flows.

Don't run so quickly through this material that you get lost in the weeds. As I've said a dozen times already, **take your time**. Even still, you'll probably finish this chapter with remaining questions. That's OK, because there's a whole book series ahead of you to keep exploring!

## Iteration

Since programs are essentially built to process data (and make decisions on that data), the patterns used to step through the data have a big impact on the program's readability.

The iterator pattern has been around for decades, and suggests a "standardized" approach to consuming data from a source one *chunk* at a time. The idea is that it's more common and helpful to iterate the data source—to progressively handle the collection of data by processing the first part, then the next, and so on, rather than handling the entire set all at once.

Imagine a data structure that represents a relational database `SELECT` query, which typically organizes the results as rows. If this query had only one or a couple of rows, you could handle the entire result set at once, and assign each row to a local variable, and perform whatever operations on that data that were appropriate.

But if the query has 100 or 1,000 (or more!) rows, you'll need iterative processing to deal with this data (typically, a loop).

The iterator pattern defines a data structure called an "iterator" that has a reference to an underlying data source (like the query result rows), which exposes a method like `next()`. Calling `next()` returns the next piece of data (i.e., a "record" or "row" from a database query).

You don't always know how many pieces of data that you will need to iterate through, so the pattern typically indicates completion by some special value or exception once you iterate through the entire set and *go past the end*.

The importance of the iterator pattern is in adhering to a *standard* way of processing data iteratively, which creates cleaner and easier to understand code, as opposed to having every data structure/source define its own custom way of handling its data.

After many years of various JS community efforts around mutually agreed-upon iteration techniques, ES6 standardized a specific protocol for the iterator pattern directly in the language. The protocol defines a `next()` method whose return is an object called an *iterator result*; the object has `value` and `done` properties, where `done` is a boolean that is `false` until the iteration over the underlying data source is complete.

## Consuming Iterators

With the ES6 iteration protocol in place, it's workable to consume a data source one value at a time, checking after each `next()` call for `done` to be `true` to stop the iteration. But this approach is rather manual, so ES6 also included several mechanisms (syntax and APIs) for standardized consumption of these iterators.

One such mechanism is the `for...of` loop:

```
// given an iterator of some data source:  
var it = /* ... */;  
  
// loop over its results one at a time  
for (let val of it) {  
    console.log(`Iterator value: ${val}`);  
}  
// Iterator value: ..  
// Iterator value: ..  
// ..
```

#### NOTE:

We'll omit the manual loop equivalent here, but it's definitely less readable than the `for...of` loop!

Another mechanism that's often used for consuming iterators is the `...` operator. This operator actually has two symmetrical forms: *spread* and *rest* (or *gather*, as I prefer). The *spread* form is an iterator-consumer.

To *spread* an iterator, you have to have *something* to spread it into. There are two possibilities in JS: an array or an argument list for a function call.

An array spread:

```
// spread an iterator into an array,  
// with each iterated value occupying  
// an array element position.  
var vals = [ ...it ];
```

A function call spread:

```
// spread an iterator into a function,  
// call with each iterated value  
// occupying an argument position.  
doSomethingUseful( ...it );
```

In both cases, the iterator-spread form of `...` follows the iterator-consumption protocol (the same as the `for...of` loop) to retrieve all available values from an iterator and place (aka, spread) them into the receiving context (array, argument list).

## Iterables

The iterator-consumption protocol is technically defined for consuming *iterables*; an iterable is a value that can be iterated over.

The protocol automatically creates an iterator instance from an iterable, and consumes *just that iterator instance* to its completion. This means a single iterable could be consumed more than once; each time, a new iterator instance would be created and used.

So where do we find iterables?

ES6 defined the basic data structure/collection types in JS as iterables. This includes strings, arrays, maps, sets, and others.

Consider:

```
// an array is an iterable  
var arr = [ 10, 20, 30 ];
```

```

for (let val of arr) {
    console.log(`Array value: ${ val }`);
}
// Array value: 10
// Array value: 20
// Array value: 30

```

Since arrays are iterables, we can shallow-copy an array using iterator consumption via the `... spread operator`:

```
var arrCopy = [ ...arr ];
```

We can also iterate the characters in a string one at a time:

```

var greeting = "Hello world!";
var chars = [ ...greeting ];

chars;
// [ "H", "e", "l", "l", "o", " ", 
//   "w", "o", "r", "l", "d", "!" ]

```

A `Map` data structure uses objects as keys, associating a value (of any type) with that object. Maps have a different default iteration than seen here, in that the iteration is not just over the map's values but instead its *entries*. An *entry* is a tuple (2-element array) including both a key and a value.

Consider:

```

// given two DOM elements, `btn1` and `btn2`

var buttonNames = new Map();
buttonNames.set(btn1,"Button 1");
buttonNames.set(btn2,"Button 2");

for (let [btn,btnName] of buttonNames) {
    btn.addEventListener("click",function onClick(){
        console.log(`Clicked ${ btnName }`);
    });
}

```

In the `for..of` loop over the default map iteration, we use the `[btn,btnName]` syntax (called "array destructuring") to break down each consumed tuple into the respective key/value pairs (`btn1 / "Button 1"` and `btn2 / "Button 2"`).

Each of the built-in iterables in JS expose a default iteration, one which likely matches your intuition. But you can also choose a more specific iteration if necessary. For example, if we want to consume only the values of the above `buttonNames` map, we can call `values()` to get a values-only iterator:

```

for (let btnName of buttonNames.values()) {
    console.log(btnName);
}
// Button 1
// Button 2

```

Or if we want the index *and* value in an array iteration, we can make an entries iterator with the `entries()` method:

```

var arr = [ 10, 20, 30 ];

for (let [idx,val] of arr.entries()) {
    console.log(`[${ idx }]: ${ val }`);
}

```

```
// [0]: 10
// [1]: 20
// [2]: 30
```

For the most part, all built-in iterables in JS have three iterator forms available: keys-only (`keys()`), values-only (`values()`), and entries (`entries()`).

Beyond just using built-in iterables, you can also ensure your own data structures adhere to the iteration protocol; doing so means you opt into the ability to consume your data with `for..of` loops and the `...` operator. "Standardizing" on this protocol means code that is overall more readily recognizable and readable.

#### NOTE:

You may have noticed a nuanced shift that occurred in this discussion. We started by talking about consuming **iterators**, but then switched to talking about iterating over **iterables**. The iteration-consumption protocol expects an *iterable*, but the reason we can provide a direct *iterator* is that an iterator is just an iterable of itself! When creating an iterator instance from an existing iterator, the iterator itself is returned.

## Closure

Perhaps without realizing it, almost every JS developer has made use of closure. In fact, closure is one of the most pervasive programming functionalities across a majority of languages. It might even be as important to understand as variables or loops; that's how fundamental it is.

Yet it feels kind of hidden, almost magical. And it's often talked about in either very abstract or very informal terms, which does little to help us nail down exactly what it is.

We need to be able to recognize where closure is used in programs, as the presence or lack of closure is sometimes the cause of bugs (or even the cause of performance issues).

So let's define closure in a pragmatic and concrete way:

Closure is when a function remembers and continues to access variables from outside its scope, even when the function is executed in a different scope.

We see two definitional characteristics here. First, closure is part of the nature of a function. Objects don't get closures, functions do. Second, to observe a closure, you must execute a function in a different scope than where that function was originally defined.

Consider:

```
function greeting(msg) {
  return function who(name) {
    console.log(`#${ msg }, ${ name }!`);
  };
}

var hello = greeting("Hello");
var howdy = greeting("Howdy");

hello("Kyle");
// Hello, Kyle!

hello("Sarah");
// Hello, Sarah!

howdy("Grant");
// Howdy, Grant!
```

First, the `greeting(..)` outer function is executed, creating an instance of the inner function `who(..)`; that function closes over the variable `msg`, which is the parameter from the outer scope of `greeting(..)`. When that inner function is returned, its reference is assigned to the `hello` variable in the outer scope. Then we call `greeting(..)` a second time, creating a new inner function instance, with a new closure over a new `msg`, and return that reference to be assigned to `howdy`.

When the `greeting(..)` function finishes running, normally we would expect all of its variables to be garbage collected (removed from memory). We'd expect each `msg` to go away, but they don't. The reason is closure. Since the inner function instances are still alive (assigned to `hello` and `howdy`, respectively), their closures are still preserving the `msg` variables.

These closures are not a snapshot of the `msg` variable's value; they are a direct link and preservation of the variable itself. That means closure can actually observe (or make!) updates to these variables over time.

```
function counter(step = 1) {
  var count = 0;
  return function increaseCount(){
    count = count + step;
    return count;
  };
}

var incBy1 = counter(1);
var incBy3 = counter(3);

incBy1();      // 1
incBy1();      // 2

incBy3();      // 3
incBy3();      // 6
incBy3();      // 9
```

Each instance of the inner `increaseCount()` function is closed over both the `count` and `step` variables from its outer `counter(..)` function's scope. `step` remains the same over time, but `count` is updated on each invocation of that inner function. Since closure is over the variables and not just snapshots of the values, these updates are preserved.

Closure is most common when working with asynchronous code, such as with callbacks. Consider:

```
function getSomeData(url) {
  ajax(url,function onResponse(resp){
    console.log(
      `Response (from ${url}): ${resp}`
    );
  });
}

getSomeData("https://some.url/wherever");
// Response (from https://some.url/wherever): ...
```

The inner function `onResponse(..)` is closed over `url`, and thus preserves and remembers it until the Ajax call returns and executes `onResponse(..)`. Even though `getSomeData(..)` finishes right away, the `url` parameter variable is kept alive in the closure for as long as needed.

It's not necessary that the outer scope be a function—it usually is, but not always—just that there be at least one variable in an outer scope accessed from an inner function:

```
for (let [idx,btn] of buttons.entries()) {
  btn.addEventListener("click",function onClick(){
    console.log(`Clicked on button ${idx}!`);
  });
}
```

Because this loop is using `let` declarations, each iteration gets new block-scoped (aka, local) `idx` and `btn` variables; the loop also creates a new inner `onclick(..)` function each time. That inner function closes over `idx`, preserving it for as long as the click handler is set on the `btn`. So when each button is clicked, its handler can print its associated index value, because the handler remembers its respective `idx` variable.

Remember: this closure is not over the value (like `1` or `3`), but over the variable `idx` itself.

Closure is one of the most prevalent and important programming patterns in any language. But that's especially true of JS; it's hard to imagine doing anything useful without leveraging closure in one way or another.

If you're still feeling unclear or shaky about closure, the majority of Book 2, *Scope & Closures* is focused on the topic.

## this Keyword

One of JS's most powerful mechanisms is also one of its most misunderstood: the `this` keyword. One common misconception is that a function's `this` refers to the function itself. Because of how `this` works in other languages, another misconception is that `this` points the instance that a method belongs to. Both are incorrect.

As discussed previously, when a function is defined, it is *attached* to its enclosing scope via closure. Scope is the set of rules that controls how references to variables are resolved.

But functions also have another characteristic besides their scope that influences what they can access. This characteristic is best described as an *execution context*, and it's exposed to the function via its `this` keyword.

Scope is static and contains a fixed set of variables available at the moment and location you define a function, but a function's execution context is dynamic, entirely dependent on **how it is called** (regardless of where it is defined or even called from).

`this` is not a fixed characteristic of a function based on the function's definition, but rather a dynamic characteristic that's determined each time the function is called.

One way to think about the *execution context* is that it's a tangible object whose properties are made available to a function while it executes. Compare that to scope, which can also be thought of as an *object*; except, the *scope object* is hidden inside the JS engine, it's always the same for that function, and its *properties* take the form of identifier variables available inside the function.

```
function classroom(teacher) {
  return function study() {
    console.log(
      `${teacher} says to study ${this.topic}`
    );
  };
}
var assignment = classroom("Kyle");
```

The outer `classroom(..)` function makes no reference to a `this` keyword, so it's just like any other function we've seen so far. But the inner `study()` function does reference `this`, which makes it a `this`-aware function. In other words, it's a function that is dependent on its *execution context*.

### NOTE:

`study()` is also closed over the `teacher` variable from its outer scope.

The inner `study()` function returned by `classroom("Kyle")` is assigned to a variable called `assignment`. So how can `assignment()` (aka `study()`) be called?

```
assignment();
// Kyle says to study undefined -- Oops :(
```

In this snippet, we call `assignment()` as a plain, normal function, without providing it any *execution context*.

Since this program is not in strict mode (see Chapter 1, "Strictly Speaking"), context-aware functions that are called **without any context specified** default the context to the global object (`window` in the browser). As there is no global variable named `topic` (and thus no such property on the global object), `this.topic` resolves to `undefined`.

Now consider:

```
var homework = {
  topic: "JS",
  assignment: assignment
};

homework.assignment();
// Kyle says to study JS
```

A copy of the `assignment` function reference is set as a property on the `homework` object, and then it's called as `homework.assignment()`. That means the `this` for that function call will be the `homework` object. Hence, `this.topic` resolves to `"JS"`.

Lastly:

```
var otherHomework = {
  topic: "Math"
};

assignment.call(otherHomework);
// Kyle says to study Math
```

A third way to invoke a function is with the `call(..)` method, which takes an object (`otherHomework` here) to use for setting the `this` reference for the function call. The property reference `this.topic` resolves to `"Math"`.

The same context-aware function invoked three different ways, gives different answers each time for what object `this` will reference.

The benefit of `this`-aware functions—and their dynamic context—is the ability to more flexibly re-use a single function with data from different objects. A function that closes over a scope can never reference a different scope or set of variables. But a function that has dynamic `this` context awareness can be quite helpful for certain tasks.

## Prototypes

Where `this` is a characteristic of function execution, a prototype is a characteristic of an object, and specifically resolution of a property access.

Think about a prototype as a linkage between two objects; the linkage is hidden behind the scenes, though there are ways to expose and observe it. This prototype linkage occurs when an object is created; it's linked to another object that already exists.

A series of objects linked together via prototypes is called the "prototype chain."

The purpose of this prototype linkage (i.e., from an object B to another object A) is so that accesses against B for properties/methods that B does not have, are *delegated* to A to handle. Delegation of property/method access allows two (or more!) objects to cooperate with each other to perform a task.

Consider defining an object as a normal literal:

```
var homework = {
  topic: "JS"
};
```

The `homework` object only has a single property on it: `topic`. However, its default prototype linkage connects to the `Object.prototype` object, which has common built-in methods on it like `toString()` and `valueof()`, among others.

We can observe this prototype linkage *delegation* from `homework` to `Object.prototype`:

```
homework.toString(); // [object Object]
```

`homework.toString()` works even though `homework` doesn't have a `toString()` method defined; the delegation invokes `Object.prototype.toString()` instead.

## Object Linkage

To define an object prototype linkage, you can create the object using the `Object.create(..)` utility:

```
var homework = {
  topic: "JS"
};

var otherHomework = Object.create(homework);

otherHomework.topic; // "JS"
```

The first argument to `Object.create(..)` specifies an object to link the newly created object to, and then returns the newly created (and linked!) object.

Figure 4 shows how the three objects (`otherHomework`, `homework`, and `Object.prototype`) are linked in a prototype chain:

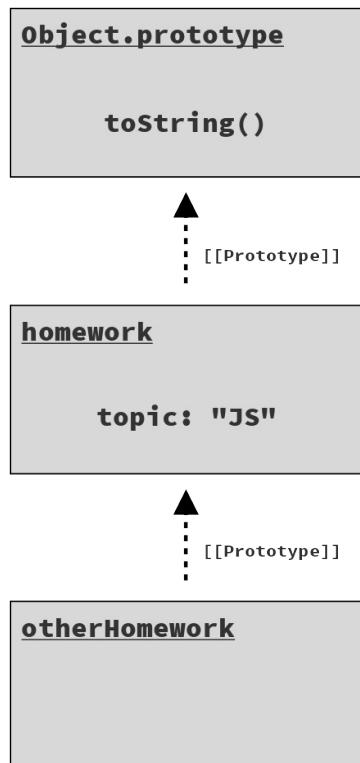


Fig. 4: Objects in a prototype chain

Delegation through the prototype chain only applies for accesses to lookup the value in a property. If you assign to a property of an object, that will apply directly to the object regardless of where that object is prototype linked to.

**TIP:**

`Object.create(null)` creates an object that is not prototype linked anywhere, so it's purely just a standalone object; in some circumstances, that may be preferable.

Consider:

```
homework.topic;
// "JS"

otherHomework.topic;
// "JS"

otherHomework.topic = "Math";
otherHomework.topic;
// "Math"

homework.topic;
// "JS" -- not "Math"
```

The assignment to `topic` creates a property of that name directly on `otherHomework`; there's no effect on the `topic` property on `homework`. The next statement then accesses `otherHomework.topic`, and we see the non-delegated answer from that new property: `"Math"`.

Figure 5 shows the objects/properties after the assignment that creates the `otherHomework.topic` property:

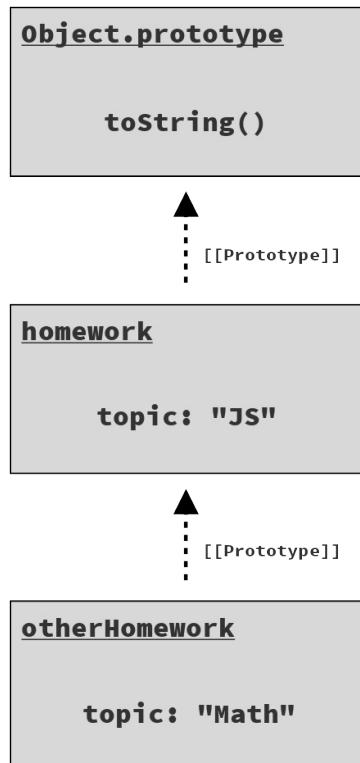


Fig. 5: Shadowed property 'topic'

The `topic` on `otherHomework` is "shadowing" the property of the same name on the `homework` object in the chain.

**NOTE:**

Another frankly more convoluted but perhaps still more common way of creating an object with a prototype linkage is using the "prototypal class" pattern, from before `class` (see Chapter 2, "Classes") was added in ES6. We'll cover this topic in more detail in Appendix A, "Prototypal 'Classes'".

## this Revisited

We covered the `this` keyword earlier, but its true importance shines when considering how it powers prototype-delegated function calls. Indeed, one of the main reasons `this` supports dynamic context based on how the function is called is so that method calls on objects which delegate through the prototype chain still maintain the expected `this`.

Consider:

```
var homework = {
  study() {
    console.log(`Please study ${ this.topic}`);
  }
};

var jsHomework = Object.create(homework);
jsHomework.topic = "JS";
jsHomework.study();
// Please study JS

var mathHomework = Object.create(homework);
mathHomework.topic = "Math";
mathHomework.study();
// Please study Math
```

The two objects `jsHomework` and `mathHomework` each prototype link to the single `homework` object, which has the `study()` function. `jsHomework` and `mathHomework` are each given their own `topic` property (see Figure 6).

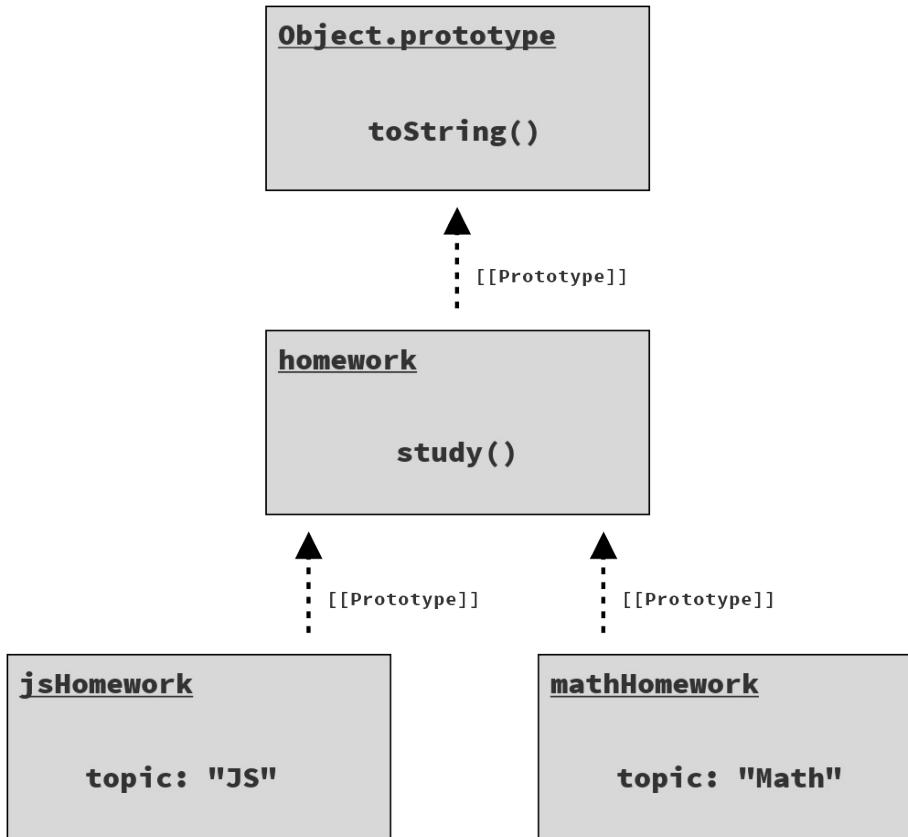


Fig. 6: Two objects linked to a common parent

`jsHomework.study()` delegates to `homework.study()`, but its `this` (`this.topic`) for that execution resolves to `jsHomework` because of how the function is called, so `this.topic` is "JS". Similarly for `mathHomework.study()` delegating to `homework.study()` but still resolving `this` to `mathHomework`, and thus `this.topic` as "Math".

The preceding code snippet would be far less useful if `this` was resolved to `homework`. Yet, in many other languages, it would seem `this` would be `homework` because the `study()` method is indeed defined on `homework`.

Unlike many other languages, JS's `this` being dynamic is a critical component of allowing prototype delegation, and indeed `class`, to work as expected!

## Asking "Why?"

The intended take-away from this chapter is that there's a lot more to JS under the hood than is obvious from glancing at the surface.

As you are *getting started* learning and knowing JS more closely, one of the most important skills you can practice and bolster is curiosity, and the art of asking "Why?" when you encounter something in the language.

Even though this chapter has gone quite deep on some of the topics, many details have still been entirely skimmed over. There's much more to learn here, and the path to that starts with you asking the *right* questions of your code. Asking the right questions is a critical skill of becoming a better developer.

In the final chapter of this book, we're going to briefly look at how JS is divided, as covered across the rest of the *You Don't Know JS Yet* book series. Also, don't skip Appendix B of this book, which has some practice code to review some of the main topics covered in this book.

# You Don't Know JS Yet: Get Started - 2nd Edition

## Chapter 4: The Bigger Picture

This book surveys what you need to be aware of as you *get started* with JS. The goal is to fill in gaps that readers newer to JS might have tripped over in their early encounters with the language. I also hope that we've hinted at enough deeper detail throughout to pique your curiosity to want to dig more into the language.

The rest of the books in this series are where we will unpack all of the rest of the language, in far greater detail than we could have done in a few brief chapters here.

Remember to take your time, though. Rather than rushing onto the next book in an attempt to churn through all the books expediently, spend some time going back over the material in this book. Spend some more time looking through code in your current projects, and comparing what you see to what's been discussed so far.

When you're ready, this final chapter divides the organization of the JS language into three main pillars, then offers a brief roadmap of what to expect from the rest of the book series, and how I suggest you proceed. Also, don't skip the appendices, especially Appendix B, "Practice, Practice, Practice!".

### Pillar 1: Scope and Closure

The organization of variables into units of scope (functions, blocks) is one of the most foundational characteristics of any language; perhaps no other characteristic has a greater impact on how programs behave.

Scopes are like buckets, and variables are like marbles you put into those buckets. The scope model of a language is like the rules that help you determine which color marbles go in which matching-color buckets.

Scopes nest inside each other, and for any given expression or statement, only variables at that level of scope nesting, or in higher/outer scopes, are accessible; variables from lower/inner scopes are hidden and inaccessible.

This is how scopes behave in most languages, which is called lexical scope. The scope unit boundaries, and how variables are organized in them, is determined at the time the program is parsed (compiled). In other words, it's an author-time decision: where you locate a function/scope in the program determines what the scope structure of that part of the program will be.

JS is lexically scoped, though many claim it isn't, because of two particular characteristics of its model that are not present in other lexically scoped languages.

The first is commonly called *hoisting*: when all variables declared anywhere in a scope are treated as if they're declared at the beginning of the scope. The other is that `var`-declared variables are function scoped, even if they appear inside a block.

Neither hoisting nor function-scoped `var` are sufficient to back the claim that JS is not lexically scoped. `let` / `const` declarations have a peculiar error behavior called the "Temporal Dead Zone" (TDZ) which results in observable but unusable variables. Though TDZ can be strange to encounter, it's *also* not an invalidation of lexical scoping. All of these are just unique parts of the language that should be learned and understood by all JS developers.

Closure is a natural result of lexical scope when the language has functions as first-class values, as JS does. When a function makes reference to variables from an outer scope, and that function is passed around as a value and executed in other scopes, it maintains access to its original scope variables; this is closure.

Across all of programming, but especially in JS, closure drives many of the most important programming patterns, including modules. As I see it, modules are as *with the grain* as you can get, when it comes to code organization in JS.

To dig further into scope, closures, and how modules work, read Book 2, *Scope & Closures*.

## Pillar 2: Prototypes

The second pillar of the language is the prototypes system. We covered this topic in-depth in Chapter 3 ("Prototypes"), but I just want to make a few more comments about its importance.

JS is one of very few languages where you have the option to create objects directly and explicitly, without first defining their structure in a class.

For many years, people implemented the class design pattern on top of prototypes—so-called "prototypal inheritance" (see Appendix A, "Prototypal 'Classes'")—and then with the advent of ES6's `class` keyword, the language doubled-down on its inclination toward OO/class-style programming.

But I think that focus has obscured the beauty and power of the prototype system: the ability for two objects to simply connect with each other and cooperate dynamically (during function/method execution) through sharing a `this` context.

Classes are just one pattern you can build on top of such power. But another approach, in a very different direction, is to simply embrace objects as objects, forget classes altogether, and let objects cooperate through the prototype chain. This is called *behavior delegation*. I think delegation is more powerful than class inheritance, as a means for organizing behavior and data in our programs.

But class inheritance gets almost all the attention. And the rest goes to functional programming (FP), as the sort of "anti-class" way of designing programs. This saddens me, because it snuffs out any chance for exploration of delegation as a viable alternative.

I encourage you to spend plenty of time deep in Book 3, *Objects & Classes*, to see how object delegation holds far more potential than we've perhaps realized. This isn't an anti- `class` message, but it is intentionally a "classes aren't the only way to use objects" message that I want more JS developers to consider.

Object delegation is, I would argue, far more *with the grain* of JS, than classes (more on *grains* in a bit).

## Pillar 3: Types and Coercion

The third pillar of JS is by far the most overlooked part of JS's nature.

The vast majority of developers have strong misconceptions about how *types* work in programming languages, and especially how they work in JS. A tidal wave of interest in the broader JS community has begun to shift to "static typing" approaches, using type-aware tooling like TypeScript or Flow.

I agree that JS developers should learn more about types, and should learn more about how JS manages type conversions. I also agree that type-aware tooling can help developers, assuming they have gained and used this knowledge in the first place!

But I don't agree at all that the inevitable conclusion of this is to decide JS's type mechanism is bad and that we need to cover up JS's types with solutions outside the language. We don't have to follow the "static typing" way to be smart and solid with types in our programs. There are other options, if you're just willing to go *against the grain* of the crowd, and *with the grain* of JS (again, more on that to come).

Arguably, this pillar is more important than the other two, in the sense that no JS program will do anything useful if it doesn't properly leverage JS's value types, as well as the conversion (coercion) of values between types.

Even if you love TypeScript/Flow, you are not going to get the most out of those tools or coding approaches if you aren't deeply familiar with how the language itself manages value types.

To learn more about JS types and coercion, check out Book 4, *Types & Grammar*. But please don't skip over this topic just because you've always heard that we should use `==` and forget about the rest.

Without learning this pillar, your foundation in JS is shaky and incomplete at best.

## With the Grain

I have some advice to share on continuing your learning journey with JS, and your path through the rest of this book series: be aware of the *grain* (recall various references to *grain* earlier in this chapter).

First, consider the *grain* (as in, wood) of how most people approach and use JS. You've probably already noticed that these books cut against that *grain* in many respects. In YDKJSY, I respect you the reader enough to explain all the parts of JS, not only some select popular parts. I believe you're both capable and deserving of that knowledge.

But that is not what you'll find from a lot of other material out there. It also means that the more you follow and adhere to the guidance from these books—that you think carefully and analyze for yourself what's best in your code—the more you will stand out. That can be a good and bad thing. If you ever want to break out from the crowd, you're going to have to break from how the crowd does it!

But I've also had many people tell me that they quoted some topic/explanation from these books during a job interview, and the interviewer told the candidate they were wrong; indeed, people have reportedly lost out on job offers as a result.

As much as possible, I endeavor in these books to provide completely accurate information about JS, informed generally from the specification itself. But I also dose out quite a bit of my opinions on how you can interpret and use JS to the best benefit in your programs. I don't present opinion as fact, or vice versa. You'll always know which is which in these books.

Facts about JS are not really up for debate. Either the specification says something, or it doesn't. If you don't like what the specification says, or my relaying of it, take that up with TC39! If you're in an interview and they claim you're wrong on the facts, ask them right then and there if you can look it up in the specification. If the interviewer won't re-consider, then you shouldn't want to work there anyway.

But if you choose to align with my opinions, you have to be prepared to back up those choices with *why* you feel that way. Don't just parrot what I say. Own your opinions. Defend them. And if someone you were hoping to work with disagrees, walk away with your head still held high. It's a big JS, and there's plenty of room for lots of different ways.

In other words, don't be afraid to go against the *grain*, as I have done with these books and all my teachings. Nobody can tell you how you will best make use of JS; that's for you to decide. I'm merely trying to empower you in coming to your own conclusions, no matter what they are.

On the other hand, there's a *grain* you really should pay attention to and follow: the *grain* of how JS works, at the language level. There are things that work well and naturally in JS, given the right practice and approach, and there are things you really shouldn't try to do in the language.

Can you make your JS program look like a Java, C#, or Perl program? What about Python or Ruby, or even PHP? To varying degrees, sure you can. But should you?

No, I don't think you should. I think you should learn and embrace the JS way, and make your JS programs as JS'y as is practical. Some will think that means sloppy and informal programming, but I don't mean that at all. I just mean that JS has a lot of patterns and idioms that are recognizably "JS," and going with that *grain* is the general path to best success.

Finally, maybe the most important *grain* to recognize is how the existing program(s) you're working on, and developers you're working with, do stuff. Don't read these books and then try to change *all that grain* in your existing projects over night. That approach will always fail.

You'll have to shift these things little by little, over time. Work on building consensus with your fellow developers on why it's important to re-visit and re-consider an approach. But do so with just one small topic at a time, and let before-and-after code comparisons do most of the talking. Bring everyone on the team together to discuss, and push for decisions that are based on analysis and evidence from the code rather than the inertia of "our senior devs have always done it this way."

That's the most important advice I can impart to help you learn JS. Always keep looking for better ways to use what JS gives us to author more readable code. Everyone who works on your code, including your future self, will thank you!

## In Order

So now you've got a broader perspective on what's left to explore in JS, and the right attitude to approach the rest of your journey.

But one of the most common practical questions I get at this point is, "What order should I read the books?" There is a straightforward answer... but it also depends.

My suggestion for most readers is to proceed through this series in this order:

1. Get started with a solid foundation of JS from *Get Started* (Book 1) -- good news, you've already almost finished this book!
2. In *Scope & Closures* (Book 2), dig into the first pillar of JS: lexical scope, how that supports closure, and how the module pattern organizes code.
3. In *Objects & Classes* (Book 3), focus on the second pillar of JS: how JS's `this` works, how object prototypes support delegation, and how prototypes enable the `class` mechanism for OO-style code organization.
4. In *Types & Grammar* (Book 4), tackle the third and final pillar of JS: types and type coercion, as well as how JS's syntax and grammar define how we write our code.
5. With the **three pillars** solidly in place, *Sync & Async* (Book 5) then explores how we use flow control to model state change in our programs, both synchronously (right away) and asynchronously (over time).
6. The series concludes with *ES.Next & Beyond* (Book 6), a forward look at the near- and mid-term future of JS, including a variety of features likely coming to your JS programs before too long.

That's the intended order to read this book series.

However, Books 2, 3, and 4 can generally be read in any order, depending on which topic you feel most curious about and comfortable exploring first. But I don't recommend you skip any of these three books—not even *Types & Grammar*, as some of you will be tempted to do!—even if you think you already have that topic down.

Book 5 (*Sync & Async*) is crucial for deeply understanding JS, but if you start digging in and find it's too intimidating, this book can be deferred until you're more experienced with the language. The more JS you've written (and struggled with!), the more you'll come to appreciate this book. So don't be afraid to come back to it at a later time.

The final book in the series, *ES.Next & Beyond*, in some respects stands alone. It can be read at the end, as I suggest, or right after *Getting Started* if you're looking for a shortcut to broaden your radar of what JS is all about. This book will also be more likely to receive updates in the future, so you'll probably want to re-visit it occasionally.

However you choose to proceed with YDKJSY, check out the appendices of this book first, especially practicing the snippets in Appendix B, "Practice, Practice, Practice!" Did I mention you should go practice!? There's no better way to learn code than to write it.

# You Don't Know JS Yet: Get Started - 2nd Edition

## Appendix A: Exploring Further

In this appendix, we're going to explore some topics from the main chapter text in a bit more detail. Think of this content as an optional preview of some of the more nuanced details covered throughout the rest of the book series.

### Values vs. References

In Chapter 2, we introduced the two main types of values: primitives and objects. But we didn't discuss yet one key difference between the two: how these values are assigned and passed around.

In many languages, the developer can choose between assigning/passing a value as the value itself, or as a reference to the value. In JS, however, this decision is entirely determined by the kind of value. That surprises a lot of developers from other languages when they start using JS.

If you assign/pass a value itself, the value is copied. For example:

```
var myName = "Kyle";
var yourName = myName;
```

Here, the `yourName` variable has a separate copy of the `"Kyle"` string from the value that's stored in `myName`. That's because the value is a primitive, and primitive values are always assigned/passed as **value copies**.

Here's how you can prove there's two separate values involved:

```
var myName = "Kyle";
var yourName = myName;
myName = "Frank";
console.log(myName);
// Frank
console.log(yourName);
// Kyle
```

See how `yourName` wasn't affected by the re-assignment of `myName` to `"Frank"`? That's because each variable holds its own copy of the value.

By contrast, references are the idea that two or more variables are pointing at the same value, such that modifying this shared value would be reflected by an access via any of those references. In JS, only object values (arrays, objects, functions, etc.) are treated as references.

Consider:

```
var myAddress = {
  street: "123 JS Blvd",
  city: "Austin",
  state: "TX"
};
```

```
var yourAddress = myAddress;

// I've got to move to a new house!
myAddress.street = "456 TS Ave";

console.log(yourAddress.street);
// 456 TS Ave
```

Because the value assigned to `myAddress` is an object, it's held/assigned by reference, and thus the assignment to the `yourAddress` variable is a copy of the reference, not the object value itself. That's why the updated value assigned to the `myAddress.street` is reflected when we access `yourAddress.street`. `myAddress` and `yourAddress` have copies of the reference to the single shared object, so an update to one is an update to both.

Again, JS chooses the value-copy vs. reference-copy behavior based on the value type. Primitives are held by value, objects are held by reference. There's no way to override this in JS, in either direction.

## So Many Function Forms

Recall this snippet from the "Functions" section in Chapter 2:

```
var awesomeFunction = function(coolThings) {
  // ...
  return amazingStuff;
};
```

The function expression here is referred to as an *anonymous function expression*, since it has no name identifier between the `function` keyword and the `(...)` parameter list. This point confuses many JS developers because as of ES6, JS performs a "name inference" on an anonymous function:

```
awesomeFunction.name;
// "awesomeFunction"
```

The `name` property of a function will reveal either its directly given name (in the case of a declaration) or its inferred name in the case of an anonymous function expression. That value is generally used by developer tools when inspecting a function value or when reporting an error stack trace.

So even an anonymous function expression *might* get a name. However, name inference only happens in limited cases such as when the function expression is assigned (with `=`). If you pass a function expression as an argument to a function call, for example, no name inference occurs; the `name` property will be an empty string, and the developer console will usually report "*anonymous function*".

Even if a name is inferred, **it's still an anonymous function**. Why? Because the inferred name is a metadata string value, not an available identifier to refer to the function. An anonymous function doesn't have an identifier to use to refer to itself from inside itself—for recursion, event unbinding, etc.

Compare the anonymous function expression form to:

```
// let awesomeFunction = ...
// const awesomeFunction = ...
var awesomeFunction = function someName(coolThings) {
  // ...
  return amazingStuff;
};

awesomeFunction.name;
// "someName"
```

This function expression is a *named function expression*, since the identifier `someName` is directly associated with the function expression at compile time; the association with the identifier `awesomeFunction` still doesn't happen until runtime at the time of that statement. Those two identifiers don't have to match; sometimes it makes sense to have them be different, other times it's better to have them be the same.

Notice also that the explicit function name, the identifier `someName`, takes precedence when assigning a *name* for the `name` property.

Should function expressions be named or anonymous? Opinions vary widely on this. Most developers tend to be unconcerned with using anonymous functions. They're shorter, and unquestionably more common in the broad sphere of JS code out there.

In my opinion, if a function exists in your program, it has a purpose; otherwise, take it out! And if it has a purpose, it has a natural name that describes that purpose.

If a function has a name, you the code author should include that name in the code, so that the reader does not have to infer that name from reading and mentally executing that function's source code. Even a trivial function body like `x * 2` has to be read to infer a name like "double" or "multBy2"; that brief extra mental work is unnecessary when you could just take a second to name the function "double" or "multBy2" *once*, saving the reader that repeated mental work every time it's read in the future.

There are, regrettably in some respects, many other function definition forms in JS as of early 2020 (maybe more in the future!).

Here are some more declaration forms:

```
// generator function declaration
function *two() { .. }

// async function declaration
async function three() { .. }

// async generator function declaration
async function *four() { .. }

// named function export declaration (ES6 modules)
export function five() { .. }
```

And here are some more of the (many!) function expression forms:

```
// IIFE
(function(){ .. })();
(function namedIIFE(){ .. })();

// asynchronous IIFE
(async function(){ .. })();
(async function namedAIIFE(){ .. })();

// arrow function expressions
var f;
f = () => 42;
f = x => x * 2;
f = (x) => x * 2;
f = (x,y) => x * y;
f = x => ({ x: x * 2 });
f = x => { return x * 2; };
f = async x => {
  var y = await doSomethingAsync(x);
  return y * 2;
};
someOperation( x => x * 2 );
// ..
```

Keep in mind that arrow function expressions are **syntactically anonymous**, meaning the syntax doesn't provide a way to provide a direct name identifier for the function. The function expression may get an inferred name, but only if it's one of the assignment forms, not in the (more common!) form of being passed as a function call argument (as in the last line of the snippet).

Since I don't think anonymous functions are a good idea to use frequently in your programs, I'm not a fan of using the `>` arrow function form. This kind of function actually has a specific purpose (i.e., handling the `this` keyword lexically), but that doesn't mean we should use it for every function we write. Use the most appropriate tool for each job.

Functions can also be specified in class definitions and object literal definitions. They're typically referred to as "methods" when in these forms, though in JS this term doesn't have much observable difference over "function":

```
class SomethingKindaGreat {
  // class methods
  coolMethod() { ... }    // no commas!
  boringMethod() { ... }
}

var EntirelyDifferent = {
  // object methods
  coolMethod() { ... },   // commas!
  boringMethod() { ... },

  // (anonymous) function expression property
  oldSchool: function() { ... }
};
```

Phew! That's a lot of different ways to define functions.

There's no simple shortcut path here; you just have to build familiarity with all the function forms so you can recognize them in existing code and use them appropriately in the code you write. Study them closely and practice!

## Coercive Conditional Comparison

Yes, that section name is quite a mouthful. But what are we talking about? We're talking about conditional expressions needing to perform coercion-oriented comparisons to make their decisions.

`if` and `? :` -ternary statements, as well as the test clauses in `while` and `for` loops, all perform an implicit value comparison. But what sort? Is it "strict" or "coercive"? Both, actually.

Consider:

```
var x = 1;

if (x) {
  // will run!
}

while (x) {
  // will run, once!
  x = false;
}
```

You might think of these `(x)` conditional expressions like this:

```
var x = 1;

if (x == true) {
  // will run!
}
```

```
while (x == true) {
    // will run, once!
    x = false;
}
```

In this specific case -- the value of `x` being `1` -- that mental model works, but it's not accurate more broadly. Consider:

```
var x = "hello";

if (x) {
    // will run!
}

if (x == true) {
    // won't run :(
}
```

Oops. So what is the `if` statement actually doing? This is the more accurate mental model:

```
var x = "hello";

if (Boolean(x) == true) {
    // will run
}

// which is the same as:

if (Boolean(x) === true) {
    // will run
}
```

Since the `Boolean(...)` function always returns a value of type boolean, the `==` vs `===` in this snippet is irrelevant; they'll both do the same thing. But the important part is to see that before the comparison, a coercion occurs, from whatever type `x` currently is, to boolean.

You just can't get away from coercions in JS comparisons. Buckle down and learn them.

## Prototypal "Classes"

In Chapter 3, we introduced prototypes and showed how we can link objects through a prototype chain.

Another way of wiring up such prototype linkages served as the (honestly, ugly) predecessor to the elegance of the ES6 `class` system (see Chapter 2, "Classes"), and is referred to as prototypal classes.

### TIP:

While this style of code is quite uncommon in JS these days, it's still perplexingly rather common to be asked about it in job interviews!

Let's first recall the `Object.create(...)` style of coding:

```
var Classroom = {
    welcome() {
        console.log("Welcome, students!");
    }
};

var mathClass = Object.create(Classroom);

mathClass.welcome();
```

```
// Welcome, students!
```

Here, a `mathClass` object is linked via its prototype to a `classroom` object. Through this linkage, the function call `mathClass.welcome()` is delegated to the method defined on `classroom`.

The prototypal class pattern would have labeled this delegation behavior "inheritance," and alternatively have defined it (with the same behavior) as:

```
function Classroom() {
  // ...
}

classroom.prototype.welcome = function hello() {
  console.log("Welcome, students!");
};

var mathClass = new Classroom();

mathClass.welcome();
// Welcome, students!
```

All functions by default reference an empty object at a property named `prototype`. Despite the confusing naming, this is **not** the function's *prototype* (where the function is prototype linked to), but rather the prototype object to *link to* when other objects are created by calling the function with `new`.

We add a `welcome` property on that empty object (called `classroom.prototype`), pointing at the `hello()` function.

Then `new Classroom()` creates a new object (assigned to `mathClass`), and prototype links it to the existing `classroom.prototype` object.

Though `mathClass` does not have a `welcome()` property/function, it successfully delegates to the function `classroom.prototype.welcome()`.

This "prototypal class" pattern is now strongly discouraged, in favor of using ES6's `class` mechanism:

```
class Classroom {
  constructor() {
    // ...
  }

  welcome() {
    console.log("Welcome, students!");
  }
}

var mathClass = new Classroom();

mathClass.welcome();
// Welcome, students!
```

Under the covers, the same prototype linkage is wired up, but this `class` syntax fits the class-oriented design pattern much more cleanly than "prototypal classes".

# You Don't Know JS Yet: Get Started - 2nd Edition

## Appendix B: Practice, Practice, Practice!

In this appendix, we'll explore some exercises and their suggested solutions. These are just to *get you started* with practice over the concepts from the book.

### Practicing Comparisons

Let's practice working with value types and comparisons (Chapter 4, Pillar 3) where coercion will need to be involved.

`scheduleMeeting(..)` should take a start time (in 24-hour format as a string "hh:mm") and a meeting duration (number of minutes). It should return `true` if the meeting falls entirely within the work day (according to the times specified in `dayStart` and `dayEnd`); return `false` if the meeting violates the work day bounds.

```
const dayStart = "07:30";
const dayEnd = "17:45";

function scheduleMeeting(startTime,durationMinutes) {
    // ..TODO..
}

scheduleMeeting("7:00",15);      // false
scheduleMeeting("07:15",30);    // false
scheduleMeeting("7:30",30);     // true
scheduleMeeting("11:30",60);    // true
scheduleMeeting("17:00",45);    // true
scheduleMeeting("17:30",30);    // false
scheduleMeeting("18:00",15);    // false
```

Try to solve this yourself first. Consider the usage of equality and relational comparison operators, and how coercion impacts this code. Once you have code that works, *compare* your solution(s) to the code in "Suggested Solutions" at the end of this appendix.

### Practicing Closure

Now let's practice with closure (Chapter 4, Pillar 1).

The `range(..)` function takes a number as its first argument, representing the first number in a desired range of numbers. The second argument is also a number representing the end of the desired range (inclusive). If the second argument is omitted, then another function should be returned that expects that argument.

```
function range(start,end) {
    // ..TODO..
}

range(3,3);      // [3]
range(3,8);      // [3,4,5,6,7,8]
range(3,0);      // []

var start3 = range(3);
var start4 = range(4);

start3(3);       // [3]
start3(8);       // [3,4,5,6,7,8]
```

```
start3(0);      // []
start4(6);      // [4, 5, 6]
```

Try to solve this yourself first.

Once you have code that works, *compare* your solution(s) to the code in "Suggested Solutions" at the end of this appendix.

## Practicing Prototypes

Finally, let's work on `this` and objects linked via prototype (Chapter 4, Pillar 2).

Define a slot machine with three reels that can individually `spin()`, and then `display()` the current contents of all the reels.

The basic behavior of a single reel is defined in the `reel` object below. But the slot machine needs individual reels—objects that delegate to `reel`, and which each have a `position` property.

A reel only *knows how to* `display()` its current slot symbol, but a slot machine typically shows three symbols per reel: the current slot (`position`), one slot above (`position - 1`), and one slot below (`position + 1`). So displaying the slot machine should end up displaying a 3 x 3 grid of slot symbols.

```
function randMax(max) {
    return Math.trunc(1E9 * Math.random()) % max;
}

var reel = {
    symbols: [
        "♣", "♥", "♦", "♠", "◎", "★", "◎", "★"
    ],
    spin() {
        if (this.position == null) {
            this.position = randMax(
                this.symbols.length - 1
            );
        }
        this.position = (
            this.position + 100 + randMax(100)
        ) % this.symbols.length;
    },
    display() {
        if (this.position == null) {
            this.position = randMax(
                this.symbols.length - 1
            );
        }
        return this.symbols[this.position];
    }
};

var slotMachine = {
    reels: [
        // this slot machine needs 3 separate reels
        // hint: Object.create(..)
    ],
    spin() {
        this.reels.forEach(function spinReel(reel){
            reel.spin();
        });
    },
    display() {
        // TODO
    }
};
```

```

slotMachine.spin();
slotMachine.display();
// ⚡ | * | *
// * | ⚡ | ⚡
// ⚡ | ⚡ | *

slotMachine.spin();
slotMachine.display();
// ⚡ | ⚡ | ⚡
// ⚡ | ⚡ | ⚡
// ⚡ | ⚡ | ⚡

```

Try to solve this yourself first.

Hints:

- Use the `%` modulo operator for wrapping `position` as you access symbols circularly around a reel.
- Use `Object.create(..)` to create an object and prototype-link it to another object. Once linked, delegation allows the objects to share `this` context during method invocation.
- Instead of modifying the reel object directly to show each of the three positions, you can use another temporary object (`Object.create(..)` again) with its own `position`, to delegate from.

Once you have code that works, *compare* your solution(s) to the code in "Suggested Solutions" at the end of this appendix.

## Suggested Solutions

Keep in mind that these suggested solutions are just that: suggestions. There's many different ways to solve these practice exercises. Compare your approach to what you see here, and consider the pros and cons of each.

Suggested solution for "Comparisons" (Pillar 3) practice:

```

const dayStart = "07:30";
const dayEnd = "17:45";

function scheduleMeeting(startTime,durationMinutes) {
    var [ , meetingStartHour, meetingStartMinutes ] =
        startTime.match(/^\d{1,2}:(\d{2})$/) || [];
    durationMinutes = Number(durationMinutes);

    if (
        typeof meetingStartHour == "string" &&
        typeof meetingStartMinutes == "string"
    ) {
        let durationHours =
            Math.floor(durationMinutes / 60);
        durationMinutes =
            durationMinutes - (durationHours * 60);
        let meetingEndHour =
            Number(meetingStartHour) + durationHours;
        let meetingEndMinutes =
            Number(meetingStartMinutes) +
            durationMinutes;

        if (meetingEndMinutes >= 60) {
            meetingEndHour = meetingEndHour + 1;
            meetingEndMinutes =
                meetingEndMinutes - 60;
        }
    }

    // re-compose fully-qualified time strings
    // (to make comparison easier)

```

```

        let meetingStart = `${{
            meetingStartHour.padStart(2, "0")
        }}:${{
            meetingStartMinutes.padStart(2, "0")
        }}`;
        let meetingEnd = `${{
            String(meetingEndHour).padStart(2, "0")
        }}:${{
            String(meetingEndMinutes).padStart(2, "0")
        }}`;

        // NOTE: since expressions are all strings,
        // comparisons here are alphabetic, but it's
        // safe here since they're fully qualified
        // time strings (ie, "07:15" < "07:30")
        return (
            meetingStart >= dayStart &&
            meetingEnd <= dayEnd
        );
    }

    return false;
}

scheduleMeeting("7:00");      // false
scheduleMeeting("07:15");     // false
scheduleMeeting("7:30");      // true
scheduleMeeting("11:30");     // true
scheduleMeeting("17:00");     // true
scheduleMeeting("17:30");     // false
scheduleMeeting("18:00");     // false

```

Suggested solution for "Closure" (Pillar 1) practice:

```

function range(start,end) {
    start = Number(start) || 0;

    if (end === undefined) {
        return function getEnd(end) {
            return getRange(start,end);
        };
    }
    else {
        end = Number(end) || 0;
        return getRange(start,end);
    }
}

// *****

function getRange(start,end) {
    var ret = [];
    for (let i = start; i <= end; i++) {
        ret.push(i);
    }
    return ret;
}

range(3,3);      // [3]
range(3,8);      // [3,4,5,6,7,8]
range(3,0);      // []

var start3 = range(3);
var start4 = range(4);

```

```

start3(3);      // [3]
start3(8);      // [3, 4, 5, 6, 7, 8]
start3(0);      // []

start4(6);      // [4, 5, 6]

```

Suggested solution for "Prototypes" (Pillar 2) practice:

```

function randMax(max) {
    return Math.trunc(1E9 * Math.random()) % max;
}

var reel = {
    symbols: [
        "♣", "♥", "♦", "♠", "⊗", "★", "₵", "*"
    ],
    spin() {
        if (this.position == null) {
            this.position = randMax(
                this.symbols.length - 1
            );
        }
        this.position = (
            this.position + 100 + randMax(100)
        ) % this.symbols.length;
    },
    display() {
        if (this.position == null) {
            this.position = randMax(
                this.symbols.length - 1
            );
        }
        return this.symbols[this.position];
    }
};

var slotMachine = {
    reels: [
        Object.create(reel),
        Object.create(reel),
        Object.create(reel)
    ],
    spin() {
        this.reels.forEach(function spinReel(reel){
            reel.spin();
        });
    },
    display() {
        var lines = [];

        // display all 3 lines on the slot machine
        for (
            let linePos = -1; linePos <= 1; linePos++
        ) {
            let line = this.reels.map(
                function getSlot(reel){
                    var slot = Object.create(reel);
                    slot.position = (
                        reel.symbols.length +
                        reel.position +
                        linePos
                    ) % reel.symbols.length;
                    return reel.display.call(slot);
                }
            );
        }
    }
};

```

```
        );
        lines.push(line.join(" | "));
    }

    return lines.join("\n");
};

slotMachine.spin();
slotMachine.display();
// C | * | *
// * | ♦ | C
// ♦ | ♥ | *

slotMachine.spin();
slotMachine.display();
// ♦ | ♦ | ♦
// ♦ | ♥ | ⊙
// ⊙ | ♦ | *
```

That's it for this book. But now it's time to look for real projects to practice these ideas on. Just keep coding, because that's the best way to learn!

## You Don't Know JS Yet: Scope & Closures - 2nd Edition



---

[Purchase ebook/PDF from Leanpub](#)

---

### Table of Contents

- Foreword (by Sarah Drasner)
- Preface
- Chapter 1: What's the Scope?
- Chapter 2: Illustrating Lexical Scope
- Chapter 3: The Scope Chain
- Chapter 4: Around the Global Scope
- Chapter 5: The (Not So) Secret Lifecycle of Variables
- Chapter 6: Limiting Scope Exposure
- Chapter 7: Using Closures
- Chapter 8: The Module Pattern
- Appendix A: Exploring Further
- Appendix B: Practice

# You Don't Know JS Yet: Scope & Closures - 2nd Edition

## Chapter 1: What's the Scope?

By the time you've written your first few programs, you're likely getting somewhat comfortable with creating variables and storing values in them. Working with variables is one of the most foundational things we do in programming!

But you may not have considered very closely the underlying mechanisms used by the engine to organize and manage these variables. I don't mean how the memory is allocated on the computer, but rather: how does JS know which variables are accessible by any given statement, and how does it handle two variables of the same name?

The answers to questions like these take the form of well-defined rules called scope. This book will dig through all aspects of scope—how it works, what it's useful for, gotchas to avoid—and then point toward common scope patterns that guide the structure of programs.

Our first step is to uncover how the JS engine processes our program **before** it runs.

## About This Book

Welcome to book 2 in the *You Don't Know JS Yet* series! If you already finished *Get Started* (the first book), you're in the right spot! If not, before you proceed I encourage you to *start there* for the best foundation.

Our focus will be the first of three pillars in the JS language: the scope system and its function closures, as well as the power of the module design pattern.

JS is typically classified as an interpreted scripting language, so it's assumed by most that JS programs are processed in a single, top-down pass. But JS is in fact parsed/compiled in a separate phase **before execution begins**. The code author's decisions on where to place variables, functions, and blocks with respect to each other are analyzed according to the rules of scope, during the initial parsing/compilation phase. The resulting scope structure is generally unaffected by runtime conditions.

JS functions are themselves first-class values; they can be assigned and passed around just like numbers or strings. But since these functions hold and access variables, they maintain their original scope no matter where in the program the functions are eventually executed. This is called closure.

Modules are a code organization pattern characterized by public methods that have privileged access (via closure) to hidden variables and functions in the internal scope of the module.

## Compiled vs. Interpreted

You may have heard of *code compilation* before, but perhaps it seems like a mysterious black box where source code slides in one end and executable programs pop out the other.

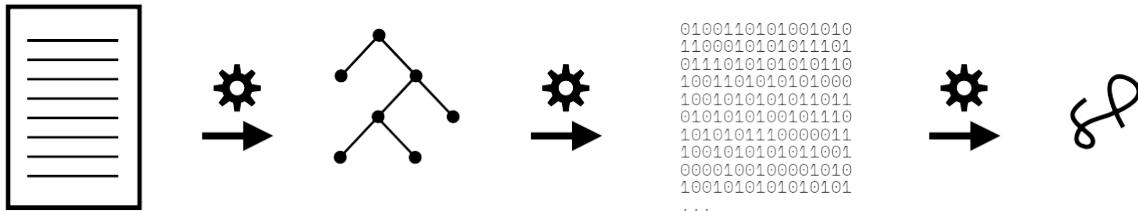
It's not mysterious or magical, though. Code compilation is a set of steps that process the text of your code and turn it into a list of instructions the computer can understand. Typically, the whole source code is transformed at once, and those resulting instructions are saved as output (usually in a file) that can later be executed.

You also may have heard that code can be *interpreted*, so how is that different from being *compiled*?

Interpretation performs a similar task to compilation, in that it transforms your program into machine-understandable instructions. But the processing model is different. Unlike a program being compiled all at once, with interpretation the source code is transformed line by line; each line or statement is executed before immediately proceeding to processing the next line of the

source code.

## Compilation:



## Interpretation:

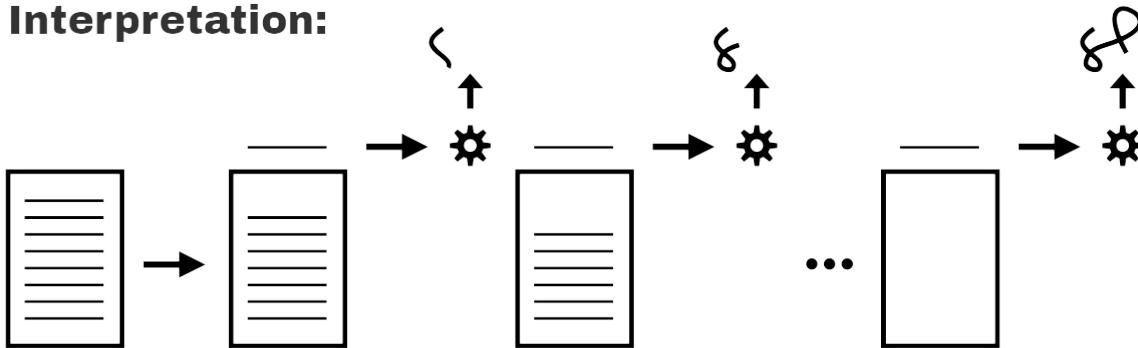


Fig. 1: Compiled vs. Interpreted Code

Figure 1 illustrates compilation vs. interpretation of programs.

Are these two processing models mutually exclusive? Generally, yes. However, the issue is more nuanced, because interpretation can actually take other forms than just operating line by line on source code text. Modern JS engines actually employ numerous variations of both compilation and interpretation in the handling of JS programs.

Recall that we surveyed this topic in Chapter 1 of the *Get Started* book. Our conclusion there is that JS is most accurately portrayed as a **compiled language**. For the benefit of readers here, the following sections will revisit and expand on that assertion.

## Compiling Code

But first, why does it even matter whether JS is compiled or not?

Scope is primarily determined during compilation, so understanding how compilation and execution relate is key in mastering scope.

In classic compiler theory, a program is processed by a compiler in three basic stages:

1. **Tokenizing/Lexing:** breaking up a string of characters into meaningful (to the language) chunks, called tokens. For instance, consider the program: `var a = 2;`. This program would likely be broken up into the following tokens: `var` , `a` , `=` , `2` , and `;`. Whitespace may or may not be persisted as a token, depending on whether it's meaningful or not.

(The difference between tokenizing and lexing is subtle and academic, but it centers on whether or not these tokens are identified in a *stateless* or *stateful* way. Put simply, if the tokenizer were to invoke stateful parsing rules to figure out whether `a` should be considered a distinct token or just part of another token, *that would be lexing*.)

2. **Parsing:** taking a stream (array) of tokens and turning it into a tree of nested elements, which collectively represent the grammatical structure of the program. This is called an Abstract Syntax Tree (AST).

For example, the tree for `var a = 2;` might start with a top-level node called `variableDeclaration`, with a child node called `Identifier` (whose value is `a`), and another child called `AssignmentExpression` which itself has a child called `NumericLiteral` (whose value is `2`).

3. **Code Generation:** taking an AST and turning it into executable code. This part varies greatly depending on the language, the platform it's targeting, and other factors.

The JS engine takes the just described AST for `var a = 2;` and turns it into a set of machine instructions to actually *create* a variable called `a` (including reserving memory, etc.), and then store a value into `a`.

#### NOTE:

The implementation details of a JS engine (utilizing system memory resources, etc.) is much deeper than we will dig here. We'll keep our focus on the observable behavior of our programs and let the JS engine manage those deeper system-level abstractions.

The JS engine is vastly more complex than just these three stages. In the process of parsing and code generation, there are steps to optimize the performance of the execution (i.e., collapsing redundant elements). In fact, code can even be re-compiled and re-optimized during the progression of execution.

So, I'm painting only with broad strokes here. But you'll see shortly why *these* details we *do* cover, even at a high level, are relevant.

JS engines don't have the luxury of an abundance of time to perform their work and optimizations, because JS compilation doesn't happen in a build step ahead of time, as with other languages. It usually must happen in mere microseconds (or less!) right before the code is executed. To ensure the fastest performance under these constraints, JS engines use all kinds of tricks (like JITs, which lazy compile and even hot re-compile); these are well beyond the "scope" of our discussion here.

## Required: Two Phases

To state it as simply as possible, the most important observation we can make about processing of JS programs is that it occurs in (at least) two phases: parsing/compilation first, then execution.

The separation of a parsing/compilation phase from the subsequent execution phase is observable fact, not theory or opinion. While the JS specification does not require "compilation" explicitly, it requires behavior that is essentially only practical with a compile-then-execute approach.

There are three program characteristics you can observe to prove this to yourself: syntax errors, early errors, and hoisting.

## Syntax Errors from the Start

Consider this program:

```
var greeting = "Hello";
console.log(greeting);

greeting = ."Hi";
// SyntaxError: unexpected token .
```

This program produces no output (`"Hello"` is not printed), but instead throws a `SyntaxError` about the unexpected `.` token right before the `"Hi"` string. Since the syntax error happens after the well-formed `console.log(..)` statement, if JS was executing top-down line by line, one would expect the `"Hello"` message being printed before the syntax error being thrown. That doesn't happen.

In fact, the only way the JS engine could know about the syntax error on the third line, before executing the first and second lines, is by the JS engine first parsing the entire program before any of it is executed.

## Early Errors

Next, consider:

```
console.log("Howdy");

saySomething("Hello", "Hi");
// Uncaught SyntaxError: Duplicate parameter name not
// allowed in this context

function saySomething(greeting, greeting) {
  "use strict";
  console.log(greeting);
}
```

The `"Howdy"` message is not printed, despite being a well-formed statement.

Instead, just like the snippet in the previous section, the `SyntaxError` here is thrown before the program is executed. In this case, it's because strict-mode (opted in for only the `saySomething(..)` function here) forbids, among many other things, functions to have duplicate parameter names; this has always been allowed in non-strict-mode.

The error thrown is not a syntax error in the sense of being a malformed string of tokens (like `."Hi"` prior), but in strict-mode is nonetheless required by the specification to be thrown as an "early error" before any execution begins.

But how does the JS engine know that the `greeting` parameter has been duplicated? How does it know that the `saySomething(..)` function is even in strict-mode while processing the parameter list (the `"use strict"` pragma appears only later, in the function body)?

Again, the only reasonable explanation is that the code must first be *fully* parsed before any execution occurs.

## Hoisting

Finally, consider:

```
function saySomething() {
  var greeting = "Hello";
  {
    greeting = "Howdy"; // error comes from here
    let greeting = "Hi";
    console.log(greeting);
  }
}

saySomething();
// ReferenceError: Cannot access 'greeting' before
// initialization
```

The noted `ReferenceError` occurs from the line with the statement `greeting = "Howdy"`. What's happening is that the `greeting` variable for that statement belongs to the declaration on the next line, `let greeting = "Hi"`, rather than to the previous `var greeting = "Hello"` statement.

The only way the JS engine could know, at the line where the error is thrown, that the *next statement* would declare a block-scoped variable of the same name (`greeting`) is if the JS engine had already processed this code in an earlier pass, and already set up all the scopes and their variable associations. This processing of scopes and declarations can only accurately be accomplished by parsing the program before execution.

The `ReferenceError` here technically comes from `greeting = "Howdy"` accessing the `greeting` variable **too early**, a conflict referred to as the Temporal Dead Zone (TDZ). Chapter 5 will cover this in more detail.

#### WARNING:

It's often asserted that `let` and `const` declarations are not hoisted, as an explanation of the TDZ behavior just illustrated. But this is not accurate. We'll come back and explain both the hoisting and TDZ of `let` / `const` in Chapter 5.

Hopefully you're now convinced that JS programs are parsed before any execution begins. But does it prove they are compiled?

This is an interesting question to ponder. Could JS parse a program, but then execute that program by *interpreting* operations represented in the AST **without** first compiling the program? Yes, that is *possible*. But it's extremely unlikely, mostly because it would be extremely inefficient performance wise.

It's hard to imagine a production-quality JS engine going to all the trouble of parsing a program into an AST, but not then converting (aka, "compiling") that AST into the most efficient (binary) representation for the engine to then execute.

Many have endeavored to split hairs with this terminology, as there's plenty of nuance and "well, actually..." interjections floating around. But in spirit and in practice, what the engine is doing in processing JS programs is **much more alike compilation** than not.

Classifying JS as a compiled language is not concerned with the distribution model for its binary (or byte-code) executable representations, but rather in keeping a clear distinction in our minds about the phase where JS code is processed and analyzed; this phase observably and indisputably happens *before* the code starts to be executed.

We need proper mental models of how the JS engine treats our code if we want to understand JS and scope effectively.

## Compiler Speak

With awareness of the two-phase processing of a JS program (compile, then execute), let's turn our attention to how the JS engine identifies variables and determines the scopes of a program as it is compiled.

First, let's examine a simple JS program to use for analysis over the next several chapters:

```
var students = [
    { id: 14, name: "Kyle" },
    { id: 73, name: "Suzy" },
    { id: 112, name: "Frank" },
    { id: 6, name: "Sarah" }
];

function getStudentName(studentID) {
    for (let student of students) {
        if (student.id == studentID) {
            return student.name;
        }
    }
}

var nextStudent = getStudentName(73);

console.log(nextStudent);
// Suzy
```

Other than declarations, all occurrences of variables/identifiers in a program serve in one of two "roles": either they're the *target* of an assignment or they're the *source* of a value.

(When I first learned compiler theory while earning my computer science degree, we were taught the terms "LHS" (aka, *target*) and "RHS" (aka, *source*) for these roles, respectively. As you might guess from the "L" and the "R", the acronyms mean "Left-Hand Side" and "Right-Hand Side", as in left and right sides of an `=` assignment operator. However, assignment targets and sources don't always literally appear on the left or right of an `=`, so it's probably clearer to think in terms of *target / source* rather than *left / right*.)

How do you know if a variable is a *target*? Check if there is a value that is being assigned to it; if so, it's a *target*. If not, then the variable is a *source*.

For the JS engine to properly handle a program's variables, it must first label each occurrence of a variable as *target* or *source*. We'll dig in now to how each role is determined.

## Targets

What makes a variable a *target*? Consider:

```
students = [ // ..
```

This statement is clearly an assignment operation; remember, the `var students` part is handled entirely as a declaration at compile time, and is thus irrelevant during execution; we left it out for clarity and focus. Same with the `nextStudent = getStudentName(73)` statement.

But there are three other *target* assignment operations in the code that are perhaps less obvious. One of them:

```
for (let student of students) {
```

That statement assigns a value to `student` for each iteration of the loop. Another *target* reference:

```
getStudentName(73)
```

But how is that an assignment to a *target*? Look closely: the argument `73` is assigned to the parameter `studentID`.

And there's one last (subtle) *target* reference in our program. Can you spot it?

..  
..  
..

Did you identify this one?

```
function getStudentName(studentID) {
```

A `function` declaration is a special case of a *target* reference. You can think of it sort of like `var getStudentName = function(studentID)`, but that's not exactly accurate. An identifier `getStudentName` is declared (at compile time), but the `= function(studentID)` part is also handled at compilation; the association between `getStudentName` and the function is automatically set up at the beginning of the scope rather than waiting for an `=` assignment statement to be executed.

### NOTE:

This automatic association of function and variable is referred to as "function hoisting", and is covered in detail in Chapter 5.

## Sources

So we've identified all five *target* references in the program. The other variable references must then be *source* references (because that's the only other option!).

In `for (let student of students)`, we said that `student` is a *target*, but `students` is a *source* reference. In the statement `if (student.id == studentID)`, both `student` and `studentID` are *source* references. `student` is also a *source* reference in `return student.name`.

In `getStudentName(73)`, `getStudentName` is a *source* reference (which we hope resolves to a function reference value). In `console.log(nextStudent)`, `console` is a *source* reference, as is `nextStudent`.

**NOTE:**

In case you were wondering, `id`, `name`, and `log` are all properties, not variable references.

What's the practical importance of understanding *targets* vs. *sources*? In Chapter 2, we'll revisit this topic and cover how a variable's role impacts its lookup (specifically, if the lookup fails).

## Cheating: Runtime Scope Modifications

It should be clear by now that scope is determined as the program is compiled, and should not generally be affected by runtime conditions. However, in non-strict-mode, there are technically still two ways to cheat this rule, modifying a program's scopes during runtime.

Neither of these techniques *should* be used—they're both dangerous and confusing, and you should be using strict-mode (where they're disallowed) anyway. But it's important to be aware of them in case you run across them in some programs.

The `eval(..)` function receives a string of code to compile and execute on the fly during the program runtime. If that string of code has a `var` or `function` declaration in it, those declarations will modify the current scope that the `eval(..)` is currently executing in:

```
function badIdea() {
    eval("var oops = 'Ugh!');");
    console.log(oops);
}
badIdea(); // Ugh!
```

If the `eval(..)` had not been present, the `oops` variable in `console.log(oops)` would not exist, and would throw a `ReferenceError`. But `eval(..)` modifies the scope of the `badIdea()` function at runtime. This is bad for many reasons, including the performance hit of modifying the already compiled and optimized scope, every time `badIdea()` runs.

The second cheat is the `with` keyword, which essentially dynamically turns an object into a local scope—its properties are treated as identifiers in that new scope's block:

```
var badIdea = { oops: "Ugh!" };

with (badIdea) {
    console.log(oops); // Ugh!
}
```

The global scope was not modified here, but `badIdea` was turned into a scope at runtime rather than compile time, and its property `oops` becomes a variable in that scope. Again, this is a terrible idea, for performance and readability reasons.

At all costs, avoid `eval(..)` (at least, `eval(..)` creating declarations) and `with`. Again, neither of these cheats is available in strict-mode, so if you just use strict-mode (you should!) then the temptation goes away!

## Lexical Scope

We've demonstrated that JS's scope is determined at compile time; the term for this kind of scope is "lexical scope". "Lexical" is associated with the "lexing" stage of compilation, as discussed earlier in this chapter.

To narrow this chapter down to a useful conclusion, the key idea of "lexical scope" is that it's controlled entirely by the placement of functions, blocks, and variable declarations, in relation to one another.

If you place a variable declaration inside a function, the compiler handles this declaration as it's parsing the function, and associates that declaration with the function's scope. If a variable is block-scope declared (`let` / `const`), then it's associated with the nearest enclosing `{ ... }` block, rather than its enclosing function (as with `var`).

Furthermore, a reference (*target* or *source* role) for a variable must be resolved as coming from one of the scopes that are *lexically available* to it; otherwise the variable is said to be "undeclared" (which usually results in an error!). If the variable is not declared in the current scope, the next outer/enclosing scope will be consulted. This process of stepping out one level of scope nesting continues until either a matching variable declaration can be found, or the global scope is reached and there's nowhere else to go.

It's important to note that compilation doesn't actually *do anything* in terms of reserving memory for scopes and variables. None of the program has been executed yet.

Instead, compilation creates a map of all the lexical scopes that lays out what the program will need while it executes. You can think of this plan as inserted code for use at runtime, which defines all the scopes (aka, "lexical environments") and registers all the identifiers (variables) for each scope.

In other words, while scopes are identified during compilation, they're not actually created until runtime, each time a scope needs to run. In the next chapter, we'll sketch out the conceptual foundations for lexical scope.

# You Don't Know JS Yet: Scope & Closures - 2nd Edition

## Chapter 2: Illustrating Lexical Scope

In Chapter 1, we explored how scope is determined during code compilation, a model called "lexical scope." The term "lexical" refers to the first stage of compilation (lexing/parsing).

To properly *reason* about our programs, it's important to have a solid conceptual foundation of how scope works. If we rely on guesses and intuition, we may accidentally get the right answers some of the time, but many other times we're far off. This isn't a recipe for success.

Like way back in grade school math class, getting the right answer isn't enough if we don't show the correct steps to get there! We need to build accurate and helpful mental models as foundation moving forward.

This chapter will illustrate *scope* with several metaphors. The goal here is to *think* about how your program is handled by the JS engine in ways that more closely align with how the JS engine actually works.

## Marbles, and Buckets, and Bubbles... Oh My!

One metaphor I've found effective in understanding scope is sorting colored marbles into buckets of their matching color.

Imagine you come across a pile of marbles, and notice that all the marbles are colored red, blue, or green. Let's sort all the marbles, dropping the red ones into a red bucket, green into a green bucket, and blue into a blue bucket. After sorting, when you later need a green marble, you already know the green bucket is where to go to get it.

In this metaphor, the marbles are the variables in our program. The buckets are scopes (functions and blocks), which we just conceptually assign individual colors for our discussion purposes. The color of each marble is thus determined by which *color* scope we find the marble originally created in.

Let's annotate the running program example from Chapter 1 with scope color labels:

```
// outer/global scope: RED

var students = [
  { id: 14, name: "Kyle" },
  { id: 73, name: "Suzy" },
  { id: 112, name: "Frank" },
  { id: 6, name: "Sarah" }
];

function getStudentName(studentID) {
  // function scope: BLUE

  for (let student of students) {
    // loop scope: GREEN

    if (student.id == studentID) {
      return student.name;
    }
  }
}

var nextStudent = getStudentName(73);
console.log(nextStudent); // Suzy
```

We've designated three scope colors with code comments: RED (outermost global scope), BLUE (scope of function `getStudentName( ... )`), and GREEN (scope of/inside the `for` loop). But it still may be difficult to recognize the boundaries of these scope buckets when looking at a code listing.

Figure 2 helps visualize the boundaries of the scopes by drawing colored bubbles (aka, buckets) around each:

```

1  var students = [
2      { id: 14, name: "Kyle" },
3      { id: 73, name: "Suzy" },
4      { id: 112, name: "Frank" },
5      { id: 6, name: "Sarah" }
6  ];
7
8  function getStudentName(studentID) {
9      for (let student of students) {
10          if (student.id == studentID) {
11              return student.name;
12          }
13      }
14
15
16  var nextStudent = getStudentName(73);
17
18  console.log(nextStudent);
19  // "Suzy"

```

*Fig. 2: Colored Scope Bubbles*

1. **Bubble 1 (RED)** encompasses the global scope, which holds three identifiers/variables: `students` (line 1), `getStudentName` (line 8), and `nextStudent` (line 16).
2. **Bubble 2 (BLUE)** encompasses the scope of the function `getStudentName( ... )` (line 8), which holds just one identifier/variable: the parameter `studentID` (line 8).
3. **Bubble 3 (GREEN)** encompasses the scope of the `for` -loop (line 9), which holds just one identifier/variable: `student` (line 9).

**NOTE:**

Technically, the parameter `studentID` is not exactly in the BLUE(2) scope. We'll unwind that confusion in "Implied Scopes" in Appendix A. For now, it's close enough to label `studentID` a BLUE(2) marble.

Scope bubbles are determined during compilation based on where the functions/blocks of scope are written, the nesting inside each other, and so on. Each scope bubble is entirely contained within its parent scope bubble—a scope is never partially in two different outer scopes.

Each marble (variable/identifier) is colored based on which bubble (bucket) it's declared in, not the color of the scope it may be accessed from (e.g., `students` on line 9 and `studentID` on line 10).

#### NOTE:

Remember we asserted in Chapter 1 that `id`, `name`, and `log` are all properties, not variables; in other words, they're not marbles in buckets, so they don't get colored based on any the rules we're discussing in this book. To understand how such property accesses are handled, see the third book in the series, *Objects & Classes*.

As the JS engine processes a program (during compilation), and finds a declaration for a variable, it essentially asks, "Which *color* scope (bubble or bucket) am I currently in?" The variable is designated as that same *color*, meaning it belongs to that bucket/bubble.

The GREEN(3) bucket is wholly nested inside of the BLUE(2) bucket, and similarly the BLUE(2) bucket is wholly nested inside the RED(1) bucket. Scopes can nest inside each other as shown, to any depth of nesting as your program needs.

References (non-declarations) to variables/identifiers are allowed if there's a matching declaration either in the current scope, or any scope above/outside the current scope, but not with declarations from lower/nested scopes.

An expression in the RED(1) bucket only has access to RED(1) marbles, **not** BLUE(2) or GREEN(3). An expression in the BLUE(2) bucket can reference either BLUE(2) or RED(1) marbles, **not** GREEN(3). And an expression in the GREEN(3) bucket has access to RED(1), BLUE(2), and GREEN(3) marbles.

We can conceptualize the process of determining these non-declaration marble colors during runtime as a lookup. Since the `students` variable reference in the `for`-loop statement on line 9 is not a declaration, it has no color. So we ask the current BLUE(2) scope bucket if it has a marble matching that name. Since it doesn't, the lookup continues with the next outer/containing scope: RED(1). The RED(1) bucket has a marble of the name `students`, so the loop-statement's `students` variable reference is determined to be a RED(1) marble.

The `if (student.id == studentID)` statement on line 10 is similarly determined to reference a GREEN(3) marble named `student` and a BLUE(2) marble `studentID`.

#### NOTE:

The JS engine doesn't generally determine these marble colors during runtime; the "lookup" here is a rhetorical device to help you understand the concepts. During compilation, most or all variable references will match already-known scope buckets, so their color is already determined, and stored with each marble reference to avoid unnecessary lookups as the program runs. More on this nuance in Chapter 3.

The key take-aways from marbles & buckets (and bubbles!):

- Variables are declared in specific scopes, which can be thought of as colored marbles from matching-color buckets.
- Any variable reference that appears in the scope where it was declared, or appears in any deeper nested scopes, will be labeled a marble of that same color—unless an intervening scope "shadows" the variable declaration; see "Shadowing" in Chapter 3.
- The determination of colored buckets, and the marbles they contain, happens during compilation. This information is used for variable (marble color) "lookups" during code execution.

## A Conversation Among Friends

Another useful metaphor for the process of analyzing variables and the scopes they come from is to imagine various conversations that occur inside the engine as code is processed and then executed. We can "listen in" on these conversations to get a better conceptual foundation for how scopes work.

Let's now meet the members of the JS engine that will have conversations as they process our program:

- *Engine*: responsible for start-to-finish compilation and execution of our JavaScript program.
- *Compiler*: one of *Engine*'s friends; handles all the dirty work of parsing and code-generation (see previous section).
- *Scope Manager*: another friend of *Engine*; collects and maintains a lookup list of all the declared variables/identifiers, and enforces a set of rules as to how these are accessible to currently executing code.

For you to *fully understand* how JavaScript works, you need to begin to *think* like *Engine* (and friends) think, ask the questions they ask, and answer their questions likewise.

To explore these conversations, recall again our running program example:

```
var students = [
  { id: 14, name: "Kyle" },
  { id: 73, name: "Suzy" },
  { id: 112, name: "Frank" },
  { id: 6, name: "Sarah" }
];

function getStudentName(studentID) {
  for (let student of students) {
    if (student.id == studentID) {
      return student.name;
    }
  }
}

var nextStudent = getStudentName(73);

console.log(nextStudent);
// Suzy
```

Let's examine how JS is going to process that program, specifically starting with the first statement. The array and its contents are just basic JS value literals (and thus unaffected by any scoping concerns), so our focus here will be on the `var students = [ .. ]` declaration and initialization-assignment parts.

We typically think of that as a single statement, but that's not how our friend *Engine* sees it. In fact, JS treats these as two distinct operations, one which *Compiler* will handle during compilation, and the other which *Engine* will handle during execution.

The first thing *Compiler* will do with this program is perform lexing to break it down into tokens, which it will then parse into a tree (AST).

Once *Compiler* gets to code generation, there's more detail to consider than may be obvious. A reasonable assumption would be that *Compiler* will produce code for the first statement such as: "Allocate memory for a variable, label it `students`, then stick a reference to the array into that variable." But that's not the whole story.

Here's the steps *Compiler* will follow to handle that statement:

1. Encountering `var students`, *Compiler* will ask *Scope Manager* to see if a variable named `students` already exists for that particular scope bucket. If so, *Compiler* would ignore this declaration and move on. Otherwise, *Compiler* will produce code that (at execution time) asks *Scope Manager* to create a new variable called `students` in that scope bucket.
2. *Compiler* then produces code for *Engine* to later execute, to handle the `students = []` assignment. The code *Engine* runs will first ask *Scope Manager* if there is a variable called `students` accessible in the current scope bucket. If not, *Engine* keeps looking elsewhere (see "Nested Scope" below). Once *Engine* finds a variable, it assigns the reference of the `[ .. ]` array to it.

In conversational form, the first phase of compilation for the program might play out between *Compiler* and *Scope Manager* like this:

**Compiler:** Hey, *Scope Manager* (of the global scope), I found a formal declaration for an identifier called `students`, ever heard of it?

**(Global) Scope Manager:** Nope, never heard of it, so I just created it for you.

**Compiler:** Hey, *Scope Manager*, I found a formal declaration for an identifier called `getStudentName`, ever heard of it?

**(Global) Scope Manager:** Nope, but I just created it for you.

**Compiler:** Hey, *Scope Manager*, `getStudentName` points to a function, so we need a new scope bucket.

**(Function) Scope Manager:** Got it, here's the scope bucket.

**Compiler:** Hey, *Scope Manager* (of the function), I found a formal parameter declaration for `studentID`, ever heard of it?

**(Function) Scope Manager:** Nope, but now it's created in this scope.

**Compiler:** Hey, *Scope Manager* (of the function), I found a `for`-loop that will need its own scope bucket.

...

The conversation is a question-and-answer exchange, where **Compiler** asks the current *Scope Manager* if an encountered identifier declaration has already been encountered. If "no," *Scope Manager* creates that variable in that scope. If the answer is "yes," then it's effectively skipped over since there's nothing more for that *Scope Manager* to do.

*Compiler* also signals when it runs across functions or block scopes, so that a new scope bucket and *Scope Manager* can be instantiated.

Later, when it comes to execution of the program, the conversation will shift to *Engine* and *Scope Manager*, and might play out like this:

**Engine:** Hey, *Scope Manager* (of the global scope), before we begin, can you look up the identifier `getStudentName` so I can assign this function to it?

**(Global) Scope Manager:** Yep, here's the variable.

**Engine:** Hey, *Scope Manager*, I found a *target* reference for `students`, ever heard of it?

**(Global) Scope Manager:** Yes, it was formally declared for this scope, so here it is.

**Engine:** Thanks, I'm initializing `students` to `undefined`, so it's ready to use.

Hey, *Scope Manager* (of the global scope), I found a *target* reference for `nextStudent`, ever heard of it?

**(Global) Scope Manager:** Yes, it was formally declared for this scope, so here it is.

**Engine:** Thanks, I'm initializing `nextStudent` to `undefined`, so it's ready to use.

Hey, *Scope Manager* (of the global scope), I found a *source* reference for `getStudentName`, ever heard of it?

**(Global) Scope Manager:** Yes, it was formally declared for this scope. Here it is.

**Engine:** Great, the value in `getStudentName` is a function, so I'm going to execute it.

**Engine:** Hey, *Scope Manager*, now we need to instantiate the function's scope.

...

This conversation is another question-and-answer exchange, where *Engine* first asks the current *Scope Manager* to look up the hoisted `getStudentName` identifier, so as to associate the function with it. *Engine* then proceeds to ask *Scope Manager* about the *target* reference for `students`, and so on.

To review and summarize how a statement like `var students = [ ... ]` is processed, in two distinct steps:

1. *Compiler* sets up the declaration of the scope variable (since it wasn't previously declared in the current scope).
2. While *Engine* is executing, to process the assignment part of the statement, *Engine* asks *Scope Manager* to look up the variable, initializes it to `undefined` so it's ready to use, and then assigns the array value to it.

## Nested Scope

When it comes time to execute the `getStudentName()` function, *Engine* asks for a *Scope Manager* instance for that function's scope, and it will then proceed to look up the parameter (`studentID`) to assign the `73` argument value to, and so on.

The function scope for `getStudentName(...)` is nested inside the global scope. The block scope of the `for`-loop is similarly nested inside that function scope. Scopes can be lexically nested to any arbitrary depth as the program defines.

Each scope gets its own *Scope Manager* instance each time that scope is executed (one or more times). Each scope automatically has all its identifiers registered at the start of the scope being executed (this is called "variable hoisting"; see Chapter 5).

At the beginning of a scope, if any identifier came from a `function` declaration, that variable is automatically initialized to its associated function reference. And if any identifier came from a `var` declaration (as opposed to `let` / `const`), that variable is automatically initialized to `undefined` so that it can be used; otherwise, the variable remains uninitialized (aka, in its "TDZ," see Chapter 5) and cannot be used until its full declaration-and-initialization are executed.

In the `for (let student of students) {` statement, `students` is a *source* reference that must be looked up. But how will that lookup be handled, since the scope of the function will not find such an identifier?

To explain, let's imagine that bit of conversation playing out like this:

**Engine:** Hey, *Scope Manager* (for the function), I have a *source* reference for `students`, ever heard of it?

**(Function) Scope Manager:** Nope, never heard of it. Try the next outer scope.

**Engine:** Hey, *Scope Manager* (for the global scope), I have a *source* reference for `students`, ever heard of it?

**(Global) Scope Manager:** Yep, it was formally declared, here it is.

...

One of the key aspects of lexical scope is that any time an identifier reference cannot be found in the current scope, the next outer scope in the nesting is consulted; that process is repeated until an answer is found or there are no more scopes to consult.

## Lookup Failures

When *Engine* exhausts all *lexically available* scopes (moving outward) and still cannot resolve the lookup of an identifier, an error condition then exists. However, depending on the mode of the program (strict-mode or not) and the role of the variable (i.e., *target* vs. *source*; see Chapter 1), this error condition will be handled differently.

## Undefined Mess

If the variable is a *source*, an unresolved identifier lookup is considered an undeclared (unknown, missing) variable, which always results in a `ReferenceError` being thrown. Also, if the variable is a *target*, and the code at that moment is running in strict-mode, the variable is considered undeclared and similarly throws a `ReferenceError`.

The error message for an undeclared variable condition, in most JS environments, will look like, "Reference Error: XYZ is not defined." The phrase "not defined" seems almost identical to the word "undefined," as far as the English language goes. But these two are very different in JS, and this error message unfortunately creates a persistent confusion.

"Not defined" really means "not declared"—or, rather, "undeclared," as in a variable that has no matching formal declaration in any *lexically available* scope. By contrast, "undefined" really means a variable was found (declared), but the variable otherwise has no other value in it at the moment, so it defaults to the `undefined` value.

To perpetuate the confusion even further, JS's `typeof` operator returns the string `"undefined"` for variable references in either state:

```
var studentName;
typeof studentName;      // "undefined"

typeof doesntExist;     // "undefined"
```

These two variable references are in very different conditions, but JS sure does muddy the waters. The terminology mess is confusing and terribly unfortunate. Unfortunately, JS developers just have to pay close attention to not mix up *which kind* of "undefined" they're dealing with!

## Global... What!?

If the variable is a *target* and strict-mode is not in effect, a confusing and surprising legacy behavior kicks in. The troublesome outcome is that the global scope's *Scope Manager* will just create an **accidental global variable** to fulfill that target assignment!

Consider:

```
function getStudentName() {
    // assignment to an undeclared variable :(
    nextStudent = "Suzy";
}

getStudentName();

console.log(nextStudent);
// "Suzy" -- oops, an accidental-global variable!
```

Here's how that *conversation* will proceed:

**Engine:** Hey, *Scope Manager* (for the function), I have a *target* reference for `nextStudent`, ever heard of it?

**(Function) Scope Manager:** Nope, never heard of it. Try the next outer scope.

**Engine:** Hey, *Scope Manager* (for the global scope), I have a *target* reference for `nextStudent`, ever heard of it?

**(Global) Scope Manager:** Nope, but since we're in non-strict-mode, I helped you out and just created a global variable for you, here it is!

Yuck.

This sort of accident (almost certain to lead to bugs eventually) is a great example of the beneficial protections offered by strict-mode, and why it's such a bad idea *not* to be using strict-mode. In strict-mode, the **Global Scope Manager** would instead have responded:

**(Global) Scope Manager:** Nope, never heard of it. Sorry, I've got to throw a `ReferenceError`.

Assigning to a never-declared variable *is* an error, so it's right that we would receive a `ReferenceError` here.

Never rely on accidental global variables. Always use strict-mode, and always formally declare your variables. You'll then get a helpful `ReferenceError` if you ever mistakenly try to assign to a not-declared variable.

## Building On Metaphors

To visualize nested scope resolution, I prefer yet another metaphor, an office building, as in Figure 3:

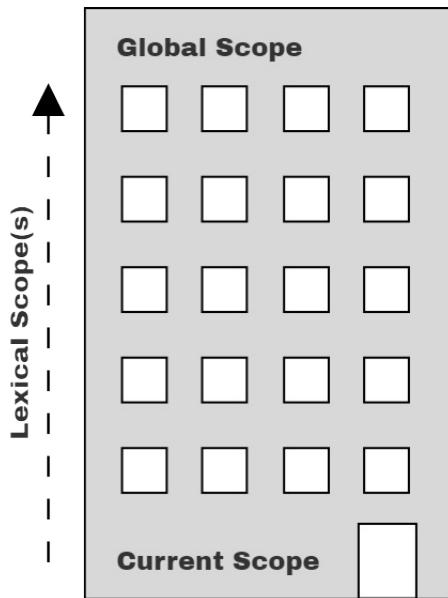


Fig. 3: Scope "Building"

The building represents our program's nested scope collection. The first floor of the building represents the currently executing scope. The top level of the building is the global scope.

You resolve a *target* or *source* variable reference by first looking on the current floor, and if you don't find it, taking the elevator to the next floor (i.e., an outer scope), looking there, then the next, and so on. Once you get to the top floor (the global scope), you either find what you're looking for, or you don't. But you have to stop regardless.

## Continue the Conversation

By this point, you should be developing richer mental models for what scope is and how the JS engine determines and uses it from your code.

Before *continuing*, go find some code in one of your projects and run through these conversations. Seriously, actually speak out loud. Find a friend and practice each role with them. If either of you find yourself confused or tripped up, spend more time reviewing this material.

As we move (up) to the next (outer) chapter, we'll explore how the lexical scopes of a program are connected in a chain.

# You Don't Know JS Yet: Scope & Closures - 2nd Edition

## Chapter 3: The Scope Chain

Chapters 1 and 2 laid down a concrete definition of *lexical scope* (and its parts) and illustrated helpful metaphors for its conceptual foundation. Before proceeding with this chapter, find someone else to explain (written or aloud), in your own words, what lexical scope is and why it's useful to understand.

That seems like a step you might skip, but I've found it really does help to take the time to reformulate these ideas as explanations to others. That helps our brains digest what we're learning!

Now it's time to dig into the nuts and bolts, so expect that things will get a lot more detailed from here forward. Stick with it, though, because these discussions really hammer home just how much we all *don't know* about scope, yet. Make sure to take your time with the text and all the code snippets provided.

To refresh the context of our running example, let's recall the color-coded illustration of the nested scope bubbles, from Chapter 2, Figure 2:

```

1  var students = [
2      { id: 14, name: "Kyle" },
3      { id: 73, name: "Suzy" },
4      { id: 112, name: "Frank" },
5      { id: 6, name: "Sarah" }
6  ];
7
8  function getStudentName(studentID) {
9      for (let student of students) {
10          if (student.id == studentID) {
11              return student.name;
12          }
13      }
14  }
15
16  var nextStudent = getStudentName(73);
17
18  console.log(nextStudent);
19  // "Suzy"

```

Fig. 2 (Ch. 2): Colored Scope Bubbles

The connections between scopes that are nested within other scopes is called the scope chain, which determines the path along which variables can be accessed. The chain is directed, meaning the lookup moves upward/outward only.

## "Lookup" Is (Mostly) Conceptual

In Figure 2, notice the color of the `students` variable reference in the `for`-loop. How exactly did we determine that it's a RED(1) marble?

In Chapter 2, we described the runtime access of a variable as a "lookup," where the *Engine* has to start by asking the current scope's *Scope Manager* if it knows about an identifier/variable, and proceeding upward/outward back through the chain of nested scopes (toward the global scope) until found, if ever. The lookup stops as soon as the first matching named declaration in a scope bucket is found.

The lookup process thus determined that `students` is a RED(1) marble, because we had not yet found a matching variable name as we traversed the scope chain, until we arrived at the final RED(1) global scope.

Similarly, `studentID` in the `if`-statement is determined to be a BLUE(2) marble.

This suggestion of a runtime lookup process works well for conceptual understanding, but it's not actually how things usually work in practice.

The color of a marble's bucket (aka, meta information of what scope a variable originates from) is *usually determined* during the initial compilation processing. Because lexical scope is pretty much finalized at that point, a marble's color will not change based on anything that can happen later during runtime.

Since the marble's color is known from compilation, and it's immutable, this information would likely be stored with (or at least accessible from) each variable's entry in the AST; that information is then used explicitly by the executable instructions that constitute the program's runtime.

In other words, *Engine* (from Chapter 2) doesn't need to lookup through a bunch of scopes to figure out which scope bucket a variable comes from. That information is already known! Avoiding the need for a runtime lookup is a key optimization benefit of lexical scope. The runtime operates more performantly without spending time on all these lookups.

But I said "...usually determined..." just a moment ago, with respect to figuring out a marble's color during compilation. So in what case would it ever *not* be known during compilation?

Consider a reference to a variable that isn't declared in any lexically available scopes in the current file—see *Get Started*, Chapter 1, which asserts that each file is its own separate program from the perspective of JS compilation. If no declaration is found, that's not *necessarily* an error. Another file (program) in the runtime may indeed declare that variable in the shared global scope.

So the ultimate determination of whether the variable was ever appropriately declared in some accessible bucket may need to be deferred to the runtime.

Any reference to a variable that's initially *undeclared* is left as an uncolored marble during that file's compilation; this color cannot be determined until other relevant file(s) have been compiled and the application runtime commences. That deferred lookup will eventually resolve the color to whichever scope the variable is found in (likely the global scope).

However, this lookup would only be needed once per variable at most, since nothing else during runtime could later change that marble's color.

The "Lookup Failures" section in Chapter 2 covers what happens if a marble is ultimately still uncolored at the moment its reference is runtime executed.

## Shadowing

"Shadowing" might sound mysterious and a little bit sketchy. But don't worry, it's completely legit!

Our running example for these chapters uses different variable names across the scope boundaries. Since they all have unique names, in a way it wouldn't matter if all of them were just stored in one bucket (like RED(1)).

Where having different lexical scope buckets starts to matter more is when you have two or more variables, each in different scopes, with the same lexical names. A single scope cannot have two or more variables with the same name; such multiple references would be assumed as just one variable.

So if you need to maintain two or more variables of the same name, you must use separate (often nested) scopes. And in that case, it's very relevant how the different scope buckets are laid out.

Consider:

```
var studentName = "Suzy";

function printStudent(studentName) {
    studentName = studentName.toUpperCase();
    console.log(studentName);
}

printStudent("Frank");
// FRANK

printStudent(studentName);
// SUZY

console.log(studentName);
// Suzy
```

#### TIP:

Before you move on, take some time to analyze this code using the various techniques/metaphors we've covered in the book. In particular, make sure to identify the marble/bubble colors in this snippet. It's good practice!

The `studentName` variable on line 1 (the `var studentName = ...` statement) creates a RED(1) marble. The same named variable is declared as a BLUE(2) marble on line 3, the parameter in the `printStudent(...)` function definition.

What color marble will `studentName` be in the `studentName = studentName.toUpperCase()` assignment statement and the `console.log(studentName)` statement? All three `studentName` references will be BLUE(2).

With the conceptual notion of the "lookup," we asserted that it starts with the current scope and works its way outward/upward, stopping as soon as a matching variable is found. The BLUE(2) `studentName` is found right away. The RED(1) `studentName` is never even considered.

This is a key aspect of lexical scope behavior, called *shadowing*. The BLUE(2) `studentName` variable (parameter) shadows the RED(1) `studentName`. So, the parameter is shadowing the (shadowed) global variable. Repeat that sentence to yourself a few times to make sure you have the terminology straight!

That's why the re-assignment of `studentName` affects only the inner (parameter) variable: the BLUE(2) `studentName`, not the global RED(1) `studentName`.

When you choose to shadow a variable from an outer scope, one direct impact is that from that scope inward/downward (through any nested scopes) it's now impossible for any marble to be colored as the shadowed variable—(RED(1), in this case). In other words, any `studentName` identifier reference will correspond to that parameter variable, never the global `studentName` variable. It's lexically impossible to reference the global `studentName` anywhere inside of the `printStudent(...)` function (or from any nested scopes).

## Global Unshadowing Trick

Please beware: leveraging the technique I'm about to describe is not very good practice, as it's limited in utility, confusing for readers of your code, and likely to invite bugs to your program. I'm covering it only because you may run across this behavior in existing programs, and understanding what's happening is critical to not getting tripped up.

It is possible to access a global variable from a scope where that variable has been shadowed, but not through a typical lexical identifier reference.

In the global scope (RED(1)), `var` declarations and `function` declarations also expose themselves as properties (of the same name as the identifier) on the *global object*—essentially an object representation of the global scope. If you've written JS for a browser environment, you probably recognize the global object as `window`. That's not *entirely* accurate, but it's good enough for our discussion. In the next chapter, we'll explore the global scope/object topic more.

Consider this program, specifically executed as a standalone .js file in a browser environment:

```
var studentName = "Suzy";

function printStudent(studentName) {
    console.log(studentName);
    console.log(window.studentName);
}

printStudent("Frank");
// "Frank"
// "Suzy"
```

Notice the `window.studentName` reference? This expression is accessing the global variable `studentName` as a property on `window` (which we're pretending for now is synonymous with the global object). That's the only way to access a shadowed variable from inside a scope where the shadowing variable is present.

The `window.studentName` is a mirror of the global `studentName` variable, not a separate snapshot copy. Changes to one are still seen from the other, in either direction. You can think of `window.studentName` as a getter/setter that accesses the actual `studentName` variable. As a matter of fact, you can even *add* a variable to the global scope by creating/setting a property on the global object.

#### **WARNING:**

Remember: just because you *can* doesn't mean you *should*. Don't shadow a global variable that you need to access, and conversely, avoid using this trick to access a global variable that you've shadowed. And definitely don't confuse readers of your code by creating global variables as `window` properties instead of with formal declarations!

This little "trick" only works for accessing a global scope variable (not a shadowed variable from a nested scope), and even then, only one that was declared with `var` or `function`.

Other forms of global scope declarations do not create mirrored global object properties:

```
var one = 1;
let notOne = 2;
const notTwo = 3;
class notThree {}

console.log(window.one);      // 1
console.log(window.notOne);   // undefined
console.log(window.notTwo);   // undefined
console.log(window.notThree); // undefined
```

Variables (no matter how they're declared!) that exist in any other scope than the global scope are completely inaccessible from a scope where they've been shadowed:

```
var special = 42;
```

```

function lookingFor(special) {
  // The identifier `special` (parameter) in this
  // scope is shadowed inside keepLooking(), and
  // is thus inaccessible from that scope.

  function keepLooking() {
    var special = 3.141592;
    console.log(special);
    console.log(window.special);
  }

  keepLooking();
}

lookingFor(112358132134);
// 3.141592
// 42

```

The global RED(1) `special` is shadowed by the BLUE(2) `special` (parameter), and the BLUE(2) `special` is itself shadowed by the GREEN(3) `special` inside `keepLooking()`. We can still access the RED(1) `special` using the indirect reference `window.special`. But there's no way for `keepLooking()` to access the BLUE(2) `special` that holds the number `112358132134`.

## Copying Is Not Accessing

I've been asked the following "But what about...?" question dozens of times. Consider:

```

var special = 42;

function lookingFor(special) {
  var another = {
    special: special
};

  function keepLooking() {
    var special = 3.141592;
    console.log(special);
    console.log(another.special); // Ooo, tricky!
    console.log(window.special);
  }

  keepLooking();
}

lookingFor(112358132134);
// 3.141592
// 112358132134
// 42

```

Oh! So does this `another` object technique disprove my claim that the `special` parameter is "completely inaccessible" from inside `keepLooking()`? No, the claim is still correct.

`special: special` is copying the value of the `special` parameter variable into another container (a property of the same name). Of course, if you put a value in another container, shadowing no longer applies (unless `another` was shadowed, too!). But that doesn't mean we're accessing the parameter `special`; it means we're accessing the copy of the value it had at that moment, by way of `another` container (object property). We cannot reassign the BLUE(2) `special` parameter to a different value from inside `keepLooking()`.

Another "But...!?" you may be about to raise: what if I'd used objects or arrays as the values instead of the numbers (`112358132134`, etc.)? Would us having references to objects instead of copies of primitive values "fix" the inaccessibility?

No. Mutating the contents of the object value via a reference copy is **not** the same thing as lexically accessing the variable itself. We still can't reassign the BLUE(2) `special` parameter.

## Illegal Shadowing

Not all combinations of declaration shadowing are allowed. `let` can shadow `var`, but `var` cannot shadow `let`:

```
function something() {
  var special = "JavaScript";

  {
    let special = 42; // totally fine shadowing
    // ...
  }
}

function another() {
  // ...

  {
    let special = "JavaScript";

    {
      var special = "JavaScript";
      // ^^^ Syntax Error
      // ...
    }
  }
}
```

Notice in the `another()` function, the inner `var special` declaration is attempting to declare a function-wide `special`, which in and of itself is fine (as shown by the `something()` function).

The syntax error description in this case indicates that `special` has already been defined, but that error message is a little misleading—again, no such error happens in `something()`, as shadowing is generally allowed just fine.

The real reason it's raised as a `SyntaxError` is because the `var` is basically trying to "cross the boundary" of (or hop over) the `let` declaration of the same name, which is not allowed.

That boundary-crossing prohibition effectively stops at each function boundary, so this variant raises no exception:

```
function another() {
  // ...

  {
    let special = "JavaScript";

    ajax("https://some.url", function callback(){
      // totally fine shadowing
      var special = "JavaScript";
      // ...
    });
  }
}
```

Summary: `let` (in an inner scope) can always shadow an outer scope's `var`. `var` (in an inner scope) can only shadow an outer scope's `let` if there is a function boundary in between.

## Function Name Scope

As you've seen by now, a `function` declaration looks like this:

```
function askQuestion() {
  // ...
}
```

And as discussed in Chapters 1 and 2, such a `function` declaration will create an identifier in the enclosing scope (in this case, the global scope) named `askQuestion`.

What about this program?

```
var askQuestion = function(){
  // ...
};
```

The same is true for the variable `askQuestion` being created. But since it's a `function` expression—a function definition used as value instead of a standalone declaration—the function itself will not "hoist" (see Chapter 5).

One major difference between `function` declarations and `function` expressions is what happens to the name identifier of the function. Consider a named `function` expression:

```
var askQuestion = function ofTheTeacher(){
  // ...
};
```

We know `askQuestion` ends up in the outer scope. But what about the `ofTheTeacher` identifier? For formal `function` declarations, the name identifier ends up in the outer/enclosing scope, so it may be reasonable to assume that's the case here. But `ofTheTeacher` is declared as an identifier **inside the function itself**:

```
var askQuestion = function ofTheTeacher() {
  console.log(ofTheTeacher);
};

askQuestion();
// function ofTheTeacher()...

console.log(ofTheTeacher);
// ReferenceError: ofTheTeacher is not defined
```

### NOTE:

Actually, `ofTheTeacher` is not exactly *in the scope of the function*. Appendix A, "Implied Scopes" will explain further.

Not only is `ofTheTeacher` declared inside the function rather than outside, but it's also defined as read-only:

```
var askQuestion = function ofTheTeacher() {
  "use strict";
  ofTheTeacher = 42; // TypeError

  //...
};

askQuestion();
// TypeError
```

Because we used strict-mode, the assignment failure is reported as a `TypeError`; in non-strict-mode, such an assignment fails silently with no exception.

What about when a `function` expression has no name identifier?

```
var askQuestion = function(){
  // ...
};
```

A `function` expression with a name identifier is referred to as a "named function expression," but one without a name identifier is referred to as an "anonymous function expression." Anonymous function expressions clearly have no name identifier that affects either scope.

#### NOTE:

We'll discuss named vs. anonymous `function` expressions in much more detail, including what factors affect the decision to use one or the other, in Appendix A.

## Arrow Functions

ES6 added an additional `function` expression form to the language, called "arrow functions":

```
var askQuestion = () => {
  // ...
};
```

The `=>` arrow function doesn't require the word `function` to define it. Also, the `( ... )` around the parameter list is optional in some simple cases. Likewise, the `{ ... }` around the function body is optional in some cases. And when the `{ ... }` are omitted, a return value is sent out without using a `return` keyword.

#### NOTE:

The attractiveness of `=>` arrow functions is often sold as "shorter syntax," and that's claimed to equate to objectively more readable code. This claim is dubious at best, and I believe outright misguided. We'll dig into the "readability" of various function forms in Appendix A.

Arrow functions are lexically anonymous, meaning they have no directly related identifier that references the function. The assignment to `askQuestion` creates an inferred name of "askQuestion", but that's **not the same thing as being non-anonymous**:

```
var askQuestion = () => {
  // ...
};

askQuestion.name; // askQuestion
```

Arrow functions achieve their syntactic brevity at the expense of having to mentally juggle a bunch of variations for different forms/conditions. Just a few, for example:

```
() => 42;

id => id.toUpperCase();

(id, name) => ({ id, name });

(...args) => {
  return args[args.length - 1];
};
```

The real reason I bring up arrow functions is because of the common but incorrect claim that arrow functions somehow behave differently with respect to lexical scope from standard `function` functions.

This is incorrect.

Other than being anonymous (and having no declarative form), => arrow functions have the same lexical scope rules as `function` functions do. An arrow function, with or without `{ ... }` around its body, still creates a separate, inner nested bucket of scope. Variable declarations inside this nested scope bucket behave the same as in a `function` scope.

## Backing Out

When a function (declaration or expression) is defined, a new scope is created. The positioning of scopes nested inside one another creates a natural scope hierarchy throughout the program, called the scope chain. The scope chain controls variable access, directionally oriented upward and outward.

Each new scope offers a clean slate, a space to hold its own set of variables. When a variable name is repeated at different levels of the scope chain, shadowing occurs, which prevents access to the outer variable from that point inward.

As we step back out from these finer details, the next chapter shifts focus to the primary scope all JS programs include: the global scope.

# You Don't Know JS Yet: Scope & Closures - 2nd Edition

## Chapter 4: Around the Global Scope

Chapter 3 mentioned the "global scope" several times, but you may still be wondering why a program's outermost scope is all that important in modern JS. The vast majority of work is now done inside of functions and modules rather than globally.

Is it good enough to just assert, "Avoid using the global scope," and be done with it?

The global scope of a JS program is a rich topic, with much more utility and nuance than you would likely assume. This chapter first explores how the global scope is (still) useful and relevant to writing JS programs today, then looks at differences in where and *how to access* the global scope in different JS environments.

Fully understanding the global scope is critical in your mastery of using lexical scope to structure your programs.

### Why Global Scope?

It's likely no surprise to readers that most applications are composed of multiple (sometimes many!) individual JS files. So how exactly do all those separate files get stitched together in a single runtime context by the JS engine?

With respect to browser-executed applications, there are three main ways.

First, if you're directly using ES modules (not transpiling them into some other module-bundle format), these files are loaded individually by the JS environment. Each module then `import`s references to whichever other modules it needs to access. The separate module files cooperate with each other exclusively through these shared imports, without needing any shared outer scope.

Second, if you're using a bundler in your build process, all the files are typically concatenated together before delivery to the browser and JS engine, which then only processes one big file. Even with all the pieces of the application co-located in a single file, some mechanism is necessary for each piece to register a *name* to be referred to by other pieces, as well as some facility for that access to occur.

In some build setups, the entire contents of the file are wrapped in a single enclosing scope, such as a wrapper function, universal module (UMD—see Appendix A), etc. Each piece can register itself for access from other pieces by way of local variables in that shared scope. For example:

```
(function wrappingOuterScope(){
  var moduleOne = (function one(){
    // ...
  })();

  var moduleTwo = (function two(){
    // ...

    function callModuleOne() {
      moduleOne.someMethod();
    }

    // ...
  })();
})();
```

As shown, the `moduleOne` and `moduleTwo` local variables inside the `wrappingOuterScope()` function scope are declared so that these modules can access each other for their cooperation.

While the scope of `wrappingOuterScope()` is a function and not the full environment global scope, it does act as a sort of "application-wide scope," a bucket where all the top-level identifiers can be stored, though not in the real global scope. It's kind of like a stand-in for the global scope in that respect.

And finally, the third way: whether a bundler tool is used for an application, or whether the (non-ES module) files are simply loaded in the browser individually (via `<script>` tags or other dynamic JS resource loading), if there is no single surrounding scope encompassing all these pieces, the **global scope** is the only way for them to cooperate with each other:

A bundled file of this sort often looks something like this:

```
var moduleOne = (function one(){
    // ...
})();
var moduleTwo = (function two(){
    // ...

    function callModuleOne() {
        moduleOne.someMethod();
    }

    // ...
})();
```

Here, since there is no surrounding function scope, these `moduleOne` and `moduleTwo` declarations are simply dropped into the global scope. This is effectively the same as if the files hadn't been concatenated, but loaded separately:

module1.js:

```
var moduleOne = (function one(){
    // ...
})();
```

module2.js:

```
var moduleTwo = (function two(){
    // ...

    function callModuleOne() {
        moduleOne.someMethod();
    }

    // ...
})();
```

If these files are loaded separately as normal standalone .js files in a browser environment, each top-level variable declaration will end up as a global variable, since the global scope is the only shared resource between these two separate files—they're independent programs, from the perspective of the JS engine.

In addition to (potentially) accounting for where an application's code resides during runtime, and how each piece is able to access the other pieces to cooperate, the global scope is also where:

- JS exposes its built-ins:
  - primitives: `undefined`, `null`, `Infinity`, `Nan`
  - natives: `Date()`, `Object()`, `String()`, etc.
  - global functions: `eval()`, `parseInt()`, etc.
  - namespaces: `Math`, `Atomics`, `JSON`
  - friends of JS: `Intl`, `WebAssembly`
- The environment hosting the JS engine exposes its own built-ins:

- `console` (and its methods)
- the DOM (`window`, `document`, etc)
- timers (`setTimeout(..)`, etc)
- web platform APIs: `navigator`, `history`, geolocation, WebRTC, etc.

These are just some of the many *globals* your programs will interact with.

**NOTE:**

Node also exposes several elements "globally," but they're technically not in the `global` scope: `require()`, `__dirname`, `module`, `URL`, and so on.

Most developers agree that the global scope shouldn't just be a dumping ground for every variable in your application. That's a mess of bugs just waiting to happen. But it's also undeniable that the global scope is an important *glue* for practically every JS application.

## Where Exactly is this Global Scope?

It might seem obvious that the global scope is located in the outermost portion of a file; that is, not inside any function or other block. But it's not quite as simple as that.

Different JS environments handle the scopes of your programs, especially the global scope, differently. It's quite common for JS developers to harbor misconceptions without even realizing it.

### Browser "Window"

With respect to treatment of the global scope, the most *pure* environment JS can be run in is as a standalone .js file loaded in a web page environment in a browser. I don't mean "pure" as in nothing automatically added—lots may be added!—but rather in terms of minimal intrusion on the code or interference with its expected global scope behavior.

Consider this .js file:

```
var studentName = "Kyle";

function hello() {
  console.log(`Hello, ${studentName}!`);
}

hello();
// Hello, Kyle!
```

This code may be loaded in a web page environment using an inline `<script>` tag, a `<script src=..>` script tag in the markup, or even a dynamically created `<script>` DOM element. In all three cases, the `studentName` and `hello` identifiers are declared in the global scope.

That means if you access the global object (commonly, `window` in the browser), you'll find properties of those same names there:

```
var studentName = "Kyle";

function hello() {
  console.log(`Hello, ${window.studentName}!`);
}

window.hello();
// Hello, Kyle!
```

That's the default behavior one would expect from a reading of the JS specification: the outer scope *is* the global scope and `studentName` is legitimately created as global variable.

That's what I mean by *pure*. But unfortunately, that won't always be true of all JS environments you encounter, and that's often surprising to JS developers.

## Globals Shadowing Globals

Recall the discussion of shadowing (and global unshadowing) from Chapter 3, where one variable declaration can override and prevent access to a declaration of the same name from an outer scope.

An unusual consequence of the difference between a global variable and a global property of the same name is that, within just the global scope itself, a global object property can be shadowed by a global variable:

```
window.something = 42;

let something = "Kyle";

console.log(something);
// Kyle

console.log(window.something);
// 42
```

The `let` declaration adds a `something` global variable but not a global object property (see Chapter 3). The effect then is that the `something` lexical identifier shadows the `something` global object property.

It's almost certainly a bad idea to create a divergence between the global object and the global scope. Readers of your code will almost certainly be tripped up.

A simple way to avoid this gotcha with global declarations: always use `var` for globals. Reserve `let` and `const` for block scopes (see "Scoping with Blocks" in Chapter 6).

## DOM Globals

I asserted that a browser-hosted JS environment has the most *pure* global scope behavior we'll see. However, it's not entirely *pure*.

One surprising behavior in the global scope you may encounter with browser-based JS applications: a DOM element with an `id` attribute automatically creates a global variable that references it.

Consider this markup:

```
<ul id="my-todo-list">
  <li id="first">Write a book</li>
  ..
</ul>
```

And the JS for that page could include:

```
first;
// <li id="first">..</li>

window["my-todo-list"];
// <ul id="my-todo-list">..</ul>
```

If the `id` value is a valid lexical name (like `first`), the lexical variable is created. If not, the only way to access that global is through the global object (`window[...]`).

The auto-registration of all `id`-bearing DOM elements as global variables is an old legacy browser behavior that nevertheless must remain because so many old sites still rely on it. My advice is never to use these global variables, even though they will always be silently created.

## What's in a (Window) Name?

Another global scope oddity in browser-based JS:

```
var name = 42;

console.log(name, typeof name);
// "42" string
```

`window.name` is a pre-defined "global" in a browser context; it's a property on the global object, so it seems like a normal global variable (yet it's anything but "normal").

We used `var` for our declaration, which **does not** shadow the pre-defined `name` global property. That means, effectively, the `var` declaration is ignored, since there's already a global scope object property of that name. As we discussed earlier, had we used `let name`, we would have shadowed `window.name` with a separate global `name` variable.

But the truly surprising behavior is that even though we assigned the number `42` to `name` (and thus `window.name`), when we then retrieve its value, it's a string `"42"`! In this case, the weirdness is because `name` is actually a pre-defined getter/setter on the `window` object, which insists on its value being a string value. Yikes!

With the exception of some rare corner cases like DOM element ID's and `window.name`, JS running as a standalone file in a browser page has some of the most *pure* global scope behavior we will encounter.

## Web Workers

Web Workers are a web platform extension on top of browser-JS behavior, which allows a JS file to run in a completely separate thread (operating system wise) from the thread that's running the main JS program.

Since these Web Worker programs run on a separate thread, they're restricted in their communications with the main application thread, to avoid/limit race conditions and other complications. Web Worker code does not have access to the DOM, for example. Some web APIs are, however, made available to the worker, such as `navigator`.

Since a Web Worker is treated as a wholly separate program, it does not share the global scope with the main JS program. However, the browser's JS engine is still running the code, so we can expect similar *purity* of its global scope behavior. Since there is no DOM access, the `window` alias for the global scope doesn't exist.

In a Web Worker, the global object reference is typically made using `self`:

```
var studentName = "Kyle";
let studentID = 42;

function hello() {
  console.log(`Hello, ${self.studentName}!`);
}

self.hello();
// Hello, Kyle!

self.studentID;
// undefined
```

Just as with main JS programs, `var` and `function` declarations create mirrored properties on the global object (aka, `self`), where other declarations (`let`, etc) do not.

So again, the global scope behavior we're seeing here is about as *pure* as it gets for running JS programs; perhaps it's even more *pure* since there's no DOM to muck things up!

## Developer Tools Console/REPL

Recall from Chapter 1 in *Get Started* that Developer Tools don't create a completely adherent JS environment. They do process JS code, but they also lean in favor of the UX interaction being most friendly to developers (aka, developer experience, or DX).

In some cases, favoring DX when typing in short JS snippets, over the normal strict steps expected for processing a full JS program, produces observable differences in code behavior between programs and tools. For example, certain error conditions applicable to a JS program may be relaxed and not displayed when the code is entered into a developer tool.

With respect to our discussions here about scope, such observable differences in behavior may include:

- The behavior of the global scope
- Hoisting (see Chapter 5)
- Block-scoping declarators (`let` / `const`, see Chapter 6) when used in the outermost scope

Although it might seem, while using the console/REPL, that statements entered in the outermost scope are being processed in the real global scope, that's not quite accurate. Such tools typically emulate the global scope position to an extent; it's emulation, not strict adherence. These tool environments prioritize developer convenience, which means that at times (such as with our current discussions regarding scope), observed behavior may deviate from the JS specification.

The take-away is that Developer Tools, while optimized to be convenient and useful for a variety of developer activities, are **not** suitable environments to determine or verify explicit and nuanced behaviors of an actual JS program context.

## ES Modules (ESM)

ES6 introduced first-class support for the module pattern (covered in Chapter 8). One of the most obvious impacts of using ESM is how it changes the behavior of the observably top-level scope in a file.

Recall this code snippet from earlier (which we'll adjust to ESM format by using the `export` keyword):

```
var studentName = "Kyle";

function hello() {
    console.log(`Hello, ${studentName}!`);
}

hello();
// Hello, Kyle!

export hello;
```

If that code is in a file that's loaded as an ES module, it will still run exactly the same. However, the observable effects, from the overall application perspective, will be different.

Despite being declared at the top level of the (module) file, in the outermost obvious scope, `studentName` and `hello` are not global variables. Instead, they are module-wide, or if you prefer, "module-global."

However, in a module there's no implicit "module-wide scope object" for these top-level declarations to be added to as properties, as there is when declarations appear in the top-level of non-module JS files. This is not to say that global variables cannot exist or be accessed in such programs. It's just that global variables don't get *created* by declaring variables in the top-level scope of a module.

The module's top-level scope is descended from the global scope, almost as if the entire contents of the module were wrapped in a function. Thus, all variables that exist in the global scope (whether they're on the global object or not!) are available as lexical identifiers from inside the module's scope.

ESM encourages a minimization of reliance on the global scope, where you import whatever modules you may need for the current module to operate. As such, you less often see usage of the global scope or its global object.

However, as noted earlier, there are still plenty of JS and web globals that you will continue to access from the global scope, whether you realize it or not!

## Node

One aspect of Node that often catches JS developers off-guard is that Node treats every single .js file that it loads, including the main one you start the Node process with, as a *module* (ES module or CommonJS module, see Chapter 8). The practical effect is that the top level of your Node programs **is never actually the global scope**, the way it is when loading a non-module file in the browser.

As of time of this writing, Node has recently added support for ES modules. But additionally, Node has from its beginning supported a module format referred to as "CommonJS", which looks like this:

```
var studentName = "Kyle";

function hello() {
    console.log(`Hello, ${studentName}!`);
}

hello();
// Hello, Kyle!

module.exports.hello = hello;
```

Before processing, Node effectively wraps such code in a function, so that the `var` and `function` declarations are contained in that wrapping function's scope, **not** treated as global variables.

Envision the preceding code as being seen by Node as this (illustrative, not actual):

```
function Module(module, require, __dirname, ...) {
    var studentName = "Kyle";

    function hello() {
        console.log(`Hello, ${studentName}!`);
    }

    hello();
    // Hello, Kyle!

    module.exports.hello = hello;
}
```

Node then essentially invokes the added `Module(...)` function to run your module. You can clearly see here why `studentName` and `hello` identifiers are not global, but rather declared in the module scope.

As noted earlier, Node defines a number of "globals" like `require()`, but they're not actually identifiers in the global scope (nor properties of the global object). They're injected in the scope of every module, essentially a bit like the parameters listed in the `Module(...)` function declaration.

So how do you define actual global variables in Node? The only way to do so is to add properties to another of Node's automatically provided "globals," which is ironically called `global`. `global` is a reference to the real global scope object, somewhat like using `window` in a browser JS environment.

Consider:

```
global.studentName = "Kyle";

function hello() {
  console.log(`Hello, ${ studentName }!`);
}

hello();
// Hello, Kyle!

module.exports.hello = hello;
```

Here we add `studentName` as a property on the `global` object, and then in the `console.log(..)` statement we're able to access `studentName` as a normal global variable.

Remember, the identifier `global` is not defined by JS; it's specifically defined by Node.

## Global This

Reviewing the JS environments we've looked at so far, a program may or may not:

- Declare a global variable in the top-level scope with `var` or `function` declarations—or `let`, `const`, and `class`.
- Also add global variables declarations as properties of the global scope object if `var` or `function` are used for the declaration.
- Refer to the global scope object (for adding or retrieving global variables, as properties) with `window`, `self`, or `global`.

I think it's fair to say that global scope access and behavior is more complicated than most developers assume, as the preceding sections have illustrated. But the complexity is never more obvious than in trying to nail down a universally applicable reference to the global scope object.

Yet another "trick" for obtaining a reference to the global scope object looks like:

```
const theGlobalScopeObject =
  (new Function("return this"))();
```

### NOTE:

A function can be dynamically constructed from code stored in a string value with the `Function()` constructor, similar to `eval(..)` (see "Cheating: Runtime Scope Modifications" in Chapter 1). Such a function will automatically be run in non-strict-mode (for legacy reasons) when invoked with the normal `()` function invocation as shown; its `this` will point at the global object. See the third book in the series, *Objects & Classes*, for more information on determining `this` bindings.

So, we have `window`, `self`, `global`, and this ugly `new Function(..)` trick. That's a lot of different ways to try to get at this global object. Each has its pros and cons.

Why not introduce yet another!?!?

As of ES2020, JS has finally defined a standardized reference to the global scope object, called `globalThis`. So, subject to the recency of the JS engines your code runs in, you can use `globalThis` in place of any of those other approaches.

We could even attempt to define a cross-environment polyfill that's safer across pre- `globalThis` JS environments, such as:

```
const theGlobalScopeObject =
  (typeof globalThis != "undefined") ? globalThis :
  (typeof global != "undefined") ? global :
  (typeof window != "undefined") ? window :
  (typeof self != "undefined") ? self :
```

```
(new Function("return this"))();
```

Phew! That's certainly not ideal, but it works if you find yourself needing a reliable global scope reference.

(The proposed name `globalThis` was fairly controversial while the feature was being added to JS. Specifically, I and many others felt the "this" reference in its name was misleading, since the reason you reference this object is to access to the global scope, never to access some sort of global/default `this` binding. There were many other names considered, but for a variety of reasons ruled out. Unfortunately, the name chosen ended up as a last resort. If you plan to interact with the global scope object in your programs, to reduce confusion, I strongly recommend choosing a better name, such as (the laughably long but accurate!) `theGlobalScopeObject` used here.)

## Globally Aware

The global scope is present and relevant in every JS program, even though modern patterns for organizing code into modules de-emphasizes much of the reliance on storing identifiers in that namespace.

Still, as our code proliferates more and more beyond the confines of the browser, it's especially important we have a solid grasp on the differences in how the global scope (and global scope object!) behave across different JS environments.

With the big picture of global scope now sharper in focus, the next chapter again descends into the deeper details of lexical scope, examining how and when variables can be used.

# You Don't Know JS Yet: Scope & Closures - 2nd Edition

## Chapter 5: The (Not So) Secret Lifecycle of Variables

By now you should have a decent grasp of the nesting of scopes, from the global scope downward—called a program's scope chain.

But just knowing which scope a variable comes from is only part of the story. If a variable declaration appears past the first statement of a scope, how will any references to that identifier *before* the declaration behave? What happens if you try to declare the same variable twice in a scope?

JS's particular flavor of lexical scope is rich with nuance in how and when variables come into existence and become available to the program.

### When Can I Use a Variable?

At what point does a variable become available to use within its scope? There may seem to be an obvious answer: *after* the variable has been declared/created. Right? Not quite.

Consider:

```
greeting();
// Hello!

function greeting() {
  console.log("Hello!");
}
```

This code works fine. You may have seen or even written code like it before. But did you ever wonder how or why it works? Specifically, why can you access the identifier `greeting` from line 1 (to retrieve and execute a function reference), even though the `greeting()` function declaration doesn't occur until line 4?

Recall Chapter 1 points out that all identifiers are registered to their respective scopes during compile time. Moreover, every identifier is *created* at the beginning of the scope it belongs to, **every time that scope is entered**.

The term most commonly used for a variable being visible from the beginning of its enclosing scope, even though its declaration may appear further down in the scope, is called **hoisting**.

But hoisting alone doesn't fully answer the question. We can see an identifier called `greeting` from the beginning of the scope, but why can we **call** the `greeting()` function before it's been declared?

In other words, how does the variable `greeting` have any value (the function reference) assigned to it, from the moment the scope starts running? The answer is a special characteristic of formal `function` declarations, called *function hoisting*. When a `function` declaration's name identifier is registered at the top of its scope, it's additionally auto-initialized to that function's reference. That's why the function can be called throughout the entire scope!

One key detail is that both *function hoisting* and `var`-flavored *variable hoisting* attach their name identifiers to the nearest enclosing **function scope** (or, if none, the global scope), not a block scope.

#### NOTE:

Declarations with `let` and `const` still hoist (see the TDZ discussion later in this chapter). But these two declaration forms attach to their enclosing block rather than just an enclosing function as with `var` and `function` declarations. See "Scoping with Blocks" in Chapter 6 for more information.

## Hoisting: Declaration vs. Expression

*Function hoisting* only applies to formal `function` declarations (specifically those which appear outside of blocks—see "FiB" in Chapter 6), not to `function` expression assignments. Consider:

```
greeting();
// TypeError

var greeting = function greeting() {
  console.log("Hello!");
}
```

Line 1 (`greeting();`) throws an error. But the *kind* of error thrown is very important to notice. A `TypeError` means we're trying to do something with a value that is not allowed. Depending on your JS environment, the error message would say something like, "'undefined' is not a function," or more helpfully, "'greeting' is not a function."

Notice that the error is **not** a `ReferenceError`. JS isn't telling us that it couldn't find `greeting` as an identifier in the scope. It's telling us that `greeting` was found but doesn't hold a function reference at that moment. Only functions can be invoked, so attempting to invoke some non-function value results in an error.

But what does `greeting` hold, if not the function reference?

In addition to being hoisted, variables declared with `var` are also automatically initialized to `undefined` at the beginning of their scope—again, the nearest enclosing function, or the global. Once initialized, they're available to be used (assigned to, retrieved from, etc.) throughout the whole scope.

So on that first line, `greeting` exists, but it holds only the default `undefined` value. It's not until line 4 that `greeting` gets assigned the function reference.

Pay close attention to the distinction here. A `function` declaration is hoisted **and initialized to its function value** (again, called *function hoisting*). A `var` variable is also hoisted, and then auto-initialized to `undefined`. Any subsequent `function` expression assignments to that variable don't happen until that assignment is processed during runtime execution.

In both cases, the name of the identifier is hoisted. But the function reference association isn't handled at initialization time (beginning of the scope) unless the identifier was created in a formal `function` declaration.

## Variable Hoisting

Let's look at another example of *variable hoisting*:

```
greeting = "Hello!";
console.log(greeting);
// Hello!

var greeting = "Howdy!";
```

Though `greeting` isn't declared until line 5, it's available to be assigned to as early as line 1. Why?

There's two necessary parts to the explanation:

- the identifier is hoisted,
- **and** it's automatically initialized to the value `undefined` from the top of the scope.

### NOTE:

Using *variable hoisting* of this sort probably feels unnatural, and many readers might rightly want to avoid relying on it in their programs. But should all hoisting (including *function hoisting*) be avoided? We'll explore these different perspectives on hoisting in more detail in Appendix A.

## Hoisting: Yet Another Metaphor

Chapter 2 was full of metaphors (to illustrate scope), but here we are faced with yet another: hoisting itself. Rather than hoisting being a concrete execution step the JS engine performs, it's more useful to think of hoisting as a visualization of various actions JS takes in setting up the program **before execution**.

The typical assertion of what hoisting means: *lifting*—like lifting a heavy weight upward—any identifiers all the way to the top of a scope. The explanation often asserted is that the JS engine will actually *rewrite* that program before execution, so that it looks more like this:

```
var greeting;           // hoisted declaration
greeting = "Hello!";   // the original line 1
console.log(greeting); // Hello!
greeting = "Howdy!";   // `var` is gone!
```

The hoisting (metaphor) proposes that JS pre-processes the original program and re-arranges it a bit, so that all the declarations have been moved to the top of their respective scopes, before execution. Moreover, the hoisting metaphor asserts that `function` declarations are, in their entirety, hoisted to the top of each scope. Consider:

```
studentName = "Suzy";
greeting();
// Hello Suzy!

function greeting() {
  console.log(`Hello ${ studentName }!`);
}
var studentName;
```

The "rule" of the hoisting metaphor is that function declarations are hoisted first, then variables are hoisted immediately after all the functions. Thus, the hoisting story suggests that program is *re-arranged* by the JS engine to look like this:

```
function greeting() {
  console.log(`Hello ${ studentName }!`);
}
var studentName;

studentName = "Suzy";
greeting();
// Hello Suzy!
```

This hoisting metaphor is convenient. Its benefit is allowing us to hand wave over the magical look-ahead pre-processing necessary to find all these declarations buried deep in scopes and somehow move (hoist) them to the top; we can just think about the program as if it's executed by the JS engine in a **single pass**, top-down.

Single-pass definitely seems more straightforward than Chapter 1's assertion of a two-phase processing.

Hoisting as a mechanism for re-ordering code may be an attractive simplification, but it's not accurate. The JS engine doesn't actually re-arrange the code. It can't magically look ahead and find declarations; the only way to accurately find them, as well as all the scope boundaries in the program, would be to fully parse the code.

Guess what parsing is? The first phase of the two-phase processing! There's no magical mental gymnastics that gets around that fact.

So if the hoisting metaphor is (at best) inaccurate, what should we do with the term? I think it's still useful—indeed, even members of TC39 regularly use it!—but I don't think we should claim it's an actual re-arrangement of source code.

**WARNING:**

Incorrect or incomplete mental models often still seem sufficient because they can occasionally lead to accidental right answers. But in the long run it's harder to accurately analyze and predict outcomes if your thinking isn't particularly aligned with how the JS engine works.

I assert that hoisting *should* be used to refer to the **compile-time operation** of generating runtime instructions for the automatic registration of a variable at the beginning of its scope, each time that scope is entered.

That's a subtle but important shift, from hoisting as a runtime behavior to its proper place among compile-time tasks.

## Re-declaration?

What do you think happens when a variable is declared more than once in the same scope? Consider:

```
var studentName = "Frank";
console.log(studentName);
// Frank

var studentName;
console.log(studentName); // ???
```

What do you expect to be printed for that second message? Many believe the second `var studentName` has re-declared the variable (and thus "reset" it), so they expect `undefined` to be printed.

But is there such a thing as a variable being "re-declared" in the same scope? No.

If you consider this program from the perspective of the hoisting metaphor, the code would be re-arranged like this for execution purposes:

```
var studentName;
var studentName; // clearly a pointless no-op!

studentName = "Frank";
console.log(studentName);
// Frank

console.log(studentName);
// Frank
```

Since hoisting is actually about registering a variable at the beginning of a scope, there's nothing to be done in the middle of the scope where the original program actually had the second `var studentName` statement. It's just a no-op(eration), a pointless statement.

**TIP:**

In the style of the conversation narrative from Chapter 2, *Compiler* would find the second `var` declaration statement and ask the *Scope Manager* if it had already seen a `studentName` identifier; since it had, there wouldn't be anything else to do.

It's also important to point out that `var studentName;` doesn't mean `var studentName = undefined;`, as most assume. Let's prove they're different by considering this variation of the program:

```
var studentName = "Frank";
console.log(studentName); // Frank

var studentName;
console.log(studentName); // Frank <--- still!

// let's add the initialization explicitly
```

```
var studentName = undefined;
console.log(studentName); // undefined <--- see!?
```

See how the explicit `= undefined` initialization produces a different outcome than assuming it happens implicitly when omitted? In the next section, we'll revisit this topic of initialization of variables from their declarations.

A repeated `var` declaration of the same identifier name in a scope is effectively a do-nothing operation. Here's another illustration, this time across a function of the same name:

```
var greeting;

function greeting() {
  console.log("Hello!");
}

// basically, a no-op
var greeting;

typeof greeting; // "function"

var greeting = "Hello!";

typeof greeting; // "string"
```

The first `greeting` declaration registers the identifier to the scope, and because it's a `var` the auto-initialization will be `undefined`. The `function` declaration doesn't need to re-register the identifier, but because of *function hoisting* it overrides the auto-initialization to use the function reference. The second `var greeting` by itself doesn't do anything since `greeting` is already an identifier and *function hoisting* already took precedence for the auto-initialization.

Actually assigning `"Hello!"` to `greeting` changes its value from the initial function `greeting()` to the string; `var` itself doesn't have any effect.

What about repeating a declaration within a scope using `let` or `const`?

```
let studentName = "Frank";

console.log(studentName);

let studentName = "Suzy";
```

This program will not execute, but instead immediately throw a `SyntaxError`. Depending on your JS environment, the error message will indicate something like: "studentName has already been declared." In other words, this is a case where attempted "re-declaration" is explicitly not allowed!

It's not just that two declarations involving `let` will throw this error. If either declaration uses `let`, the other can be either `let` or `var`, and the error will still occur, as illustrated with these two variations:

```
var studentName = "Frank";

let studentName = "Suzy";
```

and:

```
let studentName = "Frank";

var studentName = "Suzy";
```

In both cases, a `SyntaxError` is thrown on the *second* declaration. In other words, the only way to "re-declare" a variable is to use `var` for all (two or more) of its declarations.

But why disallow it? The reason for the error is not technical per se, as `var` "re-declaration" has always been allowed; clearly, the same allowance could have been made for `let`.

It's really more of a "social engineering" issue. "Re-declaration" of variables is seen by some, including many on the TC39 body, as a bad habit that can lead to program bugs. So when ES6 introduced `let`, they decided to prevent "re-declaration" with an error.

#### NOTE:

This is of course a stylistic opinion, not really a technical argument. Many developers agree with the position, and that's probably in part why TC39 included the error (as well as `let` conforming to `const`). But a reasonable case could have been made that staying consistent with `var`'s precedent was more prudent, and that such opinion-enforcement was best left to opt-in tooling like linters. In Appendix A, we'll explore whether `var` (and its associated behavior, like "re-declaration") can still be useful in modern JS.

When *Compiler* asks *Scope Manager* about a declaration, if that identifier has already been declared, and if either/both declarations were made with `let`, an error is thrown. The intended signal to the developer is "Stop relying on sloppy re-declaration!"

## Constants?

The `const` keyword is more constrained than `let`. Like `let`, `const` cannot be repeated with the same identifier in the same scope. But there's actually an overriding technical reason why that sort of "re-declaration" is disallowed, unlike `let` which disallows "re-declaration" mostly for stylistic reasons.

The `const` keyword requires a variable to be initialized, so omitting an assignment from the declaration results in a `SyntaxError`:

```
const empty; // SyntaxError
```

`const` declarations create variables that cannot be re-assigned:

```
const studentName = "Frank";
console.log(studentName);
// Frank

studentName = "Suzy"; // TypeError
```

The `studentName` variable cannot be re-assigned because it's declared with a `const`.

#### WARNING:

The error thrown when re-assigning `studentName` is a `TypeError`, not a `SyntaxError`. The subtle distinction here is actually pretty important, but unfortunately far too easy to miss. Syntax errors represent faults in the program that stop it from even starting execution. Type errors represent faults that arise during program execution. In the preceding snippet, "Frank" is printed out before we process the re-assignment of `studentName`, which then throws the error.

So if `const` declarations cannot be re-assigned, and `const` declarations always require assignments, then we have a clear technical reason why `const` must disallow any "re-declarations": any `const` "re-declaration" would also necessarily be a `const` re-assignment, which can't be allowed!

```
const studentName = "Frank";

// obviously this must be an error
const studentName = "Suzy";
```

Since `const` "re-declaration" must be disallowed (on those technical grounds), TC39 essentially felt that `let` "re-declaration" should be disallowed as well, for consistency. It's debatable if this was the best choice, but at least we have the reasoning behind the decision.

## Loops

So it's clear from our previous discussion that JS doesn't really want us to "re-declare" our variables within the same scope. That probably seems like a straightforward admonition, until you consider what it means for repeated execution of declaration statements in loops. Consider:

```
var keepGoing = true;
while (keepGoing) {
    let value = Math.random();
    if (value > 0.5) {
        keepGoing = false;
    }
}
```

Is `value` being "re-declared" repeatedly in this program? Will we get errors thrown? No.

All the rules of scope (including "re-declaration" of `let`-created variables) are applied *per scope instance*. In other words, each time a scope is entered during execution, everything resets.

Each loop iteration is its own new scope instance, and within each scope instance, `value` is only being declared once. So there's no attempted "re-declaration," and thus no error. Before we consider other loop forms, what if the `value` declaration in the previous snippet were changed to a `var`?

```
var keepGoing = true;
while (keepGoing) {
    var value = Math.random();
    if (value > 0.5) {
        keepGoing = false;
    }
}
```

Is `value` being "re-declared" here, especially since we know `var` allows it? No. Because `var` is not treated as a block-scoping declaration (see Chapter 6), it attaches itself to the global scope. So there's just one `value` variable, in the same scope as `keepGoing` (global scope, in this case). No "re-declaration" here, either!

One way to keep this all straight is to remember that `var`, `let`, and `const` keywords are effectively *removed* from the code by the time it starts to execute. They're handled entirely by the compiler.

If you mentally erase the declarator keywords and then try to process the code, it should help you decide if and when (re-)declarations might occur.

What about "re-declaration" with other loop forms, like `for`-loops?

```
for (let i = 0; i < 3; i++) {
    let value = i * 10;
    console.log(`#${i}: ${value}`);
}
// 0: 0
// 1: 10
// 2: 20
```

It should be clear that there's only one `value` declared per scope instance. But what about `i`? Is it being "re-declared"?

To answer that, consider what scope `i` is in. It might seem like it would be in the outer (in this case, global) scope, but it's not. It's in the scope of `for`-loop body, just like `value` is. In fact, you could sorta think about that loop in this more verbose equivalent form:

```
{
  // a fictional variable for illustration
  let $$i = 0;

  for (/* nothing */; $$i < 3; $$i++) {
    // here's our actual loop `i`!
    let i = $$i;

    let value = i * 10;
    console.log(`#${i}: ${value}`);
  }
  // 0: 0
  // 1: 10
  // 2: 20
}
```

Now it should be clear: the `i` and `value` variables are both declared exactly once **per scope instance**. No "re-declaration" here.

What about other `for`-loop forms?

```
for (let index in students) {
  // this is fine
}

for (let student of students) {
  // so is this
}
```

Same thing with `for..in` and `for..of` loops: the declared variable is treated as *inside* the loop body, and thus is handled per iteration (aka, per scope instance). No "re-declaration."

OK, I know you're thinking that I sound like a broken record at this point. But let's explore how `const` impacts these looping constructs. Consider:

```
var keepGoing = true;
while (keepGoing) {
  // ooo, a shiny constant!
  const value = Math.random();
  if (value > 0.5) {
    keepGoing = false;
}
}
```

Just like the `let` variant of this program we saw earlier, `const` is being run exactly once within each loop iteration, so it's safe from "re-declaration" troubles. But things get more complicated when we talk about `for`-loops.

`for..in` and `for..of` are fine to use with `const`:

```
for (const index in students) {
  // this is fine
}

for (const student of students) {
  // this is also fine
}
```

But not the general `for` -loop:

```
for (const i = 0; i < 3; i++) {
    // oops, this is going to fail with
    // a Type Error after the first iteration
}
```

What's wrong here? We could use `let` just fine in this construct, and we asserted that it creates a new `i` for each loop iteration scope, so it doesn't even seem to be a "re-declaration."

Let's mentally "expand" that loop like we did earlier:

```
{
    // a fictional variable for illustration
    const $$i = 0;

    for ( ; $$i < 3; $$i++) {
        // here's our actual loop `i`!
        const i = $$i;
        //
    }
}
```

Do you spot the problem? Our `i` is indeed just created once inside the loop. That's not the problem. The problem is the conceptual `$$i` that must be incremented each time with the `$$i++` expression. That's **re-assignment** (not "re-declaration"), which isn't allowed for constants.

Remember, this "expanded" form is only a conceptual model to help you intuit the source of the problem. You might wonder if JS could have effectively made the `const $$i = 0` instead into `let $ii = 0`, which would then allow `const` to work with our classic `for` -loop? It's possible, but then it could have introduced potentially surprising exceptions to `for` -loop semantics.

For example, it would have been a rather arbitrary (and likely confusing) nuanced exception to allow `i++` in the `for` -loop header to skirt strictness of the `const` assignment, but not allow other re-assignments of `i` inside the loop iteration, as is sometimes useful.

The straightforward answer is: `const` can't be used with the classic `for` -loop form because of the required re-assignment.

Interestingly, if you don't do re-assignment, then it's valid:

```
var keepGoing = true;

for (const i = 0; keepGoing; /* nothing here */ ) {
    keepGoing = (Math.random() > 0.5);
    //
}
```

That works, but it's pointless. There's no reason to declare `i` in that position with a `const`, since the whole point of such a variable in that position is **to be used for counting iterations**. Just use a different loop form, like a `while` loop, or use a `let !`

## Uninitialized Variables (aka, TDZ)

With `var` declarations, the variable is "hoisted" to the top of its scope. But it's also automatically initialized to the `undefined` value, so that the variable can be used throughout the entire scope.

However, `let` and `const` declarations are not quite the same in this respect.

Consider:

```
console.log(studentName);
// ReferenceError

let studentName = "Suzy";
```

The result of this program is that a `ReferenceError` is thrown on the first line. Depending on your JS environment, the error message may say something like: "Cannot access studentName before initialization."

**NOTE:**

The error message as seen here used to be much more vague or misleading. Thankfully, several of us in the community were successfully able to lobby for JS engines to improve this error message so it more accurately tells you what's wrong!

That error message is quite indicative of what's wrong: `studentName` exists on line 1, but it's not been initialized, so it cannot be used yet. Let's try this:

```
studentName = "Suzy"; // let's try to initialize it!
// ReferenceError

console.log(studentName);

let studentName;
```

Oops. We still get the `ReferenceError`, but now on the first line where we're trying to assign to (aka, initialize!) this so-called "uninitialized" variable `studentName`. What's the deal!?

The real question is, how do we initialize an uninitialized variable? For `let / const`, the **only way** to do so is with an assignment attached to a declaration statement. An assignment by itself is insufficient! Consider:

```
let studentName = "Suzy";
console.log(studentName); // Suzy
```

Here, we are initializing the `studentName` (in this case, to `"Suzy"` instead of `undefined`) by way of the `let` declaration statement form that's coupled with an assignment.

Alternatively:

```
// ..

let studentName;
// or:
// let studentName = undefined;

// ..

studentName = "Suzy";

console.log(studentName);
// Suzy
```

**NOTE:**

That's interesting! Recall from earlier, we said that `var studentName;` is *not* the same as `var studentName = undefined;`, but here with `let`, they behave the same. The difference comes down to the fact that `var studentName` automatically initializes at the top of the scope, where `let studentName` does not.

Remember that we've asserted a few times so far that *Compiler* ends up removing any `var / let / const` declarators, replacing them with the instructions at the top of each scope to register the appropriate identifiers.

So if we analyze what's going on here, we see that an additional nuance is that *Compiler* is also adding an instruction in the middle of the program, at the point where the variable `studentName` was declared, to handle that declaration's auto-initialization. We cannot use the variable at any point prior to that initialization occurring. The same goes for `const` as it does for `let`.

The term coined by TC39 to refer to this *period of time* from the entering of a scope to where the auto-initialization of the variable occurs is: Temporal Dead Zone (TDZ).

The TDZ is the time window where a variable exists but is still uninitialized, and therefore cannot be accessed in any way. Only the execution of the instructions left by *Compiler* at the point of the original declaration can do that initialization. After that moment, the TDZ is done, and the variable is free to be used for the rest of the scope.

A `var` also has technically has a TDZ, but it's zero in length and thus unobservable to our programs! Only `let` and `const` have an observable TDZ.

By the way, "temporal" in TDZ does indeed refer to *time* not *position in code*. Consider:

```
askQuestion();
// ReferenceError

let studentName = "Suzy";

function askQuestion() {
  console.log(`#${ studentName }, do you know?`);
}
```

Even though positionally the `console.log(...)` referencing `studentName` comes *after* the `let studentName` declaration, timing wise the `askQuestion()` function is invoked *before* the `let` statement is encountered, while `studentName` is still in its TDZ! Hence the error.

There's a common misconception that TDZ means `let` and `const` do not hoist. This is an inaccurate, or at least slightly misleading, claim. They definitely hoist.

The actual difference is that `let / const` declarations do not automatically initialize at the beginning of the scope, the way `var` does. The *debate* then is if the auto-initialization is *part of hoisting*, or not? I think auto-registration of a variable at the top of the scope (i.e., what I call "hoisting") and auto-initialization at the top of the scope (to `undefined`) are distinct operations and shouldn't be lumped together under the single term "hoisting."

We've already seen that `let` and `const` don't auto-initialize at the top of the scope. But let's prove that `let` and `const` *do* hoist (auto-register at the top of the scope), courtesy of our friend shadowing (see "Shadowing" in Chapter 3):

```
var studentName = "Kyle";

{
  console.log(studentName);
  // ???

  // ...

  let studentName = "Suzy";
  console.log(studentName);
  // Suzy
}
```

What's going to happen with the first `console.log(...)` statement? If `let studentName` didn't hoist to the top of the scope, then the first `console.log(...)` *should* print "Kyle", right? At that moment, it would seem, only the outer `studentName` exists, so that's the variable `console.log(...)` should access and print.

But instead, the first `console.log(..)` throws a TDZ error, because in fact, the inner scope's `studentName` **was** hoisted (auto-registered at the top of the scope). What **didn't** happen (yet!) was the auto-initialization of that inner `studentName`; it's still uninitialized at that moment, hence the TDZ violation!

So to summarize, TDZ errors occur because `let / const` declarations *do* hoist their declarations to the top of their scopes, but unlike `var`, they defer the auto-initialization of their variables until the moment in the code's sequencing where the original declaration appeared. This window of time (hint: temporal), whatever its length, is the TDZ.

How can you avoid TDZ errors?

My advice: always put your `let` and `const` declarations at the top of any scope. Shrink the TDZ window to zero (or near zero) length, and then it'll be moot.

But why is TDZ even a thing? Why didn't TC39 dictate that `let / const` auto-initialize the way `var` does? Just be patient, we'll come back to explore the *why* of TDZ in Appendix A.

## Finally Initialized

Working with variables has much more nuance than it seems at first glance. *Hoisting*, *(re)declaration*, and the *TDZ* are common sources of confusion for developers, especially those who have worked in other languages before coming to JS. Before moving on, make sure your mental model is fully grounded on these aspects of JS scope and variables.

Hoisting is generally cited as an explicit mechanism of the JS engine, but it's really more a metaphor to describe the various ways JS handles variable declarations during compilation. But even as a metaphor, hoisting offers useful structure for thinking about the life-cycle of a variable—when it's created, when it's available to use, when it goes away.

Declaration and re-declaration of variables tend to cause confusion when thought of as runtime operations. But if you shift to compile-time thinking for these operations, the quirks and *shadows* diminish.

The TDZ (temporal dead zone) error is strange and frustrating when encountered. Fortunately, TDZ is relatively straightforward to avoid if you're always careful to place `let / const` declarations at the top of any scope.

As you successfully navigate these twists and turns of variable scope, the next chapter will lay out the factors that guide our decisions to place our declarations in various scopes, especially nested blocks.

# You Don't Know JS Yet: Scope & Closures - 2nd Edition

## Chapter 6: Limiting Scope Exposure

So far our focus has been explaining the mechanics of how scopes and variables work. With that foundation now firmly in place, our attention raises to a higher level of thinking: decisions and patterns we apply across the whole program.

To begin, we're going to look at how and why we should be using different levels of scope (functions and blocks) to organize our program's variables, specifically to reduce scope over-exposure.

### Least Exposure

It makes sense that functions define their own scopes. But why do we need blocks to create scopes as well?

Software engineering articulates a fundamental discipline, typically applied to software security, called "The Principle of Least Privilege" (POLP). [POLP](#) And a variation of this principle that applies to our current discussion is typically labeled as "Least Exposure" (POLE).

POLP expresses a defensive posture to software architecture: components of the system should be designed to function with least privilege, least access, least exposure. If each piece is connected with minimum-necessary capabilities, the overall system is stronger from a security standpoint, because a compromise or failure of one piece has a minimized impact on the rest of the system.

If POLP focuses on system-level component design, the POLE *Exposure* variant focuses on a lower level; we'll apply it to how scopes interact with each other.

In following POLE, what do we want to minimize the exposure of? Simply: the variables registered in each scope.

Think of it this way: why shouldn't you just place all the variables of your program out in the global scope? That probably immediately feels like a bad idea, but it's worth considering why that is. When variables used by one part of the program are exposed to another part of the program, via scope, there are three main hazards that often arise:

- **Naming Collisions:** if you use a common and useful variable/function name in two different parts of the program, but the identifier comes from one shared scope (like the global scope), then name collision occurs, and it's very likely that bugs will occur as one part uses the variable/function in a way the other part doesn't expect.

For example, imagine if all your loops used a single global `i` index variable, and then it happens that one loop in a function is running during an iteration of a loop from another function, and now the shared `i` variable gets an unexpected value.

- **Unexpected Behavior:** if you expose variables/functions whose usage is otherwise *private* to a piece of the program, it allows other developers to use them in ways you didn't intend, which can violate expected behavior and cause bugs.

For example, if your part of the program assumes an array contains all numbers, but someone else's code accesses and modifies the array to include booleans and strings, your code may then misbehave in unexpected ways.

Worse, exposure of *private* details invites those with mal-intent to try to work around limitations you have imposed, to do things with your part of the software that shouldn't be allowed.

- **Unintended Dependency:** if you expose variables/functions unnecessarily, it invites other developers to use and depend on those otherwise *private* pieces. While that doesn't break your program today, it creates a refactoring hazard in the future, because now you cannot as easily refactor that variable or function without potentially breaking other parts of the software that you don't control.

For example, if your code relies on an array of numbers, and you later decide it's better to use some other data structure instead of an array, you now must take on the liability of adjusting other affected parts of the software.

POLE, as applied to variable/function scoping, essentially says, default to exposing the bare minimum necessary, keeping everything else as private as possible. Declare variables in as small and deeply nested of scopes as possible, rather than placing everything in the global (or even outer function) scope.

If you design your software accordingly, you have a much greater chance of avoiding (or at least minimizing) these three hazards.

Consider:

```
function diff(x,y) {
  if (x > y) {
    let tmp = x;
    x = y;
    y = tmp;
  }

  return y - x;
}

diff(3,7);      // 4
diff(7,5);      // 2
```

In this `diff(..)` function, we want to ensure that `y` is greater than or equal to `x`, so that when we subtract (`y - x`), the result is `0` or larger. If `x` is initially larger (the result would be negative!), we swap `x` and `y` using a `tmp` variable, to keep the result positive.

In this simple example, it doesn't seem to matter whether `tmp` is inside the `if` block or whether it belongs at the function level —it certainly shouldn't be a global variable! However, following the POLE principle, `tmp` should be as hidden in scope as possible. So we block scope `tmp` (using `let`) to the `if` block.

## Hiding in Plain (Function) Scope

It should now be clear why it's important to hide our variable and function declarations in the lowest (most deeply nested) scopes possible. But how do we do so?

We've already seen the `let` and `const` keywords, which are block scoped declarators; we'll come back to them in more detail shortly. But first, what about hiding `var` or `function` declarations in scopes? That can easily be done by wrapping a `function` scope around a declaration.

Let's consider an example where `function` scoping can be useful.

The mathematical operation "factorial" (notated as "6!") is the multiplication of a given integer against all successively lower integers down to `1` —actually, you can stop at `2` since multiplying `1` does nothing. In other words, "6!" is the same as "6 5!", *which is the same as "6 5 \* 4!"*, and so on. Because of the nature of the math involved, once any given integer's factorial (like "4!") has been calculated, we shouldn't need to do that work again, as it'll always be the same answer.

So if you naively calculate factorial for `6`, then later want to calculate factorial for `7`, you might unnecessarily re-calculate the factorials of all the integers from 2 up to 6. If you're willing to trade memory for speed, you can solve that wasted computation by caching each integer's factorial as it's calculated:

```
var cache = {};

function factorial(x) {
  if (x < 2) return 1;
  if (!(x in cache)) {
    cache[x] = x * factorial(x - 1);
  }
}
```

```

        return cache[x];
    }

factorial(6);
// 720

cache;
// {
//   "2": 2,
//   "3": 6,
//   "4": 24,
//   "5": 120,
//   "6": 720
// }

factorial(7);
// 5040

```

We're storing all the computed factorials in `cache` so that across multiple calls to `factorial(..)`, the previous computations remain. But the `cache` variable is pretty obviously a *private* detail of how `factorial(..)` works, not something that should be exposed in an outer scope—especially not the global scope.

**NOTE:**

`factorial(..)` here is recursive—a call to itself is made from inside—but that's just for brevity of code sake; a non-recursive implementation would yield the same scoping analysis with respect to `cache`.

However, fixing this over-exposure issue is not as simple as hiding the `cache` variable inside `factorial(..)`, as it might seem. Since we need `cache` to survive multiple calls, it must be located in a scope outside that function. So what can we do?

Define another middle scope (between the outer/global scope and the inside of `factorial(..)`) for `cache` to be located:

```

// outer/global scope

function hideTheCache() {
    // "middle scope", where we hide `cache`
    var cache = {};

    return factorial;

    // *****

    function factorial(x) {
        // inner scope
        if (x < 2) return 1;
        if (!(x in cache)) {
            cache[x] = x * factorial(x - 1);
        }
        return cache[x];
    }
}

var factorial = hideTheCache();

factorial(6);
// 720

factorial(7);
// 5040

```

The `hideTheCache()` function serves no other purpose than to create a scope for `cache` to persist in across multiple calls to `factorial(..)`. But for `factorial(..)` to have access to `cache`, we have to define `factorial(..)` inside that same scope. Then we return the function reference, as a value from `hideTheCache()`, and store it in an outer scope variable, also named

`factorial`. Now as we call `factorial(..)` (multiple times!), its persistent `cache` stays hidden yet accessible only to `factorial(..)`!

OK, but... it's going to be tedious to define (and name!) a `hideTheCache(..)` function scope each time such a need for variable/function hiding occurs, especially since we'll likely want to avoid name collisions with this function by giving each occurrence a unique name. Ugh.

#### NOTE:

The illustrated technique—caching a function's computed output to optimize performance when repeated calls of the same inputs are expected—is quite common in the Functional Programming (FP) world, canonically referred to as "memoization"; this caching relies on closure (see Chapter 7). Also, there are memory usage concerns (addressed in "A Word About Memory" in Appendix B). FP libraries will usually provide an optimized and vetted utility for memoization of functions, which would take the place of `hideTheCache(..)` here. Memoization is beyond the *scope* (pun intended!) of our discussion, but see my *Functional-Light JavaScript* book for more information.

Rather than defining a new and uniquely named function each time one of those scope-only-for-the-purpose-of-hiding-a-variable situations occurs, a perhaps better solution is to use a function expression:

```
var factorial = (function hideTheCache() {
  var cache = {};

  function factorial(x) {
    if (x < 2) return 1;
    if (!(x in cache)) {
      cache[x] = x * factorial(x - 1);
    }
    return cache[x];
  }

  return factorial;
})();

factorial(6);
// 720

factorial(7);
// 5040
```

Wait! This is still using a function to create the scope for hiding `cache`, and in this case, the function is still named `hideTheCache`, so how does that solve anything?

Recall from "Function Name Scope" (in Chapter 3), what happens to the name identifier from a `function` expression. Since `hideTheCache(..)` is defined as a `function` expression instead of a `function` declaration, its name is in its own scope—essentially the same scope as `cache`—rather than in the outer/global scope.

That means we can name every single occurrence of such a function expression the exact same name, and never have any collision. More appropriately, we can name each occurrence semantically based on whatever it is we're trying to hide, and not worry that whatever name we choose is going to collide with any other `function` expression scope in the program.

In fact, we *could* just leave off the name entirely—thus defining an "anonymous `function` expression" instead. But Appendix A will discuss the importance of names even for such scope-only functions.

## Invoking Function Expressions Immediately

There's another important bit in the previous factorial recursive program that's easy to miss: the line at the end of the `function` expression that contains `)();`.

Notice that we surrounded the entire `function` expression in a set of `( ... )`, and then on the end, we added that second `()` parentheses set; that's actually calling the `function` expression we just defined. Moreover, in this case, the first set of surrounding `( ... )` around the function expression is not strictly necessary (more on that in a moment), but we used them for readability sake anyway.

So, in other words, we're defining a `function` expression that's then immediately invoked. This common pattern has a (very creative!) name: Immediately Invoked Function Expression (IIFE).

An IIFE is useful when we want to create a scope to hide variables/functions. Since it's an expression, it can be used in **any** place in a JS program where an expression is allowed. An IIFE can be named, as with `hideTheCache()`, or (much more commonly!) unnamed/anonymous. And it can be standalone or, as before, part of another statement—`hideTheCache()` returns the `factorial()` function reference which is then `=` assigned to the variable `factorial`.

For comparison, here's an example of a standalone IIFE:

```
// outer scope

(function(){
    // inner hidden scope
})();

// more outer scope
```

Unlike earlier with `hideTheCache()`, where the outer surrounding `(...)` were noted as being an optional stylistic choice, for a standalone IIFE they're **required**; they distinguish the `function` as an expression, not a statement. For consistency, however, always surround an IIFE `function` with `( ... )`.

#### NOTE:

Technically, the surrounding `( ... )` aren't the only syntactic way to ensure the `function` in an IIFE is treated by the JS parser as a function expression. We'll look at some other options in Appendix A.

## Function Boundaries

Beware that using an IIFE to define a scope can have some unintended consequences, depending on the code around it. Because an IIFE is a full function, the function boundary alters the behavior of certain statements/constructs.

For example, a `return` statement in some piece of code would change its meaning if an IIFE is wrapped around it, because now the `return` would refer to the IIFE's function. Non-arrow function IIFEs also change the binding of a `this` keyword—more on that in the *Objects & Classes* book. And statements like `break` and `continue` won't operate across an IIFE function boundary to control an outer loop or block.

So, if the code you need to wrap a scope around has `return`, `this`, `break`, or `continue` in it, an IIFE is probably not the best approach. In that case, you might look to create the scope with a block instead of a function.

## Scoping with Blocks

You should by this point feel fairly comfortable with the merits of creating scopes to limit identifier exposure.

So far, we looked at doing this via `function` (i.e., IIFE) scope. But let's now consider using `let` declarations with nested blocks. In general, any `{ ... }` curly-brace pair which is a statement will act as a block, but **not necessarily** as a scope.

A block only becomes a scope if necessary, to contain its block-scoped declarations (i.e., `let` or `const`). Consider:

```
{
    // not necessarily a scope (yet)
```

```
// ...

// now we know the block needs to be a scope
let thisIsNowAScope = true;

for (let i = 0; i < 5; i++) {
  // this is also a scope, activated each
  // iteration
  if (i % 2 == 0) {
    // this is just a block, not a scope
    console.log(i);
  }
}
// 0 2 4
```

Not all `{ ... }` curly-brace pairs create blocks (and thus are eligible to become scopes):

- Object literals use `{ ... }` curly-brace pairs to delimit their key-value lists, but such object values are **not** scopes.
- `class` uses `{ ... }` curly-braces around its body definition, but this is not a block or scope.
- A `function` uses `{ ... }` around its body, but this is not technically a block—it's a single statement for the function body. It *is*, however, a (function) scope.
- The `{ ... }` curly-brace pair on a `switch` statement (around the set of `case` clauses) does not define a block/scope.

Other than such non-block examples, a `{ ... }` curly-brace pair can define a block attached to a statement (like an `if` or `for`), or stand alone by itself—see the outermost `{ ... }` curly brace pair in the previous snippet. An explicit block of this sort—if it has no declarations, it's not actually a scope—serves no operational purpose, though it can still be useful as a semantic signal.

Explicit standalone `{ ... }` blocks have always been valid JS syntax, but since they couldn't be a scope prior to ES6's `let` / `const`, they are quite rare. However, post ES6, they're starting to catch on a little bit.

In most languages that support block scoping, an explicit block scope is an extremely common pattern for creating a narrow slice of scope for one or a few variables. So following the POLE principle, we should embrace this pattern more widespread in JS as well; use (explicit) block scoping to narrow the exposure of identifiers to the minimum practical.

An explicit block scope can be useful even inside of another block (whether the outer block is a scope or not).

For example:

```
if (somethingHappened) {
  // this is a block, but not a scope

  {
    // this is both a block and an
    // explicit scope
    let msg = somethingHappened.message();
    notifyOthers(msg);
  }

  ...
}

recoverFromSomething();
```

Here, the `{ ... }` curly-brace pair **inside** the `if` statement is an even smaller inner explicit block scope for `msg`, since that variable is not needed for the entire `if` block. Most developers would just block-scope `msg` to the `if` block and move on. And to be fair, when there's only a few lines to consider, it's a toss-up judgement call. But as code grows, these over-exposure issues become more pronounced.

So does it matter enough to add the extra `{ ... }` pair and indentation level? I think you should follow POLE and always (within reason!) define the smallest block for each variable. So I recommend using the extra explicit block scope as shown.

Recall the discussion of TDZ errors from "Uninitialized Variables (TDZ)" (Chapter 5). My suggestion there was: to minimize the risk of TDZ errors with `let` / `const` declarations, always put those declarations at the top of their scope.

If you find yourself placing a `let` declaration in the middle of a scope, first think, "Oh, no! TDZ alert!" If this `let` declaration isn't needed in the first half of that block, you should use an inner explicit block scope to further narrow its exposure!

Another example with an explicit block scope:

```
function getNextMonthStart(dateStr) {
  var nextMonth, year;

  {
    let curMonth;
    [ , year, curMonth ] = dateStr.match(
      /(\d{4})-(\d{2})-\d{2}/
    ) || [];
    nextMonth = (Number(curMonth) % 12) + 1;
  }

  if (nextMonth == 1) {
    year++;
  }

  return `${year}-${String(nextMonth).padStart(2, "0")}
    }-01`;
}
getNextMonthStart("2019-12-25"); // 2020-01-01
```

Let's first identify the scopes and their identifiers:

1. The outer/global scope has one identifier, the function `getNextMonthStart(..)`.
2. The function scope for `getNextMonthStart(..)` has three: `dateStr` (parameter), `nextMonth`, and `year`.
3. The `{ ... }` curly-brace pair defines an inner block scope that includes one variable: `curMonth`.

So why put `curMonth` in an explicit block scope instead of just alongside `nextMonth` and `year` in the top-level function scope? Because `curMonth` is only needed for those first two statements; at the function scope level it's over-exposed.

This example is small, so the hazards of over-exposing `curMonth` are pretty limited. But the benefits of the POLE principle are best achieved when you adopt the mindset of minimizing scope exposure by default, as a habit. If you follow the principle consistently even in the small cases, it will serve you more as your programs grow.

Let's now look at an even more substantial example:

```
function sortNamesByLength(names) {
  var buckets = [];

  for (let firstName of names) {
    if (buckets[firstName.length] == null) {
      buckets[firstName.length] = [];
    }
    buckets[firstName.length].push(firstName);
  }

  // a block to narrow the scope
  {
    let sortedNames = [];

    for (let bucket of buckets) {
```

```
    if (bucket) {
        // sort each bucket alphanumerically
        bucket.sort();

        // append the sorted names to our
        // running list
        sortedNames = [
            ...sortedNames,
            ...bucket
        ];
    }

    return sortedNames;
}
}

sortNamesByLength([
    "Sally",
    "Suzy",
    "Frank",
    "John",
    "Jennifer",
    "Scott"
]);
// [ "John", "Suzy", "Frank", "Sally",
//   "Scott", "Jennifer" ]
```

There are six identifiers declared across five different scopes. Could all of these variables have existed in the single outer/global scope? Technically, yes, since they're all uniquely named and thus have no name collisions. But this would be really poor code organization, and would likely lead to both confusion and future bugs.

We split them out into each inner nested scope as appropriate. Each variable is defined at the innermost scope possible for the program to operate as desired.

`sortedNames` could have been defined in the top-level function scope, but it's only needed for the second half of this function. To avoid over-exposing that variable in a higher level scope, we again follow POLE and block-scope it in the inner explicit block scope.

## var and let

Next, let's talk about the declaration `var buckets`. That variable is used across the entire function (except the final `return` statement). Any variable that is needed across all (or even most) of a function should be declared so that such usage is obvious.

### NOTE:

The parameter `names` isn't used across the whole function, but there's no way limit the scope of a parameter, so it behaves as a function-wide declaration regardless.

So why did we use `var` instead of `let` to declare the `buckets` variable? There's both semantic and technical reasons to choose `var` here.

Stylistically, `var` has always, from the earliest days of JS, signaled "variable that belongs to a whole function." As we asserted in "Lexical Scope" (Chapter 1), `var` attaches to the nearest enclosing function scope, no matter where it appears. That's true even if `var` appears inside a block:

```
function diff(x,y) {
    if (x > y) {
        var tmp = x;      // `tmp` is function-scoped
        x = y;
        y = tmp;
    }
}
```

```

    return y - x;
}

```

Even though `var` is inside a block, its declaration is function-scoped (to `diff(..)`), not block-scoped.

While you can declare `var` inside a block (and still have it be function-scoped), I would recommend against this approach except in a few specific cases (discussed in Appendix A). Otherwise, `var` should be reserved for use in the top-level scope of a function.

Why not just use `let` in that same location? Because `var` is visually distinct from `let` and therefore signals clearly, "this variable is function-scoped." Using `let` in the top-level scope, especially if not in the first few lines of a function, and when all the other declarations in blocks use `let`, does not visually draw attention to the difference with the function-scoped declaration.

In other words, I feel `var` better communicates function-scoped than `let` does, and `let` both communicates (and achieves!) block-scoping where `var` is insufficient. As long as your programs are going to need both function-scoped and block-scoped variables, the most sensible and readable approach is to use both `var` and `let` together, each for their own best purpose.

There are other semantic and operational reasons to choose `var` or `let` in different scenarios. We'll explore the case for `var` and `let` in more detail in Appendix A.

#### **WARNING:**

My recommendation to use both `var` and `let` is clearly controversial and contradicts the majority. It's far more common to hear assertions like, "var is broken, let fixes it" and, "never use var, let is the replacement." Those opinions are valid, but they're merely opinions, just like mine. `var` is not factually broken or deprecated; it has worked since early JS and it will continue to work as long as JS is around.

## Where To `let` ?

My advice to reserve `var` for (mostly) only a top-level function scope means that most other declarations should use `let`. But you may still be wondering how to decide where each declaration in your program belongs?

POLE already guides you on those decisions, but let's make sure we explicitly state it. The way to decide is not based on which keyword you want to use. The way to decide is to ask, "What is the most minimal scope exposure that's sufficient for this variable?"

Once that is answered, you'll know if a variable belongs in a block scope or the function scope. If you decide initially that a variable should be block-scoped, and later realize it needs to be elevated to be function-scoped, then that dictates a change not only in the location of that variable's declaration, but also the declarator keyword used. The decision-making process really should proceed like that.

If a declaration belongs in a block scope, use `let`. If it belongs in the function scope, use `var` (again, just my opinion).

But another way to sort of visualize this decision making is to consider the pre-ES6 version of a program. For example, let's recall `diff(..)` from earlier:

```

function diff(x,y) {
  var tmp;

  if (x > y) {
    tmp = x;
    x = y;
    y = tmp;
  }

  return y - x;
}

```

In this version of `diff(..)`, `tmp` is clearly declared in the function scope. Is that appropriate for `tmp`? I would argue, no. `tmp` is only needed for those few statements. It's not needed for the `return` statement. It should therefore be block-scoped.

Prior to ES6, we didn't have `let` so we couldn't *actually* block-scope it. But we could do the next-best thing in signaling our intent:

```
function diff(x,y) {
  if (x > y) {
    // `tmp` is still function-scoped, but
    // the placement here semantically
    // signals block-scoping
    var tmp = x;
    x = y;
    y = tmp;
  }

  return y - x;
}
```

Placing the `var` declaration for `tmp` inside the `if` statement signals to the reader of the code that `tmp` belongs to that block. Even though JS doesn't enforce that scoping, the semantic signal still has benefit for the reader of your code.

Following this perspective, you can find any `var` that's inside a block of this sort and switch it to `let` to enforce the semantic signal already being sent. That's proper usage of `let` in my opinion.

Another example that was historically based on `var` but which should now pretty much always use `let` is the `for` loop:

```
for (var i = 0; i < 5; i++) {
  // do something
}
```

No matter where such a loop is defined, the `i` should basically always be used only inside the loop, in which case POLE dictates it should be declared with `let` instead of `var`:

```
for (let i = 0; i < 5; i++) {
  // do something
}
```

Almost the only case where switching a `var` to a `let` in this way would "break" your code is if you were relying on accessing the loop's iterator (`i`) outside/after the loop, such as:

```
for (var i = 0; i < 5; i++) {
  if (checkValue(i)) {
    break;
  }

  if (i < 5) {
    console.log("The loop stopped early!");
  }
}
```

This usage pattern is not terribly uncommon, but most feel it smells like poor code structure. A preferable approach is to use another outer-scoped variable for that purpose:

```
var lastI;

for (let i = 0; i < 5; i++) {
  lastI = i;
  if (checkValue(i)) {
```

```
        break;
    }

}

if (lastI < 5) {
    console.log("The loop stopped early!");
}
```

`lastI` is needed across this whole scope, so it's declared with `var`. `i` is only needed in (each) loop iteration, so it's declared with `let`.

## What's the Catch?

So far we've asserted that `var` and parameters are function-scoped, and `let / const` signal block-scoped declarations. There's one little exception to call out: the `catch` clause.

Since the introduction of `try..catch` back in ES3 (in 1999), the `catch` clause has used an additional (little-known) block-scoping declaration capability:

```
try {
    doesntExist();
}
catch (err) {
    console.log(err);
    // ReferenceError: 'doesntExist' is not defined
    // ^^^^ message printed from the caught exception

    let onlyHere = true;
    var outerVariable = true;
}

console.log(outerVariable);      // true

console.log(err);
// ReferenceError: 'err' is not defined
// ^^^^ this is another thrown (uncaught) exception
```

The `err` variable declared by the `catch` clause is block-scoped to that block. This `catch` clause block can hold other block-scoped declarations via `let`. But a `var` declaration inside this block still attaches to the outer function/global scope.

ES2019 (recently, at the time of writing) changed `catch` clauses so their declaration is optional; if the declaration is omitted, the `catch` block is no longer (by default) a scope; it's still a block, though!

So if you need to react to the condition *that an exception occurred* (so you can gracefully recover), but you don't care about the error value itself, you can omit the `catch` declaration:

```
try {
    doOptionOne();
}
catch { // catch-declaration omitted
    doOptionTwoInstead();
}
```

This is a small but delightful simplification of syntax for a fairly common use case, and may also be slightly more performant in removing an unnecessary scope!

## Function Declarations in Blocks (FiB)

We've seen now that declarations using `let` or `const` are block-scoped, and `var` declarations are function-scoped. So what about `function` declarations that appear directly inside blocks? As a feature, this is called "FiB."

We typically think of `function` declarations like they're the equivalent of a `var` declaration. So are they function-scoped like `var` is?

No and yes. I know... that's confusing. Let's dig in:

```
if (false) {
  function ask() {
    console.log("Does this run?");
  }
}
ask();
```

What do you expect for this program to do? Three reasonable outcomes:

1. The `ask()` call might fail with a `ReferenceError` exception, because the `ask` identifier is block-scoped to the `if` block scope and thus isn't available in the outer/global scope.
2. The `ask()` call might fail with a `TypeError` exception, because the `ask` identifier exists, but it's `undefined` (since the `if` statement doesn't run) and thus not a callable function.
3. The `ask()` call might run correctly, printing out the "Does it run?" message.

Here's the confusing part: depending on which JS environment you try that code snippet in, you may get different results! This is one of those few crazy areas where existing legacy behavior betrays a predictable outcome.

The JS specification says that `function` declarations inside of blocks are block-scoped, so the answer should be (1). However, most browser-based JS engines (including v8, which comes from Chrome but is also used in Node) will behave as (2), meaning the identifier is scoped outside the `if` block but the function value is not automatically initialized, so it remains `undefined`.

Why are browser JS engines allowed to behave contrary to the specification? Because these engines already had certain behaviors around FiB before ES6 introduced block scoping, and there was concern that changing to adhere to the specification might break some existing website JS code. As such, an exception was made in Appendix B of the JS specification, which allows certain deviations for browser JS engines (only!).

#### NOTE:

You wouldn't typically categorize Node as a browser JS environment, since it usually runs on a server. But Node's v8 engine is shared with Chrome (and Edge) browsers. Since v8 is first a browser JS engine, it adopts this Appendix B exception, which then means that the browser exceptions are extended to Node.

One of the most common use cases for placing a `function` declaration in a block is to conditionally define a function one way or another (like with an `if..else` statement) depending on some environment state. For example:

```
if (typeof Array.isArray != "undefined") {
  function isArray(a) {
    return Array.isArray(a);
  }
} else {
  function isArray(a) {
    return Object.prototype.toString.call(a)
      == "[object Array]";
  }
}
```

It's tempting to structure code this way for performance reasons, since the `typeof Array.isArray` check is only performed once, as opposed to defining just one `isArray(..)` and putting the `if` statement inside it—the check would then run unnecessarily on every call.

**WARNING:**

In addition to the risks of FiB deviations, another problem with conditional-definition of functions is it's harder to debug such a program. If you end up with a bug in the `isArray(..)` function, you first have to figure out *which* `isArray(..)` implementation is actually running! Sometimes, the bug is that the wrong one was applied because the conditional check was incorrect! If you define multiple versions of a function, that program is always harder to reason about and maintain.

In addition to the previous snippets, several other FiB corner cases are lurking; such behaviors in various browsers and non-browser JS environments (JS engines that aren't browser based) will likely vary. For example:

```
if (true) {
    function ask() {
        console.log("Am I called?");
    }
}

if (true) {
    function ask() {
        console.log("Or what about me?");
    }
}

for (let i = 0; i < 5; i++) {
    function ask() {
        console.log("Or is it one of these?");
    }
}

ask();

function ask() {
    console.log("Wait, maybe, it's this one?");
}
```

Recall that function hoisting as described in "When Can I Use a Variable?" (in Chapter 5) might suggest that the final `ask()` in this snippet, with "Wait, maybe..." as its message, would hoist above the call to `ask()`. Since it's the last function declaration of that name, it should "win," right? Unfortunately, no.

It's not my intention to document all these weird corner cases, nor to try to explain why each of them behaves a certain way. That information is, in my opinion, arcane legacy trivia.

My real concern with FiB is, what advice can I give to ensure your code behaves predictably in all circumstances?

As far as I'm concerned, the only practical answer to avoiding the vagaries of FiB is to simply avoid FiB entirely. In other words, never place a `function` declaration directly inside any block. Always place `function` declarations anywhere in the top-level scope of a function (or in the global scope).

So for the earlier `if..else` example, my suggestion is to avoid conditionally defining functions if at all possible. Yes, it may be slightly less performant, but this is the better overall approach:

```
function isArray(a) {
    if (typeof Array.isArray != "undefined") {
        return Array.isArray(a);
    }
    else {
        return Object.prototype.toString.call(a)
            == "[object Array]";
    }
}
```

If that performance hit becomes a critical path issue for your application, I suggest you consider this approach:

```
var isArray = function isArray(a) {
    return Array.isArray(a);
};

// override the definition, if you must
if (typeof Array.isArray == "undefined") {
    isArray = function isArray(a) {
        return Object.prototype.toString.call(a)
        == "[object Array]";
    };
}
```

It's important to notice that here I'm placing a `function expression`, not a declaration, inside the `if` statement. That's perfectly fine and valid, for `function` expressions to appear inside blocks. Our discussion about FiB is about avoiding `function declarations` in blocks.

Even if you test your program and it works correctly, the small benefit you may derive from using FiB style in your code is far outweighed by the potential risks in the future for confusion by other developers, or variances in how your code runs in other JS environments.

FiB is not worth it, and should be avoided.

## Blocked Over

The point of lexical scoping rules in a programming language is so we can appropriately organize our program's variables, both for operational as well as semantic code communication purposes.

And one of the most important organizational techniques is to ensure that no variable is over-exposed to unnecessary scopes (POLE). Hopefully you now appreciate block scoping much more deeply than before.

Hopefully by you feel like you're standing on much more solid ground with understanding lexical scope. From that base, the next chapter jumps into the weighty topic of closure.

 POLP. *Principle of Least Privilege*, [https://en.wikipedia.org/wiki/Principle\\_of\\_least\\_privilege](https://en.wikipedia.org/wiki/Principle_of_least_privilege), 3 March 2020. ↵

# You Don't Know JS Yet: Scope & Closures - 2nd Edition

## Appendix A: Exploring Further

We will now explore a number of nuances and edges around many of the topics covered in the main text of this book. This appendix is optional, supporting material.

Some people find diving too deeply into the nuanced corner cases and varying opinions creates nothing but noise and distraction—supposedly, developers are better served by sticking to the commonly-tread paths. My approach has been criticized as being impractical and counterproductive. I understand and appreciate that perspective, even if I don't necessarily share it.

I believe it's better to be empowered by knowledge of how things work than to just gloss over details with assumptions and lack of curiosity. Ultimately, you will encounter situations where something bubbles up from a piece you hadn't explored. In other words, you won't get to spend all your time riding on the smooth *happy path*. Wouldn't you rather be prepared for the inevitable bumps of off-roading?

These discussions will also be more heavily influenced by my opinions than the main text was, so keep that in mind as you consume and consider what is presented. This appendix is a bit like a collection of mini-blog posts that elaborate on various book topics. It's long and deep in the weeds, so take your time and don't rush through everything here.

## Implied Scopes

Scopes are sometimes created in non-obvious places. In practice, these implied scopes don't often impact your program behavior, but it's still useful to know they're happening. Keep an eye out for the following surprising scopes:

- Parameter scope
- Function name scope

### Parameter Scope

The conversation metaphor in Chapter 2 implies that function parameters are basically the same as locally declared variables in the function scope. But that's not always true.

Consider:

```
// outer/global scope: RED(1)

function getStudentName(studentID) {
    // function scope: BLUE(2)

    // ...
}
```

Here, `studentID` is considered a "simple" parameter, so it does behave as a member of the `BLUE(2)` function scope. But if we change it to be a non-simple parameter, that's no longer technically the case. Parameter forms considered non-simple include parameters with default values, rest parameters (using `...`), and destructured parameters.

Consider:

```
// outer/global scope: RED(1)

function getStudentName(/*BLUE(2)*/ studentID = 0) {
    // function scope: GREEN(3)
```

```
// ...
}
```

Here, the parameter list essentially becomes its own scope, and the function's scope is then nested inside *that* scope.

Why? What difference does it make? The non-simple parameter forms introduce various corner cases, so the parameter list becomes its own scope to more effectively deal with them.

Consider:

```
function getStudentName(studentID = maxID, maxID) {
    // ...
}
```

Assuming left-to-right operations, the default `= maxID` for the `studentID` parameter requires a `maxID` to already exist (and to have been initialized). This code produces a TDZ error (Chapter 5). The reason is that `maxID` is declared in the parameter scope, but it's not yet been initialized because of the order of parameters. If the parameter order is flipped, no TDZ error occurs:

```
function getStudentName(maxID, studentID = maxID) {
    // ...
}
```

The complication gets even more *in the weeds* if we introduce a function expression into the default parameter position, which then can create its own closure (Chapter 7) over parameters in this implied parameter scope:

```
function whatsTheDealHere(id, defaultID = () => id) {
    id = 5;
    console.log( defaultID() );
}

whatsTheDealHere(3);
// 5
```

That snippet probably makes sense, because the `defaultID()` arrow function closes over the `id` parameter/variable, which we then re-assign to `5`. But now let's introduce a shadowing definition of `id` in the function scope:

```
function whatsTheDealHere(id, defaultID = () => id) {
    var id = 5;
    console.log( defaultID() );
}

whatsTheDealHere(3);
// 3
```

Uh oh! The `var id = 5` is shadowing the `id` parameter, but the closure of the `defaultID()` function is over the parameter, not the shadowing variable in the function body. This proves there's a scope bubble around the parameter list.

But it gets even crazier than that!

```
function whatsTheDealHere(id, defaultID = () => id) {
    var id;

    console.log(`local variable 'id': ${ id }`);
    console.log(
        `parameter 'id' (closure): ${ defaultID() }`
    );

    console.log("reassigning 'id' to 5");
```

```

id = 5;

console.log(`local variable 'id': ${ id }`);
console.log(
  `parameter 'id' (closure): ${ defaultID() }`
);
}

whatsTheDealHere(3);
// local variable 'id': 3  <--- Huh!? Weird!
// parameter 'id' (closure): 3
// reassigning 'id' to 5
// local variable 'id': 5
// parameter 'id' (closure): 3

```

The strange bit here is the first console message. At that moment, the shadowing `id` local variable has just been `var id` declared, which Chapter 5 asserts is typically auto-initialized to `undefined` at the top of its scope. Why doesn't it print `undefined`?

In this specific corner case (for legacy compat reasons), JS doesn't auto-initialize `id` to `undefined`, but rather to the value of the `id` parameter (`3`)!

Though the two `id`s look at that moment like they're one variable, they're actually still separate (and in separate scopes). The `id = 5` assignment makes the divergence observable, where the `id` parameter stays `3` and the local variable becomes `5`.

My advice to avoid getting bitten by these weird nuances:

- Never shadow parameters with local variables
- Avoid using a default parameter function that closes over any of the parameters

At least now you're aware and can be careful about the fact that the parameter list is its own scope if any of the parameters are non-simple.

## Function Name Scope

In the "Function Name Scope" section in Chapter 3, I asserted that the name of a function expression is added to the function's own scope. Recall:

```

var askQuestion = function ofTheTeacher(){
  // ..
};

```

It's true that `ofTheTeacher` is not added to the enclosing scope (where `askQuestion` is declared), but it's also not *just* added to the scope of the function, the way you're likely assuming. It's another strange corner case of implied scope.

The name identifier of a function expression is in its own implied scope, nested between the outer enclosing scope and the main inner function scope.

If `ofTheTeacher` was in the function's scope, we'd expect an error here:

```

var askQuestion = function ofTheTeacher(){
  // why is this not a duplicate declaration error?
  let ofTheTeacher = "Confused, yet?";
};

```

The `let` declaration form does not allow re-declaration (see Chapter 5). But this is perfectly legal shadowing, not re-declaration, because the two `ofTheTeacher` identifiers are in separate scopes.

You'll rarely run into any case where the scope of a function's name identifier matters. But again, it's good to know how these mechanisms actually work. To avoid being bitten, never shadow function name identifiers.

## Anonymous vs. Named Functions

As discussed in Chapter 3, functions can be expressed either in named or anonymous form. It's vastly more common to use the anonymous form, but is that a good idea?

As you contemplate naming your functions, consider:

- Name inference is incomplete
- Lexical names allow self-reference
- Names are useful descriptions
- Arrow functions have no lexical names
- IIFEs also need names

## Explicit or Inferred Names?

Every function in your program has a purpose. If it doesn't have a purpose, take it out, because you're just wasting space. If it *does* have a purpose, there *is* a name for that purpose.

So far many readers likely agree with me. But does that mean we should always put that name into the code? Here's where I'll raise more than a few eyebrows. I say, unequivocally, yes!

First of all, "anonymous" showing up in stack traces is just not all that helpful to debugging:

```
btn.addEventListener("click",function(){
  setTimeout(function(){
    ["a",42].map(function(v){
      console.log(v.toUpperCase());
    });
  },100);
});
// Uncaught TypeError: Cannot read property
// 'toUpperCase' of null
//     at myProgram.js:4
//     at Array.map (<anonymous>)
//     at myProgram.js:3
```

Ugh. Compare to what is reported if I give the functions names:

```
btn.addEventListener("click",function onClick(){
  setTimeout(function waitAMoment(){
    ["a",42].map(function allUpper(v){
      console.log(v.toUpperCase());
    });
  },100);
});
// Uncaught TypeError: v.toUpperCase is not a function
//     at allUpper (myProgram.js:4)
//     at Array.map (<anonymous>)
//     at waitAMoment (myProgram.js:3)
```

See how `waitAMoment` and `allUpper` names appear and give the stack trace more useful information/context for debugging? The program is more debuggable if we use reasonable names for all our functions.

### NOTE:

The unfortunate "<anonymous>" that still shows up refers to the fact that the implementation of `Array.map(...)` isn't

present in our program, but is built into the JS engine. It's not from any confusion our program introduces with readability shortcuts.

By the way, let's make sure we're on the same page about what a named function is:

```
function thisIsNamed() {
  // ...
}

ajax("some.url",function thisIsAlsoNamed(){
  // ...
});

var notNamed = function(){
  // ...
};

makeRequest({
  data: 42,
  cb /* also not a name */: function(){
    // ...
  }
});

var stillNotNamed = function butThisIs(){
  // ...
};
```

"But wait!", you say. Some of those *are* named, right!?

```
var notNamed = function(){
  // ...
};

var config = {
  cb: function(){
    // ...
  }
};

notNamed.name;
// notNamed

config.cb.name;
// cb
```

These are referred to as *inferred* names. Inferred names are fine, but they don't really address the full concern I'm discussing.

## Missing Names?

Yes, these inferred names might show up in stack traces, which is definitely better than "anonymous" showing up. But...

```
function ajax(url,cb) {
  console.log(cb.name);
}

ajax("some.url",function(){
  // ...
});
// ""
```

Oops. Anonymous `function` expressions passed as callbacks are incapable of receiving an inferred name, so `cb.name` holds just the empty string `""`. The vast majority of all `function` expressions, especially anonymous ones, are used as callback arguments; none of these get a name. So relying on name inference is incomplete, at best.

And it's not just callbacks that fall short with inference:

```
var config = {};

config.cb = function(){
    // ...
};

config.cb.name;
// ""

var [ noName ] = [ function(){}
];
noName.name
// ""
```

Any assignment of a `function` expression that's not a *simple assignment* will also fail name inferencing. So, in other words, unless you're careful and intentional about it, essentially almost all anonymous `function` expressions in your program will in fact have no name at all.

Name inference is just... not enough.

And even if a `function` expression *does* get an inferred name, that still doesn't count as being a full named function.

## Who am I?

Without a lexical name identifier, the function has no internal way to refer to itself. Self-reference is important for things like recursion and event handling:

```
// broken
runOperation(function(num){
    if (num <= 1) return 1;
    return num * oopsNoNameToCall(num - 1);
});

// also broken
btn.addEventListener("click",function(){
    console.log("should only respond to one click!");
    btn.removeEventListener("click",oopsNoNameHere);
});
```

Leaving off the lexical name from your callback makes it harder to reliably self-reference the function. You *could* declare a variable in an enclosing scope that references the function, but this variable is *controlled* by that enclosing scope—it could be re-assigned, etc.—so it's not as reliable as the function having its own internal self-reference.

## Names are Descriptors

Lastly, and I think most importantly of all, leaving off a name from a function makes it harder for the reader to tell what the function's purpose is, at a quick glance. They have to read more of the code, including the code inside the function, and the surrounding code outside the function, to figure it out.

Consider:

```
[ 1, 2, 3, 4, 5 ].filter(function(v){
    return v % 2 == 1;
});
```

```
// [ 1, 3, 5 ]

[ 1, 2, 3, 4, 5 ].filter(function keepOnlyOdds(v){
  return v % 2 == 1;
});
// [ 1, 3, 5 ]
```

There's just no reasonable argument to be made that **omitting** the name `keepOnlyOdds` from the first callback more effectively communicates to the reader the purpose of this callback. You saved 13 characters, but lost important readability information. The name `keepOnlyOdds` very clearly tells the reader, at a quick first glance, what's happening.

The JS engine doesn't care about the name. But human readers of your code absolutely do.

Can the reader look at `v % 2 == 1` and figure out what it's doing? Sure. But they have to infer the purpose (and name) by mentally executing the code. Even a brief pause to do so slows down reading of the code. A good descriptive name makes this process almost effortless and instant.

Think of it this way: how many times does the author of this code need to figure out the purpose of a function before adding the name to the code? About once. Maybe two or three times if they need to adjust the name. But how many times will readers of this code have to figure out the name/purpose? Every single time this line is ever read. Hundreds of times? Thousands? More?

No matter the length or complexity of the function, my assertion is, the author should figure out a good descriptive name and add it to the code. Even the one-liner functions in `map(..)` and `then(..)` statements should be named:

```
lookupTheRecords(someData)
  .then(function extractSalesRecords(resp){
    return resp.allSales;
})
  .then(storeRecords);
```

The name `extractSalesRecords` tells the reader the purpose of this `then(..)` handler *better* than just inferring that purpose from mentally executing `return resp.allSales`.

The only excuse for not including a name on a function is either laziness (don't want to type a few extra characters) or uncreativity (can't come up with a good name). If you can't figure out a good name, you likely don't understand the function and its purpose yet. The function is perhaps poorly designed, or it does too many things, and should be re-worked. Once you have a well-designed, single-purpose function, its proper name should become evident.

Here's a trick I use: while first writing a function, if I don't fully understand its purpose and can't think of a good name to use, I just use `TODO` as the name. That way, later when reviewing my code, I'm likely to find those name placeholders, and I'm more inclined (and more prepared!) to go back and figure out a better name, rather than just leave it as `TODO`.

All functions need names. Every single one. No exceptions. Any name you omit is making the program harder to read, harder to debug, harder to extend and maintain later.

## Arrow Functions

Arrow functions are **always** anonymous, even if (rarely) they're used in a way that gives them an inferred name. I just spent several pages explaining why anonymous functions are a bad idea, so you can probably guess what I think about arrow functions.

Don't use them as a general replacement for regular functions. They're more concise, yes, but that brevity comes at the cost of omitting key visual delimiters that help our brains quickly parse out what we're reading. And, to the point of this discussion, they're anonymous, which makes them worse for readability from that angle as well.

Arrow functions have a purpose, but that purpose is not to save keystrokes. Arrow functions have *lexical this* behavior, which is somewhat beyond the bounds of our discussion in this book.

Briefly: arrow functions don't define a `this` identifier keyword at all. If you use a `this` inside an arrow function, it behaves exactly as any other variable reference, which is that the scope chain is consulted to find a function scope (non-arrow function) where it is defined, and to use that one.

In other words, arrow functions treat `this` like any other lexical variable.

If you're used to hacks like `var self = this`, or if you prefer to call `.bind(this)` on inner `function` expressions, just to force them to inherit a `this` from an outer function like it was a lexical variable, then `=>` arrow functions are absolutely the better option. They're designed specifically to fix that problem.

So, in the rare cases you need *lexical this*, use an arrow function. It's the best tool for that job. But just be aware that in doing so, you're accepting the downsides of an anonymous function. You should expend additional effort to mitigate the readability cost, such as more descriptive variable names and code comments.

## IIFE Variations

All functions should have names. I said that a few times, right!? That includes IIFEs.

```
(function(){
    // don't do this!
})();

(function doThisInstead(){
    // ..
})();
```

How do we come up with a name for an IIFE? Identify what the IIFE is there for. Why do you need a scope in that spot? Are you hiding a cache variable for student records?

```
var getStudents = (function StoreStudentRecords(){
    var studentRecords = [];

    return function getStudents() {
        // ..
    }
})();
```

I named the IIFE `storeStudentRecords` because that's what it's doing: storing student records. Every IIFE should have a name. No exceptions.

IIFEs are typically defined by placing `( .. )` around the `function` expression, as shown in those previous snippets. But that's not the only way to define an IIFE. Technically, the only reason we're using that first surrounding set of `( .. )` is just so the `function` keyword isn't in a position to qualify as a `function` declaration to the JS parser. But there are other syntactic ways to avoid being parsed as a declaration:

```
!function thisIsAnIIFE(){
    // ..
}();

+function soIsThisOne(){
    // ..
}();

~function andThisOneToo(){
    // ..
}();
```

The `!`, `+`, `-`, and several other unary operators (operators with one operand) can all be placed in front of `function` to turn it into an expression. Then the final `()` call is valid, which makes it an IIFE.

I actually kind of like using the `void` unary operator when defining a standalone IIFE:

```
void function yepItsAnIIFE() {
  // ...
}();
```

The benefit of `void` is, it clearly communicates at the beginning of the function that this IIFE won't be returning any value.

However you define your IIFEs, show them some love by giving them names.

## Hoisting: Functions and Variables

Chapter 5 articulated both *function hoisting* and *variable hoisting*. Since hoisting is often cited as mistake in the design of JS, I wanted to briefly explore why both these forms of hoisting *can* be beneficial and should still be considered.

Give hoisting a deeper level of consideration by considering the merits of:

- Executable code first, function declarations last
- Semantic placement of variable declarations

### Function Hoisting

To review, this program works because of *function hoisting*:

```
getStudents();

// ...

function getStudents() {
  // ...
}
```

The `function` declaration is hoisted during compilation, which means that `getStudents` is an identifier declared for the entire scope. Additionally, the `getStudents` identifier is auto-initialized with the function reference, again at the beginning of the scope.

Why is this useful? The reason I prefer to take advantage of *function hoisting* is that it puts the *executable* code in any scope at the top, and any further declarations (functions) below. This means it's easier to find the code that will run in any given area, rather than having to scroll and scroll, hoping to find a trailing `}` marking the end of a scope/function somewhere.

I take advantage of this inverse positioning in all levels of scope:

```
getStudents();

// *****

function getStudents() {
  var whatever = doSomething();

  // other stuff

  return whatever;
}

// *****

function doSomething() {
```

```
// ...
}
```

When I first open a file like that, the very first line is executable code that kicks off its behavior. That's very easy to spot! Then, if I ever need to go find and inspect `getStudents()`, I like that its first line is also executable code. Only if I need to see the details of `doSomething()` do I go and find its definition down below.

In other words, I think *function hoisting* makes code more readable through a flowing, progressive reading order, from top to bottom.

## Variable Hoisting

What about *variable hoisting*?

Even though `let` and `const` hoist, you cannot use those variables in their TDZ (see Chapter 5). So, the following discussion only applies to `var` declarations. Before I continue, I'll admit: in almost all cases, I completely agree that *variable hoisting* is a bad idea:

```
pleaseDontDoThis = "bad idea";

// much later
var pleaseDontDoThis;
```

While that kind of inverted ordering was helpful for *function hoisting*, here I think it usually makes code harder to reason about.

But there's one exception that I've found, somewhat rarely, in my own coding. It has to do with where I place my `var` declarations inside a CommonJS module definition.

Here's how I typically structure my module definitions in Node:

```
// dependencies
var aModuleINeed = require("very-helpful");
var anotherModule = require("kinda-helpful");

// public API
var publicAPI = Object.assign(module.exports, {
  getStudents,
  addStudents,
  // ...
});

// ****
// private implementation

var cache = {};
var otherData = [];

function getStudents() {
  // ...
}

function addStudents() {
  // ...
}
```

Notice how the `cache` and `otherData` variables are in the "private" section of the module layout? That's because I don't plan to expose them publicly. So I organize the module so they're located alongside the other hidden implementation details of the module.

But I've had a few rare cases where I needed the assignments of those values to happen *above*, before I declare the exported public API of the module. For instance:

```
// public API
var publicAPI = Object.assign(module.exports, {
  getStudents,
  addStudents,
  refreshData: refreshData.bind(null,cache)
});
```

I need the `cache` variable to have already been assigned a value, because that value is used in the initialization of the public API (the `.bind(...)` partial-application).

Should I just move the `var cache = {}` up to the top, above this public API initialization? Well, perhaps. But now it's less obvious that `var cache` is a *private* implementation detail. Here's the compromise I've (somewhat rarely) used:

```
cache = {};
// used here, but declared below

// public API
var publicAPI = Object.assign(module.exports, {
  getStudents,
  addStudents,
  refreshData: refreshData.bind(null,cache)
});

// ****
// private implementation

var cache /* = {} */;
```

See the *variable hoisting*? I've declared the `cache` down where it belongs, logically, but in this rare case I've used it earlier up above, in the area where its initialization is needed. I even left a hint at the value that's assigned to `cache` in a code comment.

That's literally the only case I've ever found for leveraging *variable hoisting* to assign a variable earlier in a scope than its declaration. But I think it's a reasonable exception to employ with caution.

## The Case for `var`

Speaking of *variable hoisting*, let's have some real talk for a bit about `var`, a favorite villain devs love to blame for many of the woes of JS development. In Chapter 5, we explored `let` / `const` and promised we'd revisit where `var` falls in the whole mix.

As I lay out the case, don't miss:

- `var` was never broken
- `let` is your friend
- `const` has limited utility
- The best of both worlds: `var` and `let`

## Don't Throw Out `var`

`var` is fine, and works just fine. It's been around for 25 years, and it'll be around and useful and functional for another 25 years or more. Claims that `var` is broken, deprecated, outdated, dangerous, or ill-designed are bogus bandwagoning.

Does that mean `var` is the right declarator for every single declaration in your program? Certainly not. But it still has its place in your programs. Refusing to use it because someone on the team chose an aggressive linter opinion that chokes on `var` is cutting off your nose to spite your face.

OK, now that I've got you really riled up, let me try to explain my position.

For the record, I'm a fan of `let`, for block-scoped declarations. I really dislike TDZ and I think that was a mistake. But `let` itself is great. I use it often. In fact, I probably use it as much or more than I use `var`.

## `const` -antly Confused

`const` on the other hand, I don't use as often. I'm not going to dig into all the reasons why, but it comes down to `const` not *carrying its own weight*. That is, while there's a tiny bit of benefit of `const` in some cases, that benefit is outweighed by the long history of troubles around `const` confusion in a variety of languages, long before it ever showed up in JS.

`const` pretends to create values that can't be mutated—a misconception that's extremely common in developer communities across many languages—whereas what it really does is prevent re-assignment.

```
const studentIDs = [ 14, 73, 112 ];

// later

studentIDs.push(6); // whoa, wait... what!?
```

Using a `const` with a mutable value (like an array or object) is asking for a future developer (or reader of your code) to fall into the trap you set, which was that they either didn't know, or sorta forgot, that *value immutability* isn't at all the same thing as *assignment immutability*.

I just don't think we should set those traps. The only time I ever use `const` is when I'm assigning an already-immutable value (like `42` or `"Hello, friends!"`), and when it's clearly a "constant" in the sense of being a named placeholder for a literal value, for semantic purposes. That's what `const` is best used for. That's pretty rare in my code, though.

If variable re-assignment were a big deal, then `const` would be more useful. But variable re-assignment just isn't that big of a deal in terms of causing bugs. There's a long list of things that lead to bugs in programs, but "accidental re-assignment" is way, way down that list.

Combine that with the fact that `const` (and `let`) are supposed to be used in blocks, and blocks are supposed to be short, and you have a really small area of your code where a `const` declaration is even applicable. A `const` on line 1 of your ten-line block only tells you something about the next nine lines. And the thing it tells you is already obvious by glancing down at those nine lines: the variable is never on the left-hand side of an `=`; it's not re-assigned.

That's it, that's all `const` really does. Other than that, it's not very useful. Stacked up against to the significant confusion of value vs. assignment immutability, `const` loses a lot of its luster.

A `let` (or `var`!) that's never re-assigned is already behaviorally a "constant", even though it doesn't have the compiler guarantee. That's good enough in most cases.

## `var` and `let`

In my mind, `const` is pretty rarely useful, so this is only two-horse race between `let` and `var`. But it's not really a race either, because there doesn't have to be just one winner. They can both win... different races.

The fact is, you should be using both `var` and `let` in your programs. They are not interchangeable: you shouldn't use `var` where a `let` is called for, but you also shouldn't use `let` where a `var` is most appropriate.

So where should we still use `var`? Under what circumstances is it a better choice than `let`?

For one, I always use `var` in the top-level scope of any function, regardless of whether that's at the beginning, middle, or end of the function. I also use `var` in the global scope, though I try to minimize usage of the global scope.

Why use `var` for function scoping? Because that's exactly what `var` does. There literally is no better tool for the job of function scoping a declaration than a declarator that has, for 25 years, done exactly that.

You *could* use `let` in this top-level scope, but it's not the best tool for that job. I also find that if you use `let` everywhere, then it's less obvious which declarations are designed to be localized and which ones are intended to be used throughout the function.

By contrast, I rarely use a `var` inside a block. That's what `let` is for. Use the best tool for the job. If you see a `let`, it tells you that you're dealing with a localized declaration. If you see `var`, it tells you that you're dealing with a function-wide declaration. Simple as that.

```
function getStudents(data) {
  var studentRecords = [];

  for (let record of data.records) {
    let id = `student-${record.id}`;
    studentRecords.push({
      id,
      record.name
    });
  }

  return studentRecords;
}
```

The `studentRecords` variable is intended for use across the whole function. `var` is the best declarator to tell the reader that. By contrast, `record` and `id` are intended for use only in the narrower scope of the loop iteration, so `let` is the best tool for that job.

In addition to this *best tool* semantic argument, `var` has a few other characteristics that, in certain limited circumstances, make it more powerful.

One example is when a loop is exclusively using a variable, but its conditional clause cannot see block-scoped declarations inside the iteration:

```
function commitAction() {
  do {
    let result = commit();
    var done = result && result.code == 1;
  } while (!done);
}
```

Here, `result` is clearly only used inside the block, so we use `let`. But `done` is a bit different. It's only useful for the loop, but the `while` clause cannot see `let` declarations that appear inside the loop. So we compromise and use `var`, so that `done` is hoisted to the outer scope where it can be seen.

The alternative—declaring `done` outside the loop—separates it from where it's first used, and either necessitates picking a default value to assign, or worse, leaving it unassigned and thus looking ambiguous to the reader. I think `var` inside the loop is preferable here.

Another helpful characteristic of `var` is seen with declarations inside unintended blocks. Unintended blocks are blocks that are created because the syntax requires a block, but where the intent of the developer is not really to create a localized scope. The best illustration of unintended scope is the `try..catch` statement:

```
function getStudents() {
  try {
    // not really a block scope
    var records = fromCache("students");
  }
  catch (err) {
    // oops, fall back to a default
```

```

        var records = [];
    }
    // ...
}

```

There are other ways to structure this code, yes. But I think this is the *best* way, given various trade-offs.

I don't want to declare `records` (with `var` or `let`) outside of the `try` block, and then assign to it in one or both blocks. I prefer initial declarations to always be as close as possible (ideally, same line) to the first usage of the variable. In this simple example, that would only be a couple of lines distance, but in real code it can grow to many more lines. The bigger the gap, the harder it is to figure out what variable from what scope you're assigning to. `var` used at the actual assignment makes it less ambiguous.

Also notice I used `var` in both the `try` and `catch` blocks. That's because I want to signal to the reader that no matter which path is taken, `records` always gets declared. Technically, that works because `var` is hoisted once to the function scope. But it's still a nice semantic signal to remind the reader what either `var` ensures. If `var` were only used in one of the blocks, and you were only reading the other block, you wouldn't as easily discover where `records` was coming from.

This is, in my opinion, a little superpower of `var`. Not only can it escape the unintentional `try..catch` blocks, but it's allowed to appear multiple times in a function's scope. You can't do that with `let`. It's not bad, it's actually a little helpful feature. Think of `var` more like a declarative annotation that's reminding you, each usage, where the variable comes from. "Ah ha, right, it belongs to the whole function."

This repeated-annotation superpower is useful in other cases:

```

function getStudents() {
    var data = [];

    // do something with data
    // ... 50 more lines of code ...

    // purely an annotation to remind us
    var data;

    // use data again
    // ...
}

```

The second `var data` is not re-declaring `data`, it's just annotating for the readers' benefit that `data` is a function-wide declaration. That way, the reader doesn't need to scroll up 50+ lines of code to find the initial declaration.

I'm perfectly fine with re-using variables for multiple purposes throughout a function scope. I'm also perfectly fine with having two usages of a variable be separated by quite a few lines of code. In both cases, the ability to safely "re-declare" (annotate) with `var` helps make sure I can tell where my `data` is coming from, no matter where I am in the function.

Again, sadly, `let` cannot do this.

There are other nuances and scenarios when `var` turns out to offer some assistance, but I'm not going to belabor the point any further. The takeaway is that `var` can be useful in our programs alongside `let` (and the occasional `const`). Are you willing to creatively use the tools the JS language provides to tell a richer story to your readers?

Don't just throw away a useful tool like `var` because someone shamed you into thinking it wasn't cool anymore. Don't avoid `var` because you got confused once years ago. Learn these tools and use them each for what they're best at.

## What's the Deal with TDZ?

The TDZ (temporal dead zone) was explained in Chapter 5. We illustrated how it occurs, but we skimmed over any explanation of *why* it was necessary to introduce in the first place. Let's look briefly at the motivations of TDZ.

Some breadcrumbs in the TDZ origin story:

- `const` s should never change
- It's all about time
- Should `let` behave more like `const` or `var` ?

## Where It All Started

TDZ comes from `const`, actually.

During early ES6 development work, TC39 had to decide whether `const` (and `let`) were going to hoist to the top of their blocks. They decided these declarations would hoist, similar to how `var` does. Had that not been the case, I think some of the fear was confusion with mid-scope shadowing, such as:

```
let greeting = "Hi!";

{
    // what should print here?
    console.log(greeting);

    // .. a bunch of lines of code ..

    // now shadowing the `greeting` variable
    let greeting = "Hello, friends!";

    // ..
}
```

What should we do with that `console.log(..)` statement? Would it make any sense to JS devs for it to print "Hi!"? Seems like that could be a gotcha, to have shadowing kick in only for the second half of the block, but not the first half. That's not very intuitive, JS-like behavior. So `let` and `const` have to hoist to the top of the block, visible throughout.

But if `let` and `const` hoist to the top of the block (like `var` hoists to the top of a function), why don't `let` and `const` auto-initialize (to `undefined`) the way `var` does? Here was the main concern:

```
{
    // what should print here?
    console.log(studentName);

    // later

    const studentName = "Frank";

    // ..
}
```

Let's imagine that `studentName` not only hoisted to the top of this block, but was also auto-initialized to `undefined`. For the first half of the block, `studentName` could be observed to have the `undefined` value, such as with our `console.log(..)` statement. Once the `const studentName = ..` statement is reached, now `studentName` is assigned `"Frank"`. From that point forward, `studentName` can't ever be re-assigned.

But, is it strange or surprising that a constant observably has two different values, first `undefined`, then `"Frank"`? That does seem to go against what we think a `const` ant means; it should only ever be observable with one value.

So... now we have a problem. We can't auto-initialize `studentName` to `undefined` (or any other value for that matter). But the variable has to exist throughout the whole scope. What do we do with the period of time from when it first exists (beginning of scope) and when it's assigned its value?

We call this period of time the "dead zone," as in the "temporal dead zone" (TDZ). To prevent confusion, it was determined that any sort of access of a variable while in its TDZ is illegal and must result in the TDZ error.

OK, that line of reasoning does make some sense, I must admit.

## Who let the TDZ Out?

But that's just `const`. What about `let`?

Well, TC39 made the decision: since we need a TDZ for `const`, we might as well have a TDZ for `let` as well. *In fact, if we make let have a TDZ, then we discourage all that ugly variable hoisting people do.* So there was a consistency perspective and, perhaps, a bit of social engineering to shift developers' behavior.

My counter-argument would be: if you're favoring consistency, be consistent with `var` instead of `const`; `let` is definitely more like `var` than `const`. That's especially true since they had already chosen consistency with `var` for the whole hoisting-to-the-top-of-the-scope thing. Let `const` be its own unique deal with a TDZ, and let the answer to TDZ purely be: just avoid the TDZ by always declaring your constants at the top of the scope. I think this would have been more reasonable.

But alas, that's not how it landed. `let` has a TDZ because `const` needs a TDZ, because `let` and `const` mimic `var` in their hoisting to the top of the (block) scope. There ya go. Too circular? Read it again a few times.

## Are Synchronous Callbacks Still Closures?

Chapter 7 presented two different models for tackling closure:

- Closure is a function instance remembering its outer variables even as that function is passed around and **invoked in** other scopes.
- Closure is a function instance and its scope environment being preserved in-place while any references to it are passed around and **invoked from** other scopes.

These models are not wildly divergent, but they do approach from a different perspective. And that different perspective changes what we identify as a closure.

Don't get lost following this rabbit trail through closures and callbacks:

- Calling back to what (or where)?
- Maybe "synchronous callback" isn't the best label
- **IIF** functions don't move around, why would they need closure?
- Deferring over time is key to closure

## What is a Callback?

Before we revisit closure, let me spend a brief moment addressing the word "callback." It's a generally accepted norm that saying "callback" is synonymous with both *asynchronous callbacks* and *synchronous callbacks*. I don't think I agree that this is a good idea, so I want to explain why and propose we move away from that to another term.

Let's first consider an *asynchronous callback*, a function reference that will be invoked at some future *later* point. What does "callback" mean, in this case?

It means that the current code has finished or paused, suspended itself, and that when the function in question is invoked later, execution is entering back into the suspended program, resuming it. Specifically, the point of re-entry is the code that was wrapped in the function reference:

```
setTimeout(function waitForASecond(){
    // this is where JS should call back into
    // the program when the timer has elapsed
}, 1000);

// this is where the current program finishes
// or suspends
```

In this context, "calling back" makes a lot of sense. The JS engine is resuming our suspended program by *calling back in* at a specific location. OK, so a callback is asynchronous.

## Synchronous Callback?

But what about *synchronous callbacks*? Consider:

```
function getLabels(studentIDs) {
    return studentIDs.map(
        function formatIDLabel(id){
            return `Student ID: ${String(id).padStart(6)}
        }
    );
}

getLabels([ 14, 73, 112, 6 ]);
// [
//   "Student ID: 000014",
//   "Student ID: 000073",
//   "Student ID: 000112",
//   "Student ID: 000006"
// ]
```

Should we refer to `formatIDLabel(..)` as a callback? Is the `map(..)` utility really *calling back* into our program by invoking the function we provided?

There's nothing to *call back into* per se, because the program hasn't paused or exited. We're passing a function (reference) from one part of the program to another part of the program, and then it's immediately invoked.

There's other established terms that might match what we're doing—passing in a function (reference) so that another part of the program can invoke it on our behalf. You might think of this as *Dependency Injection* (DI) or *Inversion of Control* (IoC).

DI can be summarized as passing in necessary part(s) of functionality to another part of the program so that it can invoke them to complete its work. That's a decent description for the `map(..)` call above, isn't it? The `map(..)` utility knows to iterate over the list's values, but it doesn't know what to *do* with those values. That's why we pass it the `formatIDLabel(..)` function. We pass in the dependency.

IoC is a pretty similar, related concept. Inversion of control means that instead of the current area of your program controlling what's happening, you hand control off to another part of the program. We wrapped the logic for computing a label string in the function `formatIDLabel(..)`, then handed invocation control to the `map(..)` utility.

Notably, Martin Fowler cites IoC as the difference between a framework and a library: with a library, you call its functions; with a framework, it calls your functions. [fowlerIOC](#)

In the context of our discussion, either DI or IoC could work as an alternative label for a *synchronous callback*.

But I have a different suggestion. Let's refer to (the functions formerly known as) *synchronous callbacks*, as *inter-invoked functions* (IIFs). Yes, exactly, I'm playing off IIFEs. These kinds of functions are *inter-invoked*, meaning: another entity invokes them, as opposed to IIFEs, which invoke themselves immediately.

What's the relationship between an *asynchronous callback* and an IIF? An *asynchronous callback* is an IIF that's invoked asynchronously instead of synchronously.

## Synchronous Closure?

Now that we've re-labeled *synchronous callbacks* as IIFs, we can return to our main question: are IIFs an example of closure? Obviously, the IIF would have to reference variable(s) from an outer scope for it to have any chance of being a closure. The `formatIDLabel(...)` IIF from earlier does not reference any variables outside its own scope, so it's definitely not a closure.

What about an IIF that does have external references, is that closure?

```
function printLabels(labels) {
  var list = document.getElementById("labelsList");

  labels.forEach(
    function renderLabel(label){
      var li = document.createElement("li");
      li.innerText = label;
      list.appendChild(li);
    }
  );
}
```

The inner `renderLabel(...)` IIF references `list` from the enclosing scope, so it's an IIF that *could* have closure. But here's where the definition/model we choose for closure matters:

- If `renderLabel(...)` is a **function that gets passed somewhere else**, and that function is then invoked, then yes, `renderLabel(...)` is exercising a closure, because closure is what preserved its access to its original scope chain.
- But if, as in the alternative conceptual model from Chapter 7, `renderLabel(...)` stays in place, and only a reference to it is passed to `forEach(...)`, is there any need for closure to preserve the scope chain of `renderLabel(...)`, while it executes synchronously right inside its own scope?

No. That's just normal lexical scope.

To understand why, consider this alternative form of `printLabels(..)`:

```
function printLabels(labels) {
  var list = document.getElementById("labelsList");

  for (let label of labels) {
    // just a normal function call in its own
    // scope, right? That's not really closure!
    renderLabel(label);
  }

  // *****
}

function renderLabel(label) {
  var li = document.createElement("li");
  li.innerText = label;
  list.appendChild(li);
}
```

These two versions of `printLabels(..)` are essentially the same.

The latter one is definitely not an example of closure, at least not in any useful or observable sense. It's just lexical scope. The former version, with `forEach(..)` calling our function reference, is essentially the same thing. It's also not closure, but rather just a plain ol' lexical scope function call.

## Defer to Closure

By the way, Chapter 7 briefly mentioned partial application and currying (which *do* rely on closure!). This is a interesting scenario where manual currying can be used:

```
function printLabels(labels) {
  var list = document.getElementById("labelsList");
  var renderLabel = renderTo(list);

  // definitely closure this time!
  labels.forEach( renderLabel(label) );

  // *****
}

function renderTo(list) {
  return function createLabel(label){
    var li = document.createElement("li");
    li.innerText = label;
    list.appendChild(li);
  };
}
}
```

The inner function `createLabel(..)`, which we assign to `renderLabel`, is closed over `list`, so closure is definitely being utilized.

Closure allows us to remember `list` for later, while we defer execution of the actual label-creation logic from the `renderTo(..)` call to the subsequent `forEach(..)` invocations of the `createLabel(..)` IIF. That may only be a brief moment here, but any amount of time could pass, as closure bridges from call to call.

## Classic Module Variations

Chapter 8 explained the classic module pattern, which can look like this:

```
var StudentList = (function defineModule(Student){
  var elems = [];

  var publicAPI = {
    renderList() {
      // ..
    }
  };

  return publicAPI;
})(Student);
```

Notice that we're passing `Student` (another module instance) in as a dependency. But there's lots of useful variations on this module form you may encounter. Some hints for recognizing these variations:

- Does the module know about its own API?
- Even if we use a fancy module loader, it's just a classic module
- Some modules need to work universally

## Where's My API?

First, most classic modules don't define and use a `publicAPI` the way I have shown in this code. Instead, they typically look like:

```
var StudentList = (function defineModule(Student){
  var elems = [];

  return {
    renderList() {
      // ...
    }
  };
})(Student);
```

The only difference here is directly returning the object that serves as the public API for the module, as opposed to first saving it to an inner `publicAPI` variable. This is by far how most classic modules are defined.

But I strongly prefer, and always use myself, the former `publicAPI` form. Two reasons:

- `publicAPI` is a semantic descriptor that aids readability by making it more obvious what the purpose of the object is.
- Storing an inner `publicAPI` variable that references the same external public API object returned, can be useful if you need to access or modify the API during the lifetime of the module.

For example, you may want to call one of the publicly exposed functions, from inside the module. Or, you may want to add or remove methods depending on certain conditions, or update the value of an exposed property.

Whatever the case may be, it just seems rather silly to me that we *wouldn't* maintain a reference to access our own API. Right?

## Asynchronous Module Definition (AMD)

Another variation on the classic module form is AMD-style modules (popular several years back), such as those supported by the RequireJS utility:

```
define([ "./Student" ], function StudentList(Student){
  var elems = [];

  return {
    renderList() {
      // ...
    }
  };
});
```

If you look closely at `StudentList(..)`, it's a classic module factory function. Inside the machinery of `define(..)` (provided by RequireJS), the `StudentList(..)` function is executed, passing to it any other module instances declared as dependencies. The return value is an object representing the public API for the module.

This is based on exactly the same principles (including how the closure works!) as we explored with classic modules.

## Universal Modules (UMD)

The final variation we'll look at is UMD, which is less a specific, exact format and more a collection of very similar formats. It was designed to create better interop (without any build-tool conversion) for modules that may be loaded in browsers, by AMD-style loaders, or in Node. I personally still publish many of my utility libraries using a form of UMD.

Here's the typical structure of a UMD:

```
(function UMD(name,context,definition){
    // loaded by an AMD-style loader?
    if (
        typeof define === "function" &&
        define.amd
    ) {
        define(definition);
    }
    // in Node?
    else if (
        typeof module !== "undefined" &&
        module.exports
    ) {
        module.exports = definition(name,context);
    }
    // assume standalone browser script
    else {
        context[name] = definition(name,context);
    }
})( "StudentList",this,function DEF(name,context){
    var elems = [];

    return {
        renderList() {
            // ...
        }
    };
});
```

Though it may look a bit unusual, UMD is really just an IIFE.

What's different is that the main `function` expression part (at the top) of the IIFE contains a series of `if..else if` statements to detect which of the three supported environments the module is being loaded in.

The final `()` that normally invokes an IIFE is being passed three arguments: `"StudentsList"`, `this`, and another `function` expression. If you match those arguments to their parameters, you'll see they are: `name`, `context`, and `definition`, respectively. `"StudentList"` (`name`) is the name label for the module, primarily in case it's defined as a global variable. `this` (`context`) is generally the `window` (aka, global object; see Chapter 4) for defining the module by its name.

`definition(..)` is invoked to actually retrieve the definition of the module, and you'll notice that, sure enough, that's just a classic module form!

There's no question that as of the time of this writing, ESM (ES Modules) are becoming popular and widespread rapidly. But with millions and millions of modules written over the last 20 years, all using some pre-ESM variation of classic modules, they're still very important to be able to read and understand when you come across them.

fowlerIOC. *Inversion of Control*, Martin Fowler, <https://martinfowler.com/bliki/InversionOfControl.html>, 26 June 2005. ↵

# You Don't Know JS Yet: Scope & Closures - 2nd Edition

## Appendix B: Practice

This appendix aims to give you some challenging and interesting exercises to test and solidify your understanding of the main topics from this book. It's a good idea to try out the exercises yourself—in an actual code editor!—instead of skipping straight to the solutions at the end. No cheating!

These exercises don't have a specific right answer that you have to get exactly. Your approach may differ some (or a lot!) from the solutions presented, and that's OK.

There's no judging you on how you write your code. My hope is that you come away from this book feeling confident that you can tackle these sorts of coding tasks built on a strong foundation of knowledge. That's the only objective, here. If you're happy with your code, I am, too!

### Buckets of Marbles

Remember Figure 2 from back in Chapter 2?

```

1 var students = [
2   { id: 14, name: "Kyle" },
3   { id: 73, name: "Suzy" },
4   { id: 112, name: "Frank" },
5   { id: 6, name: "Sarah" }
6 ];
7
8 function getStudentName(studentID) {
9   for (let student of students) {
10     if (student.id == studentID) {
11       return student.name;
12     }
13   }
14 }
15
16 var nextStudent = getStudentName(73);
17
18 console.log(nextStudent);
19 // "Suzy"

```

*Fig. 2 (Ch. 2): Colored Scope Bubbles*

This exercise asks you to write a program—any program!—that contains nested functions and block scopes, which satisfies these constraints:

- If you color all the scopes (including the global scope!) different colors, you need at least six colors. Make sure to add a code comment labeling each scope with its color.

BONUS: identify any implied scopes your code may have.

- Each scope has at least one identifier.

- Contains at least two function scopes and at least two block scopes.
- At least one variable from an outer scope must be shadowed by a nested scope variable (see Chapter 3).
- At least one variable reference must resolve to a variable declaration at least two levels higher in the scope chain.

**TIP:**

You *can* just write junk foo/bar/baz-type code for this exercise, but I suggest you try to come up with some sort of non-trivial real-ish code that at least does something kind of reasonable.

Try the exercise for yourself, then check out the suggested solution at the end of this appendix.

## Closure (PART 1)

Let's first practice closure with some common computer-math operations: determining if a value is prime (has no divisors other than 1 and itself), and generating a list of prime factors (divisors) for a given number.

For example:

```
isPrime(11);      // true
isPrime(12);      // false

factorize(11);    // [ 11 ]
factorize(12);    // [ 3, 2, 2 ] --> 3*2*2=12
```

Here's an implementation of `isPrime(..)`, adapted from the Math.js library: [MathJSisPrime](#)

```
function isPrime(v) {
  if (v <= 3) {
    return v > 1;
  }
  if (v % 2 == 0 || v % 3 == 0) {
    return false;
  }
  var vSqrt = Math.sqrt(v);
  for (let i = 5; i <= vSqrt; i += 6) {
    if (v % i == 0 || v % (i + 2) == 0) {
      return false;
    }
  }
  return true;
}
```

And here's a somewhat basic implementation of `factorize(..)` (not to be confused with `factorial(..)` from Chapter 6):

```
function factorize(v) {
  if (!isPrime(v)) {
    let i = Math.floor(Math.sqrt(v));
    while (v % i != 0) {
      i--;
    }
    return [
      ...factorize(i),
      ...factorize(v / i)
    ];
  }
  return [v];
}
```

**NOTE:**

I call this basic because it's not optimized for performance. It's binary-recursive (which isn't tail-call optimizable), and it creates a lot of intermediate array copies. It also doesn't order the discovered factors in any way. There are many, many other algorithms for this task, but I wanted to use something short and roughly understandable for our exercise.

If you were to call `isPrime(4327)` multiple times in a program, you can see that it would go through all its dozens of comparison/computation steps every time. If you consider `factorize(..)`, it's calling `isPrime(..)` many times as it computes the list of factors. And there's a good chance most of those calls are repeats. That's a lot of wasted work!

The first part of this exercise is to use closure to implement a cache to remember the results of `isPrime(..)`, so that the primality (`true` or `false`) of a given number is only ever computed once. Hint: we already showed this sort of caching in Chapter 6 with `factorial(..)`.

If you look at `factorize(..)`, it's implemented with recursion, meaning it calls itself repeatedly. That again means we may likely see a lot of wasted calls to compute prime factors for the same number. So the second part of the exercise is to use the same closure cache technique for `factorize(..)`.

Use separate closures for caching of `isPrime(..)` and `factorize(..)`, rather than putting them inside a single scope.

Try the exercise for yourself, then check out the suggested solution at the end of this appendix.

## A Word About Memory

I want to share a little quick note about this closure cache technique and the impacts it has on your application's performance.

We can see that in saving the repeated calls, we improve computation speed (in some cases, by a dramatic amount). But this usage of closure is making an explicit trade-off that you should be very aware of.

The trade-off is memory. We're essentially growing our cache (in memory) unboundedly. If the functions in question were called many millions of times with mostly unique inputs, we'd be chewing up a lot of memory. This can definitely be worth the expense, but only if we think it's likely we see repetition of common inputs so that we're taking advantage of the cache.

If most every call will have a unique input, and the cache is essentially never *used* to any benefit, this is an inappropriate technique to employ.

It also might be a good idea to have a more sophisticated caching approach, such as an LRU (least recently used) cache, that limits its size; as it runs up to the limit, an LRU evicts the values that are... well, least recently used!

The downside here is that LRU is quite non-trivial in its own right. You'll want to use a highly optimized implementation of LRU, and be keenly aware of all the trade-offs at play.

## Closure (PART 2)

In this exercise, we're going to again practice closure by defining a `toggle(..)` utility that gives us a value toggler.

You will pass one or more values (as arguments) into `toggle(..)`, and get back a function. That returned function will alternate/rotate between all the passed-in values in order, one at a time, as it's called repeatedly.

```
function toggle(* .. *) {
  // ..
}

var hello = toggle("hello");
var onOff = toggle("on", "off");
var speed = toggle("slow", "medium", "fast");

hello();      // "hello"
hello();      // "hello"
```

```

onOff();      // "on"
onOff();      // "off"
onOff();      // "on"

speed();      // "slow"
speed();      // "medium"
speed();      // "fast"
speed();      // "slow"

```

The corner case of passing in no values to `toggle(...)` is not very important; such a toggler instance could just always return `undefined`.

Try the exercise for yourself, then check out the suggested solution at the end of this appendix.

## Closure (PART 3)

In this third and final exercise on closure, we're going to implement a basic calculator. The `calculator()` function will produce an instance of a calculator that maintains its own state, in the form of a function (`calc(...)`, below):

```

function calculator() {
  // ...
}

var calc = calculator();

```

Each time `calc(...)` is called, you'll pass in a single character that represents a keypress of a calculator button. To keep things more straightforward, we'll restrict our calculator to supporting entering only digits (0-9), arithmetic operations (+, -, \*, /), and "=" to compute the operation. Operations are processed strictly in the order entered; there's no "( )" grouping or operator precedence.

We don't support entering decimals, but the divide operation can result in them. We don't support entering negative numbers, but the "-" operation can result in them. So, you should be able to produce any negative or decimal number by first entering an operation to compute it. You can then keep computing with that value.

The return of `calc(...)` calls should mimic what would be shown on a real calculator, like reflecting what was just pressed, or computing the total when pressing "=".

For example:

```

calc("4");      // 4
calc("+");     // +
calc("7");      // 7
calc("3");      // 3
calc("-");     // -
calc("2");      // 2
calc("=");     // 75
calc("*");     // *
calc("4");      // 4
calc("=");     // 300
calc("5");      // 5
calc("-");     // -
calc("5");      // 5
calc("=");     // 0

```

Since this usage is a bit clumsy, here's a `useCalc(...)` helper, that runs the calculator with characters one at a time from a string, and computes the display each time:

```

function useCalc(calc, keys) {
  return [...keys].reduce(

```

```

function showDisplay(display, key) {
    var ret = String( calc(key) );
    return (
        display +
        (
            (ret != "" && key == "=") ?
                "=" :
                ""
        ) +
        ret
    );
}

useCalc(calc, "4+3=");
useCalc(calc, "+9=");
useCalc(calc, "*8=");
useCalc(calc, "7*2*3=");
useCalc(calc, "1/0=");
useCalc(calc, "+3=");
useCalc(calc, "51=");

```

The most sensible usage of this `useCalc(...)` helper is to always have `"+"` be the last character entered.

Some of the formatting of the totals displayed by the calculator require special handling. I'm providing this `formatTotal(..)` function, which your calculator should use whenever it's going to return a current computed total (after an `"+"` is entered):

```

function formatTotal(display) {
    if (Number.isFinite(display)) {
        // constrain display to max 11 chars
        let maxDigits = 11;
        // reserve space for "e+" notation?
        if (Math.abs(display) > 99999999999) {
            maxDigits -= 6;
        }
        // reserve space for "-"?
        if (display < 0) {
            maxDigits--;
        }

        // whole number?
        if (Number.isInteger(display)) {
            display = display
                .toPrecision(maxDigits)
                .replace(/\.\d+$/,"");
        }
        // decimal
        else {
            // reserve space for "."
            maxDigits--;
            // reserve space for leading "0"?
            if (
                Math.abs(display) >= 0 &&
                Math.abs(display) < 1
            ) {
                maxDigits--;
            }
            display = display
                .toPrecision(maxDigits)
                .replace(/\d+$/,"");
        }
    }
    else {
        display = "ERR";
    }
}

```

```
    return display;
}
```

Don't worry too much about how `formatTotal(..)` works. Most of its logic is a bunch of handling to limit the calculator display to 11 characters max, even if negatives, repeating decimals, or even "e+" exponential notation is required.

Again, don't get too mired in the mud around calculator-specific behavior. Focus on the *memory* of closure.

Try the exercise for yourself, then check out the suggested solution at the end of this appendix.

## Modules

This exercise is to convert the calculator from Closure (PART 3) into a module.

We're not adding any additional functionality to the calculator, only changing its interface. Instead of calling a single function `calc(..)`, we'll be calling specific methods on the public API for each "keypress" of our calculator. The outputs stay the same.

This module should be expressed as a classic module factory function called `calculator()`, instead of a singleton IIFE, so that multiple calculators can be created if desired.

The public API should include the following methods:

- `number(..)` (input: the character/number "pressed")
- `plus()`
- `minus()`
- `mult()`
- `div()`
- `eq()`

Usage would look like:

```
var calc = calculator();

calc.number("4");      // 4
calc.plus();          // +
calc.number("7");      // 7
calc.number("3");      // 3
calc.minus();          // -
calc.number("2");      // 2
calc.eq();             // 75
```

`formatTotal(..)` remains the same from that previous exercise. But the `useCalc(..)` helper needs to be adjusted to work with the module API:

```
function useCalc(calc, keys) {
  var keyMappings = {
    "+": "plus",
    "-": "minus",
    "*": "mult",
    "/": "div",
    "=": "eq"
  };

  return [...keys].reduce(
    function showDisplay(display, key){
      var fn = keyMappings[key] || "number";
      var ret = String( calc[fn](key) );
      return (
        display +
        (
          ret
        )
      );
    }
  );
}
```

```

        (ret != "" && key == "=") ?
        "=" :
        ""
    ) +
    ret
);
},
""
);
}

useCalc(calc, "4+3=");
useCalc(calc, "+9=");
useCalc(calc, "*8=");
useCalc(calc, "7*2*3=");
useCalc(calc, "1/0=");
useCalc(calc, "+3=");
useCalc(calc, "51=");
// 4+3=7
// +9=16
// *5=128
// 7*2*3=42
// 1/0=ERR
// +3=ERR
// 51=

```

Try the exercise for yourself, then check out the suggested solution at the end of this appendix.

As you work on this exercise, also spend some time considering the pros/cons of representing the calculator as a module as opposed to the closure-function approach from the previous exercise.

BONUS: write out a few sentences explaining your thoughts.

BONUS #2: try converting your module to other module formats, including: UMD, CommonJS, and ESM (ES Modules).

## Suggested Solutions

Hopefully you've tried out the exercises before you're reading this far. No cheating!

Remember, each suggested solution is just one of a bunch of different ways to approach the problems. They're not "the right answer," but they do illustrate a reasonable way to approach each exercise.

The most important benefit you can get from reading these suggested solutions is to compare them to your code and analyze why we each made similar or different choices. Don't get into too much bikeshedding; try to stay focused on the main topic rather than the small details.

### Suggested: Buckets of Marbles

The *Buckets of Marbles Exercise* can be solved like this:

```

// RED(1)
const howMany = 100;

// Sieve of Eratosthenes
function findPrimes(howMany) {
    // BLUE(2)
    var sieve = Array(howMany).fill(true);
    var max = Math.sqrt(howMany);

    for (let i = 2; i < max; i++) {
        // GREEN(3)
        if (sieve[i]) {
            // ORANGE(4)
            let j = Math.pow(i, 2);
            for (let k = j; k < howMany; k += i) {
                // PURPLE(5)
                sieve[k] = false;
            }
        }
    }
}

```

```

    }

    return sieve
      .map(function getPrime(flag, prime){
        // PINK(6)
        if (flag) return prime;
        return flag;
      })
      .filter(function onlyPrimes(v){
        // YELLOW(7)
        return !!v;
      })
      .slice(1);
}

findPrimes(howMany);
// [
//   2, 3, 5, 7, 11, 13, 17,
//   19, 23, 29, 31, 37, 41,
//   43, 47, 53, 59, 61, 67,
//   71, 73, 79, 83, 89, 97
// ]

```

## Suggested: Closure (PART 1)

The *Closure Exercise (PART 1)* for `isPrime(..)` and `factorize(..)`, can be solved like this:

```

var isPrime = (function isPrime(v){
  var primes = {};

  return function isPrime(v) {
    if (v in primes) {
      return primes[v];
    }
    if (v <= 3) {
      return (primes[v] = v > 1);
    }
    if (v % 2 == 0 || v % 3 == 0) {
      return (primes[v] = false);
    }
    let vSqrt = Math.sqrt(v);
    for (let i = 5; i <= vSqrt; i += 6) {
      if (v % i == 0 || v % (i + 2) == 0) {
        return (primes[v] = false);
      }
    }
    return (primes[v] = true);
  };
})();

var factorize = (function factorize(v){
  var factors = {};

  return function findFactors(v) {
    if (v in factors) {
      return factors[v];
    }
    if (!isPrime(v)) {
      let i = Math.floor(Math.sqrt(v));
      while (v % i != 0) {
        i--;
      }
      return (factors[v] = [
        ...findFactors(i),
        ...findFactors(v / i)
      ]);
    }
  }
});

```

```

        return (factors[v] = [v]);
    };
})();

```

The general steps I used for each utility:

1. Wrap an IIFE to define the scope for the cache variable to reside.
2. In the underlying call, first check the cache, and if a result is already known, return.
3. At each place where a `return` was happening originally, assign to the cache and just return the results of that assignment operation—this is a space savings trick mostly just for brevity in the book.

I also renamed the inner function from `factorize(...)` to `findFactors(...)`. That's not technically necessary, but it helps it make clearer which function the recursive calls invoke.

## Suggested: Closure (PART 2)

The *Closure Exercise (PART 2)* `toggle(...)` can be solved like this:

```

function toggle(...vals) {
  var unset = {};
  var cur = unset;

  return function next(){
    // save previous value back at
    // the end of the list
    if (cur != unset) {
      vals.push(cur);
    }
    cur = vals.shift();
    return cur;
  };
}

var hello = toggle("hello");
var onOff = toggle("on", "off");
var speed = toggle("slow", "medium", "fast");

hello();      // "hello"
hello();      // "hello"

onOff();      // "on"
onOff();      // "off"
onOff();      // "on"

speed();      // "slow"
speed();      // "medium"
speed();      // "fast"
speed();      // "slow"

```

## Suggested: Closure (PART 3)

The *Closure Exercise (PART 3)* `calculator()` can be solved like this:

```

// from earlier:
//
// function useCalc(..) { .. }
// function formatTotal(..) { .. }

function calculator() {
  var currentTotal = 0;
  var currentVal = "";

```

```

var currentOper = "=";

return pressKey;

// *****

function pressKey(key){
    // number key?
    if (/^\d/.test(key)) {
        currentVal += key;
        return key;
    }
    // operator key?
    else if (/[+*-/]/.test(key)) {
        // multiple operations in a series?
        if (
            currentOper != "=" &&
            currentVal != ""
        ) {
            // implied '=' keypress
            pressKey("=");
        }
        else if (currentVal != "") {
            currentTotal = Number(currentVal);
        }
        currentOper = key;
        currentVal = "";
        return key;
    }
    // = key?
    else if (
        key == "=" &&
        currentOper != "="
    ) {
        currentTotal = op(
            currentTotal,
            currentOper,
            Number(currentVal)
        );
        currentOper = "=";
        currentVal = "";
        return formatTotal(currentTotal);
    }
    return "";
};

function op(val1,oper,val2) {
    var ops = {
        // NOTE: using arrow functions
        // only for brevity in the book
        "+": (v1,v2) => v1 + v2,
        "-": (v1,v2) => v1 - v2,
        "*": (v1,v2) => v1 * v2,
        "/": (v1,v2) => v1 / v2
    };
    return ops[oper](val1,val2);
}

var calc = calculator();

useCalc(calc,"4+3=");           // 4+3=7
useCalc(calc,"+9=");           // +9=16
useCalc(calc,"*8=");           // *5=128
useCalc(calc,"7*2*3=");         // 7*2*3=42
useCalc(calc,"1/0=");           // 1/0=ERR
useCalc(calc,"+3=");           // +3=ERR
useCalc(calc,"51=");           // 51

```

**NOTE:**

Remember: this exercise is about closure. Don't focus too much on the actual mechanics of a calculator, but rather on whether you are properly *remembering* the calculator state across function calls.

**Suggested: Modules**

The *Modules Exercise* `calculator()` can be solved like this:

```
// from earlier:
//
// function useCalc(..) { .. }
// function formatTotal(..) { .. }

function calculator() {
    var currentTotal = 0;
    var currentVal = "";
    var currentOper = "=";

    var publicAPI = {
        number,
        eq,
        plus() { return operator("+"); },
        minus() { return operator("-"); },
        mult() { return operator("*"); },
        div() { return operator("/"); }
    };

    return publicAPI;
}

// *****

function number(key) {
    // number key?
    if (/^\d/.test(key)) {
        currentVal += key;
        return key;
    }
}

function eq() {
    // = key?
    if (currentOper != "=") {
        currentTotal = op(
            currentTotal,
            currentOper,
            Number(currentVal)
        );
        currentOper = "=";
        currentVal = "";
        return formatTotal(currentTotal);
    }
    return "";
}

function operator(key) {
    // multiple operations in a series?
    if (
        currentOper != "=" &&
        currentVal != ""
    ) {
        // implied '=' keypress
        eq();
    }
    else if (currentVal != "") {
        currentTotal = Number(currentVal);
    }
}
```

```

        }
        currentOper = key;
        currentValue = "";
        return key;
    }

    function op(val1,oper,val2) {
        var ops = {
            // NOTE: using arrow functions
            // only for brevity in the book
            "+": (v1,v2) => v1 + v2,
            "-": (v1,v2) => v1 - v2,
            "*": (v1,v2) => v1 * v2,
            "/": (v1,v2) => v1 / v2
        };
        return ops[oper](val1,val2);
    }

    var calc = calculator();

    useCalc(calc,"4+3=");
    useCalc(calc,"+9=");
    useCalc(calc,"*8=");
    useCalc(calc,"7*2*3=");
    useCalc(calc,"1/0=");
    useCalc(calc,"+3=");
    useCalc(calc,"51=");

```

That's it for this book, congratulations on your achievement! When you're ready, move on to Book 3, *Objects & Classes*.

MathJSisPrime . *Math.js: isPrime(..)*, <https://github.com/josdejong/mathjs/blob/develop/src/function/utils/isPrime.js>, 3

March 2020. ↪

# You Don't Know JS Yet: Objects & Classes - 2nd Edition

<b>NOTE:</b>
Work in progress

## Table of Contents

- Foreword (by TBA)
- Preface
- Chapter 1: `this` Or That?
- Chapter 2: `this` All Makes Sense Now!
- Chapter 3: Objects
- Chapter 4: Mixing (Up) "Class" Objects
- Chapter 5: Prototypes
- Chapter 6: Behavior Delegation
- Appendix A: ES6 `class`

# You Don't Know JS Yet: Objects & Classes - 2nd Edition

## Chapter 1: this Or That?

<b>NOTE:</b>
Work in progress

<b>NOTE:</b>
Everything below here is previous text from 1st edition, and is only here for reference while 2nd edition work is underway. <b>Please ignore this stuff.</b>

Everything below here is previous text from 1st edition, and is only here for reference while 2nd edition work is underway.  
**Please ignore this stuff.**

One of the most confused mechanisms in JavaScript is the `this` keyword. It's a special identifier keyword that's automatically defined in the scope of every function, but what exactly it refers to bedevils even seasoned JavaScript developers.

Any sufficiently *advanced* technology is indistinguishable from magic. -- Arthur C. Clarke

JavaScript's `this` mechanism isn't actually *that* advanced, but developers often paraphrase that quote in their own mind by inserting "complex" or "confusing", and there's no question that without lack of clear understanding, `this` can seem downright magical in *your* confusion.

**Note:** The word "this" is a terribly common pronoun in general discourse. So, it can be very difficult, especially verbally, to determine whether we are using "this" as a pronoun or using it to refer to the actual keyword identifier. For clarity, I will always use `this` to refer to the special keyword, and "this" or *this* otherwise.

## Why `this` ?

If the `this` mechanism is so confusing, even to seasoned JavaScript developers, one may wonder why it's even useful? Is it more trouble than it's worth? Before we jump into the *how*, we should examine the *why*.

Let's try to illustrate the motivation and utility of `this` :

```
function identify() {
    return this.name.toUpperCase();
}

function speak() {
    var greeting = "Hello, I'm " + identify.call( this );
    console.log( greeting );
}
```

```

}

var me = {
  name: "Kyle"
};

var you = {
  name: "Reader"
};

identify.call( me ); // KYLE
identify.call( you ); // READER

speak.call( me ); // Hello, I'm KYLE
speak.call( you ); // Hello, I'm READER

```

If the *how* of this snippet confuses you, don't worry! We'll get to that shortly. Just set those questions aside briefly so we can look into the *why* more clearly.

This code snippet allows the `identify()` and `speak()` functions to be re-used against multiple *context* (`me` and `you`) objects, rather than needing a separate version of the function for each object.

Instead of relying on `this`, you could have explicitly passed in a context object to both `identify()` and `speak()`.

```

function identify(context) {
  return context.name.toUpperCase();
}

function speak(context) {
  var greeting = "Hello, I'm " + identify( context );
  console.log( greeting );
}

identify( you ); // READER
speak( me ); // Hello, I'm KYLE

```

However, the `this` mechanism provides a more elegant way of implicitly "passing along" an object reference, leading to cleaner API design and easier re-use.

The more complex your usage pattern is, the more clearly you'll see that passing context around as an explicit parameter is often messier than passing around a `this` context. When we explore objects and prototypes, you will see the helpfulness of a collection of functions being able to automatically reference the proper context object.

## Confusions

We'll soon begin to explain how `this` *actually* works, but first we must dispel some misconceptions about how it *doesn't* actually work.

The name "this" creates confusion when developers try to think about it too literally. There are two meanings often assumed, but both are incorrect.

### Itself

The first common temptation is to assume `this` refers to the function itself. That's a reasonable grammatical inference, at least.

Why would you want to refer to a function from inside itself? The most common reasons would be things like recursion (calling a function from inside itself) or having an event handler that can unbind itself when it's first called.

Developers new to JS's mechanisms often think that referencing the function as an object (all functions in JavaScript are objects!) lets you store *state* (values in properties) between function calls. While this is certainly possible and has some limited uses, the rest of the book will expound on many other patterns for *better* places to store state besides the function object.

But for just a moment, we'll explore that pattern, to illustrate how `this` doesn't let a function get a reference to itself like we might have assumed.

Consider the following code, where we attempt to track how many times a function (`foo`) was called:

```
function foo(num) {
  console.log("foo: " + num);

  // keep track of how many times `foo` is called
  this.count++;
}

foo.count = 0;

var i;

for (i=0; i<10; i++) {
  if (i > 5) {
    foo(i);
  }
}
// foo: 6
// foo: 7
// foo: 8
// foo: 9

// how many times was `foo` called?
console.log(foo.count); // 0 -- WTF?
```

`foo.count` is still `0`, even though the four `console.log` statements clearly indicate `foo(..)` was in fact called four times. The frustration stems from a *too literal* interpretation of what `this` (in `this.count++`) means.

When the code executes `foo.count = 0`, indeed it's adding a property `count` to the function object `foo`. But for the `this.count` reference inside of the function, `this` is not in fact pointing *at all* to that function object, and so even though the property names are the same, the root objects are different, and confusion ensues.

**Note:** A responsible developer *should* ask at this point, "If I was incrementing a `count` property but it wasn't the one I expected, which `count` was I incrementing?" In fact, were she to dig deeper, she would find that she had accidentally created a global variable `count` (see Chapter 2 for *how* that happened!), and it currently has the value `Nan`. Of course, once she identifies this peculiar outcome, she then has a whole other set of questions: "How was it global, and why did it end up `Nan` instead of some proper count value?" (see Chapter 2).

Instead of stopping at this point and digging into why the `this` reference doesn't seem to be behaving as *expected*, and answering those tough but important questions, many developers simply avoid the issue altogether, and hack toward some other solution, such as creating another object to hold the `count` property:

```
function foo(num) {
  console.log("foo: " + num);

  // keep track of how many times `foo` is called
  data.count++;
}

var data = {
  count: 0
};

var i;
```

```

for (i=0; i<10; i++) {
  if (i > 5) {
    foo( i );
  }
}
// foo: 6
// foo: 7
// foo: 8
// foo: 9

// how many times was `foo` called?
console.log( data.count ); // 4

```

While it is true that this approach "solves" the problem, unfortunately it simply ignores the real problem -- lack of understanding what `this` means and how it works -- and instead falls back to the comfort zone of a more familiar mechanism: lexical scope.

**Note:** Lexical scope is a perfectly fine and useful mechanism; I am not belittling the use of it, by any means (see "*Scope & Closures*" title of this book series). But constantly *guessing* at how to use `this`, and usually being *wrong*, is not a good reason to retreat back to lexical scope and never learn *why* `this` eludes you.

To reference a function object from inside itself, `this` by itself will typically be insufficient. You generally need a reference to the function object via a lexical identifier (variable) that points at it.

Consider these two functions:

```

function foo() {
  foo.count = 4; // `foo` refers to itself
}

setTimeout( function(){
  // anonymous function (no name), cannot
  // refer to itself
}, 10 );

```

In the first function, called a "named function", `foo` is a reference that can be used to refer to the function from inside itself.

But in the second example, the function callback passed to `setTimeout(..)` has no name identifier (so called an "anonymous function"), so there's no proper way to refer to the function object itself.

**Note:** The old-school but now deprecated and frowned-upon `arguments.callee` reference inside a function *also* points to the function object of the currently executing function. This reference is typically the only way to access an anonymous function's object from inside itself. The best approach, however, is to avoid the use of anonymous functions altogether, at least for those which require a self-reference, and instead use a named function (expression). `arguments.callee` is deprecated and should not be used.

So another solution to our running example would have been to use the `foo` identifier as a function object reference in each place, and not use `this` at all, which works:

```

function foo(num) {
  console.log( "foo: " + num );

  // keep track of how many times `foo` is called
  foo.count++;
}

foo.count = 0;

var i;

for (i=0; i<10; i++) {
  if (i > 5) {

```

```

        foo( i );
    }
}

// foo: 6
// foo: 7
// foo: 8
// foo: 9

// how many times was `foo` called?
console.log( foo.count ); // 4

```

However, that approach similarly side-steps *actual* understanding of `this` and relies entirely on the lexical scoping of variable `foo`.

Yet another way of approaching the issue is to force `this` to actually point at the `foo` function object:

```

function foo(num) {
    console.log( "foo: " + num );

    // keep track of how many times `foo` is called
    // Note: `this` IS actually `foo` now, based on
    // how `foo` is called (see below)
    this.count++;
}

foo.count = 0;

var i;

for (i=0; i<10; i++) {
    if (i > 5) {
        // using `call(..)`, we ensure the `this`
        // points at the function object (`foo`) itself
        foo.call( foo, i );
    }
}
// foo: 6
// foo: 7
// foo: 8
// foo: 9

// how many times was `foo` called?
console.log( foo.count ); // 4

```

**Instead of avoiding `this`, we embrace it.** We'll explain in a little bit *how* such techniques work much more completely, so don't worry if you're still a bit confused!

## Its Scope

The next most common misconception about the meaning of `this` is that it somehow refers to the function's scope. It's a tricky question, because in one sense there is some truth, but in the other sense, it's quite misguided.

To be clear, `this` does not, in any way, refer to a function's **lexical scope**. It is true that internally, scope is kind of like an object with properties for each of the available identifiers. But the scope "object" is not accessible to JavaScript code. It's an inner part of the *Engine's implementation*.

Consider code which attempts (and fails!) to cross over the boundary and use `this` to implicitly refer to a function's lexical scope:

```

function foo() {
    var a = 2;
    this.bar();
}

```

```

}

function bar() {
  console.log( this.a );
}

foo(); //undefined

```

There's more than one mistake in this snippet. While it may seem contrived, the code you see is a distillation of actual real-world code that has been exchanged in public community help forums. It's a wonderful (if not sad) illustration of just how misguided `this` assumptions can be.

Firstly, an attempt is made to reference the `bar()` function via `this.bar()`. It is almost certainly an *accident* that it works, but we'll explain the *how* of that shortly. The most natural way to have invoked `bar()` would have been to omit the leading `this.` and just make a lexical reference to the identifier.

However, the developer who writes such code is attempting to use `this` to create a bridge between the lexical scopes of `foo()` and `bar()`, so that `bar()` has access to the variable `a` in the inner scope of `foo()`. **No such bridge is possible.** You cannot use a `this` reference to look something up in a lexical scope. It is not possible.

Every time you feel yourself trying to mix lexical scope look-ups with `this`, remind yourself: *there is no bridge*.

## What's `this` ?

Having set aside various incorrect assumptions, let us now turn our attention to how the `this` mechanism really works.

We said earlier that `this` is not an author-time binding but a runtime binding. It is contextual based on the conditions of the function's invocation. `this` binding has nothing to do with where a function is declared, but has instead everything to do with the manner in which the function is called.

When a function is invoked, an activation record, otherwise known as an execution context, is created. This record contains information about where the function was called from (the call-stack), *how* the function was invoked, what parameters were passed, etc. One of the properties of this record is the `this` reference which will be used for the duration of that function's execution.

In the next chapter, we will learn to find a function's **call-site** to determine how its execution will bind `this`.

## Review (TL;DR)

`this` binding is a constant source of confusion for the JavaScript developer who does not take the time to learn how the mechanism actually works. Guesses, trial-and-error, and blind copy-n-paste from Stack Overflow answers is not an effective or proper way to leverage *this* important `this` mechanism.

To learn `this`, you first have to learn what `this` is *not*, despite any assumptions or misconceptions that may lead you down those paths. `this` is neither a reference to the function itself, nor is it a reference to the function's *lexical* scope.

`this` is actually a binding that is made when a function is invoked, and *what* it references is determined entirely by the call-site where the function is called.

# You Don't Know JS Yet: Objects & Classes - 2nd Edition

## Chapter 2: this All Makes Sense Now!

<b>NOTE:</b>
Work in progress

<b>NOTE:</b>
Everything below here is previous text from 1st edition, and is only here for reference while 2nd edition work is underway. <b>Please ignore this stuff.</b>

In Chapter 1, we discarded various misconceptions about `this` and learned instead that `this` is a binding made for each function invocation, based entirely on its **call-site** (how the function is called).

## Call-site

To understand `this` binding, we have to understand the call-site: the location in code where a function is called (**not where it's declared**). We must inspect the call-site to answer the question: what's `this` `this` a reference to?

Finding the call-site is generally: "go locate where a function is called from", but it's not always that easy, as certain coding patterns can obscure the *true* call-site.

What's important is to think about the **call-stack** (the stack of functions that have been called to get us to the current moment in execution). The call-site we care about is *in* the invocation *before* the currently executing function.

Let's demonstrate call-stack and call-site:

```
function baz() {
  // call-stack is: `baz`
  // so, our call-site is in the global scope

  console.log( "baz" );
  bar(); // <- call-site for `bar`
}

function bar() {
  // call-stack is: `baz` -> `bar`
  // so, our call-site is in `baz`
```

```

console.log( "bar" );
foo(); // <-- call-site for `foo`
}

function foo() {
  // call-stack is: `baz` -> `bar` -> `foo`
  // so, our call-site is in `bar`

  console.log( "foo" );
}

baz(); // <-- call-site for `baz`

```

Take care when analyzing code to find the actual call-site (from the call-stack), because it's the only thing that matters for `this` binding.

**Note:** You can visualize a call-stack in your mind by looking at the chain of function calls in order, as we did with the comments in the above snippet. But this is painstaking and error-prone. Another way of seeing the call-stack is using a debugger tool in your browser. Most modern desktop browsers have built-in developer tools, which includes a JS debugger. In the above snippet, you could have set a breakpoint in the tools for the first line of the `foo()` function, or simply inserted the `debugger;` statement on that first line. When you run the page, the debugger will pause at this location, and will show you a list of the functions that have been called to get to that line, which will be your call stack. So, if you're trying to diagnose `this` binding, use the developer tools to get the call-stack, then find the second item from the top, and that will show you the real call-site.

## Nothing But Rules

We turn our attention now to *how* the call-site determines where `this` will point during the execution of a function.

You must inspect the call-site and determine which of 4 rules applies. We will first explain each of these 4 rules independently, and then we will illustrate their order of precedence, if multiple rules *could* apply to the call-site.

### Default Binding

The first rule we will examine comes from the most common case of function calls: standalone function invocation. Think of `this` rule as the default catch-all rule when none of the other rules apply.

Consider this code:

```

function foo() {
  console.log( this.a );
}

var a = 2;

foo(); // 2

```

The first thing to note, if you were not already aware, is that variables declared in the global scope, as `var a = 2` is, are synonymous with global-object properties of the same name. They're not copies of each other, they *are* each other. Think of it as two sides of the same coin.

Secondly, we see that when `foo()` is called, `this.a` resolves to our global variable `a`. Why? Because in this case, the *default binding* for `this` applies to the function call, and so points `this` at the global object.

How do we know that the *default binding* rule applies here? We examine the call-site to see how `foo()` is called. In our snippet, `foo()` is called with a plain, un-decorated function reference. None of the other rules we will demonstrate will apply here, so the *default binding* applies instead.

If `strict mode` is in effect, the global object is not eligible for the *default binding*, so the `this` is instead set to `undefined`.

```
function foo() {
  "use strict";

  console.log( this.a );
}

var a = 2;

foo(); // TypeError: `this` is `undefined`
```

A subtle but important detail is: even though the overall `this` binding rules are entirely based on the call-site, the global object is **only** eligible for the *default binding* if the **contents** of `foo()` are **not** running in `strict mode`; the `strict mode` state of the call-site of `foo()` is irrelevant.

```
function foo() {
  console.log( this.a );
}

var a = 2;

(function(){
  "use strict";

  foo(); // 2
})();
```

**Note:** Intentionally mixing `strict mode` and non-`strict mode` together in your own code is generally frowned upon. Your entire program should probably either be **Strict** or **non-Strict**. However, sometimes you include a third-party library that has different **Strict**ness than your own code, so care must be taken over these subtle compatibility details.

## Implicit Binding

Another rule to consider is: does the call-site have a context object, also referred to as an owning or containing object, though *these* alternate terms could be slightly misleading.

Consider:

```
function foo() {
  console.log( this.a );
}

var obj = {
  a: 2,
  foo: foo
};

obj.foo(); // 2
```

Firstly, notice the manner in which `foo()` is declared and then later added as a reference property onto `obj`. Regardless of whether `foo()` is initially declared *on* `obj`, or is added as a reference later (as this snippet shows), in neither case is the **function** really "owned" or "contained" by the `obj` object.

However, the call-site *uses* the `obj` context to **reference** the function, so you *could* say that the `obj` object "owns" or "contains" the **function reference** at the time the function is called.

Whatever you choose to call this pattern, at the point that `foo()` is called, it's preceded by an object reference to `obj`. When there is a context object for a function reference, the *implicit binding* rule says that it's *that* object which should be used for the function call's `this` binding.

Because `obj` is the `this` for the `foo()` call, `this.a` is synonymous with `obj.a`.

Only the top/last level of an object property reference chain matters to the call-site. For instance:

```
function foo() {
  console.log( this.a );
}

var obj2 = {
  a: 42,
  foo: foo
};

var obj1 = {
  a: 2,
  obj2: obj2
};

obj1.obj2.foo(); // 42
```

## Implicitly Lost

One of the most common frustrations that `this` binding creates is when an *implicitly bound* function loses that binding, which usually means it falls back to the *default binding*, of either the global object or `undefined`, depending on `strict mode`.

Consider:

```
function foo() {
  console.log( this.a );
}

var obj = {
  a: 2,
  foo: foo
};

var bar = obj.foo; // function reference/alias!

var a = "oops, global"; // `a` also property on global object

bar(); // "oops, global"
```

Even though `bar` appears to be a reference to `obj.foo`, in fact, it's really just another reference to `foo` itself. Moreover, the call-site is what matters, and the call-site is `bar()`, which is a plain, un-decorated call and thus the *default binding* applies.

The more subtle, more common, and more unexpected way this occurs is when we consider passing a callback function:

```
function foo() {
  console.log( this.a );
}

function doFoo(fn) {
  // `fn` is just another reference to `foo`

  fn(); // <-- call-site!
}

var obj = {
  a: 2,
```

```

    foo: foo
};

var a = "oops, global"; // `a` also property on global object

doFoo( obj.foo ); // "oops, global"

```

Parameter passing is just an implicit assignment, and since we're passing a function, it's an implicit reference assignment, so the end result is the same as the previous snippet.

What if the function you're passing your callback to is not your own, but built-in to the language? No difference, same outcome.

```

function foo() {
  console.log( this.a );
}

var obj = {
  a: 2,
  foo: foo
};

var a = "oops, global"; // `a` also property on global object

setTimeout( obj.foo, 100 ); // "oops, global"

```

Think about this crude theoretical pseudo-implementation of `setTimeout()` provided as a built-in from the JavaScript environment:

```

function setTimeout(fn,delay) {
  // wait (somehow) for `delay` milliseconds
  fn(); // <-- call-site!
}

```

It's quite common that our function callbacks *lose* their `this` binding, as we've just seen. But another way that `this` can surprise us is when the function we've passed our callback to intentionally changes the `this` for the call. Event handlers in popular JavaScript libraries are quite fond of forcing your callback to have a `this` which points to, for instance, the DOM element that triggered the event. While that may sometimes be useful, other times it can be downright infuriating. Unfortunately, these tools rarely let you choose.

Either way the `this` is changed unexpectedly, you are not really in control of how your callback function reference will be executed, so you have no way (yet) of controlling the call-site to give your intended binding. We'll see shortly a way of "fixing" that problem by *fixing* the `this`.

## Explicit Binding

With *implicit binding* as we just saw, we had to mutate the object in question to include a reference on itself to the function, and use this property function reference to indirectly (implicitly) bind `this` to the object.

But, what if you want to force a function call to use a particular object for the `this` binding, without putting a property function reference on the object?

"All" functions in the language have some utilities available to them (via their `[[Prototype]]` -- more on that later) which can be useful for this task. Specifically, functions have `call(..)` and `apply(..)` methods. Technically, JavaScript host environments sometimes provide functions which are special enough (a kind way of putting it!) that they do not have such functionality. But those are few. The vast majority of functions provided, and certainly all functions you will create, do have access to `call(..)` and `apply(..)`.

How do these utilities work? They both take, as their first parameter, an object to use for the `this`, and then invoke the function with that `this` specified. Since you are directly stating what you want the `this` to be, we call it *explicit binding*.

Consider:

```
function foo() {
  console.log( this.a );
}

var obj = {
  a: 2
};

foo.call( obj ); // 2
```

Invoking `foo` with *explicit binding* by `foo.call(..)` allows us to force its `this` to be `obj`.

If you pass a simple primitive value (of type `string`, `boolean`, or `number`) as the `this` binding, the primitive value is wrapped in its object-form (`new String(..)`, `new Boolean(..)`, or `new Number(..)`, respectively). This is often referred to as "boxing".

**Note:** With respect to `this` binding, `call(..)` and `apply(..)` are identical. They *do* behave differently with their additional parameters, but that's not something we care about presently.

Unfortunately, *explicit binding* alone still doesn't offer any solution to the issue mentioned previously, of a function "losing" its intended `this` binding, or just having it paved over by a framework, etc.

## Hard Binding

But a variation pattern around *explicit binding* actually does the trick. Consider:

```
function foo() {
  console.log( this.a );
}

var obj = {
  a: 2
};

var bar = function() {
  foo.call( obj );
};

bar(); // 2
setTimeout( bar, 100 ); // 2

// `bar` hard binds `foo`'s `this` to `obj`
// so that it cannot be overridden
bar.call( window ); // 2
```

Let's examine how this variation works. We create a function `bar()` which, internally, manually calls `foo.call(obj)`, thereby forcibly invoking `foo` with `obj` binding for `this`. No matter how you later invoke the function `bar`, it will always manually invoke `foo` with `obj`. This binding is both explicit and strong, so we call it *hard binding*.

The most typical way to wrap a function with a *hard binding* creates a pass-thru of any arguments passed and any return value received:

```
function foo(something) {
  console.log( this.a, something );
  return this.a + something;
}
```

```

var obj = {
  a: 2
};

var bar = function() {
  return foo.apply( obj, arguments );
};

var b = bar( 3 ); // 2 3
console.log( b ); // 5

```

Another way to express this pattern is to create a re-usable helper:

```

function foo(something) {
  console.log( this.a, something );
  return this.a + something;
}

// simple `bind` helper
function bind(fn, obj) {
  return function() {
    return fn.apply( obj, arguments );
  };
}

var obj = {
  a: 2
};

var bar = bind( foo, obj );

var b = bar( 3 ); // 2 3
console.log( b ); // 5

```

Since *hard binding* is such a common pattern, it's provided with a built-in utility as of ES5: `Function.prototype.bind`, and it's used like this:

```

function foo(something) {
  console.log( this.a, something );
  return this.a + something;
}

var obj = {
  a: 2
};

var bar = foo.bind( obj );

var b = bar( 3 ); // 2 3
console.log( b ); // 5

```

`bind(..)` returns a new function that is hard-coded to call the original function with the `this` context set as you specified.

**Note:** As of ES6, the hard-bound function produced by `bind(..)` has a `.name` property that derives from the original *target function*. For example: `bar = foo.bind(..)` should have a `bar.name` value of "bound `foo`", which is the function call name that should show up in a stack trace.

## API Call "Contexts"

Many libraries' functions, and indeed many new built-in functions in the JavaScript language and host environment, provide an optional parameter, usually called "context", which is designed as a work-around for you not having to use `bind(..)` to ensure your callback function uses a particular `this`.

For instance:

```
function foo(el) {
  console.log( el, this.id );
}

var obj = {
  id: "awesome"
};

// use `obj` as `this` for `foo(..)` calls
[1, 2, 3].forEach( foo, obj ); // 1 awesome 2 awesome 3 awesome
```

Internally, these various functions almost certainly use *explicit binding* via `call(..)` or `apply(..)`, saving you the trouble.

## new Binding

The fourth and final rule for `this` binding requires us to re-think a very common misconception about functions and objects in JavaScript.

In traditional class-oriented languages, "constructors" are special methods attached to classes, that when the class is instantiated with a `new` operator, the constructor of that class is called. This usually looks something like:

```
something = new MyClass(..);
```

JavaScript has a `new` operator, and the code pattern to use it looks basically identical to what we see in those class-oriented languages; most developers assume that JavaScript's mechanism is doing something similar. However, there really is *no connection* to class-oriented functionality implied by `new` usage in JS.

First, let's re-define what a "constructor" in JavaScript is. In JS, constructors are **just functions** that happen to be called with the `new` operator in front of them. They are not attached to classes, nor are they instantiating a class. They are not even special types of functions. They're just regular functions that are, in essence, hijacked by the use of `new` in their invocation.

For example, the `Number(..)` function acting as a constructor, quoting from the ES5.1 spec:

### 15.7.2 The Number Constructor

When `Number` is called as part of a `new` expression it is a constructor: it initialises the newly created object.

So, pretty much any ol' function, including the built-in object functions like `Number(..)` (see Chapter 3) can be called with `new` in front of it, and that makes that function call a *constructor call*. This is an important but subtle distinction: there's really no such thing as "constructor functions", but rather construction calls of functions.

When a function is invoked with `new` in front of it, otherwise known as a constructor call, the following things are done automatically:

1. a brand new object is created (aka, constructed) out of thin air
2. *the newly constructed object is `[[Prototype]]`-linked*
3. the newly constructed object is set as the `this` binding for that function call
4. unless the function returns its own alternate `object`, the `new`-invoked function call will *automatically* return the newly constructed object.

Steps 1, 3, and 4 apply to our current discussion. We'll skip over step 2 for now and come back to it in Chapter 5.

Consider this code:

```
function foo(a) {
  this.a = a;
}

var bar = new foo( 2 );
console.log( bar.a ); // 2
```

By calling `foo(...)` with `new` in front of it, we've constructed a new object and set that new object as the `this` for the call of `foo(...)`. So `new` is the final way that a function call's `this` can be bound. We'll call this *new binding*.

## Everything In Order

So, now we've uncovered the 4 rules for binding `this` in function calls. All you need to do is find the call-site and inspect it to see which rule applies. But, what if the call-site has multiple eligible rules? There must be an order of precedence to these rules, and so we will next demonstrate what order to apply the rules.

It should be clear that the *default binding* is the lowest priority rule of the 4. So we'll just set that one aside.

Which is more precedent, *implicit binding* or *explicit binding*? Let's test it:

```
function foo() {
  console.log( this.a );
}

var obj1 = {
  a: 2,
  foo: foo
};

var obj2 = {
  a: 3,
  foo: foo
};

obj1.foo(); // 2
obj2.foo(); // 3

obj1.foo.call( obj2 ); // 3
obj2.foo.call( obj1 ); // 2
```

So, *explicit binding* takes precedence over *implicit binding*, which means you should ask **first** if *explicit binding* applies before checking for *implicit binding*.

Now, we just need to figure out where *new binding* fits in the precedence.

```
function foo(something) {
  this.a = something;
}

var obj1 = {
  foo: foo
};

var obj2 = {};

obj1.foo( 2 );
console.log( obj1.a ); // 2

obj1.foo.call( obj2, 3 );
```

```

console.log( obj2.a ); // 3

var bar = new obj1.foo( 4 );
console.log( obj1.a ); // 2
console.log( bar.a ); // 4

```

OK, *new binding* is more precedent than *implicit binding*. But do you think *new binding* is more or less precedent than *explicit binding*?

**Note:** `new` and `call` / `apply` cannot be used together, so `new foo.call(obj1)` is not allowed, to test *new binding* directly against *explicit binding*. But we can still use a *hard binding* to test the precedence of the two rules.

Before we explore that in a code listing, think back to how *hard binding* physically works, which is that

`Function.prototype.bind(..)` creates a new wrapper function that is hard-coded to ignore its own `this` binding (whatever it may be), and use a manual one we provide.

By that reasoning, it would seem obvious to assume that *hard binding* (which is a form of *explicit binding*) is more precedent than *new binding*, and thus cannot be overridden with `new`.

Let's check:

```

function foo(something) {
  this.a = something;
}

var obj1 = {};

var bar = foo.bind( obj1 );
bar( 2 );
console.log( obj1.a ); // 2

var baz = new bar( 3 );
console.log( obj1.a ); // 2
console.log( baz.a ); // 3

```

Whoa! `bar` is hard-bound against `obj1`, but `new bar(3)` did **not** change `obj1.a` to be `3` as we would have expected. Instead, the *hard bound* (to `obj1`) call to `bar(..)` is able to be overridden with `new`. Since `new` was applied, we got the newly created object back, which we named `baz`, and we see in fact that `baz.a` has the value `3`.

This should be surprising if you go back to our "fake" bind helper:

```

function bind(fn, obj) {
  return function() {
    fn.apply( obj, arguments );
  };
}

```

If you reason about how the helper's code works, it does not have a way for a `new` operator call to override the hard-binding to `obj` as we just observed.

But the built-in `Function.prototype.bind(..)` as of ES5 is more sophisticated, quite a bit so in fact. Here is the (slightly reformatted) polyfill provided by the MDN page for `bind(..)`:

```

if (!Function.prototype.bind) {
  Function.prototype.bind = function(oThis) {
    if (typeof this !== "function") {
      // closest thing possible to the ECMAScript 5
      // internal IsCallable function
      throw new TypeError( "Function.prototype.bind - what " +
        "is trying to be bound is not callable" );
    }
  };
}

```

```

        }

        var aArgs = Array.prototype.slice.call( arguments, 1 ),
            fToBind = this,
            fNOP = function(){},
            fBound = function(){
                return fToBind.apply(
                    (
                        this instanceof fNOP &&
                        oThis ? this : oThis
                    ),
                    aArgs.concat( Array.prototype.slice.call( arguments ) )
                );
            }
        ;

        fNOP.prototype = this.prototype;
        fBound.prototype = new fNOP();

        return fBound;
    );
}

```

**Note:** The `bind(..)` polyfill shown above differs from the built-in `bind(..)` in ES5 with respect to hard-bound functions that will be used with `new` (see below for why that's useful). Because the polyfill cannot create a function without a `.prototype` as the built-in utility does, there's some nuanced indirection to approximate the same behavior. Tread carefully if you plan to use `new` with a hard-bound function and you rely on this polyfill.

The part that's allowing `new` overriding is:

```

this instanceof fNOP &&
oThis ? this : oThis

// ... and:

fNOP.prototype = this.prototype;
fBound.prototype = new fNOP();

```

We won't actually dive into explaining how this trickery works (it's complicated and beyond our scope here), but essentially the utility determines whether or not the hard-bound function has been called with `new` (resulting in a newly constructed object being its `this`), and if so, it uses *that* newly created `this` rather than the previously specified *hard binding* for `this`.

Why is `new` being able to override *hard binding* useful?

The primary reason for this behavior is to create a function (that can be used with `new` for constructing objects) that essentially ignores the `this` *hard binding* but which presets some or all of the function's arguments. One of the capabilities of `bind(..)` is that any arguments passed after the first `this` binding argument are defaulted as standard arguments to the underlying function (technically called "partial application", which is a subset of "currying").

For example:

```

function foo(p1,p2) {
    this.val = p1 + p2;
}

// using `null` here because we don't care about
// the `this` hard-binding in this scenario, and
// it will be overridden by the `new` call anyway!
var bar = foo.bind( null, "p1" );

var baz = new bar( "p2" );

baz.val; // p1p2

```

## Determining `this`

Now, we can summarize the rules for determining `this` from a function call's call-site, in their order of precedence. Ask these questions in this order, and stop when the first rule applies.

1. Is the function called with `new` (**new binding**)? If so, `this` is the newly constructed object.

```
var bar = new foo()
```

2. Is the function called with `call` or `apply` (**explicit binding**), even hidden inside a `bind` **hard binding**? If so, `this` is the explicitly specified object.

```
var bar = foo.call( obj2 )
```

3. Is the function called with a context (**implicit binding**), otherwise known as an owning or containing object? If so, `this` is *that* context object.

```
var bar = obj1.foo()
```

4. Otherwise, default the `this` (**default binding**). If in `strict mode`, pick `undefined`, otherwise pick the `global` object.

```
var bar = foo()
```

That's it. That's *all it takes* to understand the rules of `this` binding for normal function calls. Well... almost.

## Binding Exceptions

As usual, there are some *exceptions* to the "rules".

The `this`-binding behavior can in some scenarios be surprising, where you intended a different binding but you end up with binding behavior from the *default binding* rule (see previous).

### Ignored `this`

If you pass `null` or `undefined` as a `this` binding parameter to `call`, `apply`, or `bind`, those values are effectively ignored, and instead the *default binding* rule applies to the invocation.

```
function foo() {
  console.log( this.a );
}

var a = 2;

foo.call( null ); // 2
```

Why would you intentionally pass something like `null` for a `this` binding?

It's quite common to use `apply(..)` for spreading out arrays of values as parameters to a function call. Similarly, `bind(..)` can curry parameters (pre-set values), which can be very helpful.

```
function foo(a,b) {
  console.log( "a:" + a + ", b:" + b );
}

// spreading out array as parameters
foo.apply( null, [2, 3] ); // a:2, b:3

// currying with `bind(..)`
var bar = foo.bind( null, 2 );
```

```
bar( 3 ); // a:2, b:3
```

Both these utilities require a `this` binding for the first parameter. If the functions in question don't care about `this`, you need a placeholder value, and `null` might seem like a reasonable choice as shown in this snippet.

**Note:** We don't cover it in this book, but ES6 has the `...` spread operator which will let you syntactically "spread out" an array as parameters without needing `apply(..)`, such as `foo(...[1,2])`, which amounts to `foo(1,2)` -- syntactically avoiding a `this` binding if it's unnecessary. Unfortunately, there's no ES6 syntactic substitute for currying, so the `this` parameter of the `bind(..)` call still needs attention.

However, there's a slight hidden "danger" in always using `null` when you don't care about the `this` binding. If you ever use that against a function call (for instance, a third-party library function that you don't control), and that function *does* make a `this` reference, the *default binding* rule means it might inadvertently reference (or worse, mutate!) the `global` object (`window` in the browser).

Obviously, such a pitfall can lead to a variety of *very difficult* to diagnose/track-down bugs.

## Safer `this`

Perhaps a somewhat "safer" practice is to pass a specifically set up object for `this` which is guaranteed not to be an object that can create problematic side effects in your program. Borrowing terminology from networking (and the military), we can create a "DMZ" (de-militarized zone) object -- nothing more special than a completely empty, non-delegated (see Chapters 5 and 6) object.

If we always pass a DMZ object for ignored `this` bindings we don't think we need to care about, we're sure any hidden/unexpected usage of `this` will be restricted to the empty object, which insulates our program's `global` object from side-effects.

Since this object is totally empty, I personally like to give it the variable name `ø` (the lowercase mathematical symbol for the empty set). On many keyboards (like US-layout on Mac), this symbol is easily typed with `⌥ + o` (option+o). Some systems also let you set up hotkeys for specific symbols. If you don't like the `ø` symbol, or your keyboard doesn't make that as easy to type, you can of course call it whatever you want.

Whatever you call it, the easiest way to set it up as **totally empty** is `Object.create(null)` (see Chapter 5).

`Object.create(null)` is similar to `{}`, but without the delegation to `Object.prototype`, so it's "more empty" than just `{}`.

```
function foo(a,b) {
  console.log( "a:" + a + ", b:" + b );
}

// our DMZ empty object
var ø = Object.create( null );

// spreading out array as parameters
foo.apply( ø, [2, 3] ); // a:2, b:3

// currying with `bind(..)`
var bar = foo.bind( ø, 2 );
bar( 3 ); // a:2, b:3
```

Not only functionally "safer", there's a sort of stylistic benefit to `ø`, in that it semantically conveys "I want the `this` to be empty" a little more clearly than `null` might. But again, name your DMZ object whatever you prefer.

## Indirection

Another thing to be aware of is you can (intentionally or not!) create "indirect references" to functions, and in those cases, when that function reference is invoked, the *default binding* rule also applies.

One of the most common ways that *indirect references* occur is from an assignment:

```
function foo() {
  console.log( this.a );
}

var a = 2;
var o = { a: 3, foo: foo };
var p = { a: 4 };

o.foo(); // 3
(p.foo = o.foo)(); // 2
```

The *result value* of the assignment expression `p.foo = o.foo` is a reference to just the underlying function object. As such, the effective call-site is just `foo()`, not `p.foo()` or `o.foo()` as you might expect. Per the rules above, the *default binding* rule applies.

Reminder: regardless of how you get to a function invocation using the *default binding* rule, the `strict mode` status of the **contents** of the invoked function making the `this` reference -- not the function call-site -- determines the *default binding* value: either the `global` object if in non-`strict mode` or `undefined` if in `strict mode`.

## Softening Binding

We saw earlier that *hard binding* was one strategy for preventing a function call falling back to the *default binding* rule inadvertently, by forcing it to be bound to a specific `this` (unless you use `new` to override it!). The problem is, *hard-binding* greatly reduces the flexibility of a function, preventing manual `this` override with either the *implicit binding* or even subsequent *explicit binding* attempts.

It would be nice if there was a way to provide a different default for *default binding* (not `global` or `undefined`), while still leaving the function able to be manually `this` bound via *implicit binding* or *explicit binding* techniques.

We can construct a so-called *soft binding* utility which emulates our desired behavior.

```
if (!Function.prototype.softBind) {
  Function.prototype.softBind = function(obj) {
    var fn = this,
      curried = [].slice.call( arguments, 1 ),
      bound = function bound() {
        return fn.apply(
          (!this ||
            (typeof window !== "undefined" &&
              this === window) ||
            (typeof global !== "undefined" &&
              this === global)
          ) ? obj : this,
          curried.concat.apply( curried, arguments )
        );
      };
    bound.prototype = Object.create( fn.prototype );
    return bound;
  };
}
```

The `softBind(..)` utility provided here works similarly to the built-in ES5 `bind(..)` utility, except with our *soft binding* behavior. It wraps the specified function in logic that checks the `this` at call-time and if it's `global` or `undefined`, uses a pre-specified alternate `default ( obj )`. Otherwise the `this` is left untouched. It also provides optional currying (see the `bind(..)` discussion earlier).

Let's demonstrate its usage:

```
function foo() {
  console.log("name: " + this.name);
}

var obj = { name: "obj" },
  obj2 = { name: "obj2" },
  obj3 = { name: "obj3" };

var fooOBJ = foo.softBind( obj );

fooOBJ(); // name: obj

obj2.foo = foo.softBind(obj);
obj2.foo(); // name: obj2  <---- look!!!

fooOBJ.call( obj3 ); // name: obj3  <---- look!

setTimeout( obj2.foo, 10 ); // name: obj  <---- falls back to soft-binding
```

The soft-bound version of the `foo()` function can be manually `this`-bound to `obj2` or `obj3` as shown, but it falls back to `obj` if the *default binding* would otherwise apply.

## Lexical `this`

Normal functions abide by the 4 rules we just covered. But ES6 introduces a special kind of function that does not use these rules: arrow-function.

Arrow-functions are signified not by the `function` keyword, but by the `=>` so called "fat arrow" operator. Instead of using the four standard `this` rules, arrow-functions adopt the `this` binding from the enclosing (function or global) scope.

Let's illustrate arrow-function lexical scope:

```
function foo() {
  // return an arrow function
  return (a) => {
    // `this` here is lexically adopted from `foo()`
    console.log( this.a );
  };
}

var obj1 = {
  a: 2
};

var obj2 = {
  a: 3
};

var bar = foo.call( obj1 );
bar.call( obj2 ); // 2, not 3!
```

The arrow-function created in `foo()` lexically captures whatever `foo()`'s `this` is at its call-time. Since `foo()` was `this`-bound to `obj1`, `bar` (a reference to the returned arrow-function) will also be `this`-bound to `obj1`. The lexical binding of an arrow-function cannot be overridden (even with `new`!).

The most common use-case will likely be in the use of callbacks, such as event handlers or timers:

```
function foo() {
  setTimeout(() => {
```

```

        // `this` here is lexically adopted from `foo()`
        console.log( this.a );
    }, 100);
}

var obj = {
    a: 2
};

foo.call( obj ); // 2

```

While arrow-functions provide an alternative to using `bind(..)` on a function to ensure its `this`, which can seem attractive, it's important to note that they essentially are disabling the traditional `this` mechanism in favor of more widely-understood lexical scoping. Pre-ES6, we already have a fairly common pattern for doing so, which is basically almost indistinguishable from the spirit of ES6 arrow-functions:

```

function foo() {
    var self = this; // lexical capture of `this`
    setTimeout( function(){
        console.log( self.a );
    }, 100 );
}

var obj = {
    a: 2
};

foo.call( obj ); // 2

```

While `self = this` and arrow-functions both seem like good "solutions" to not wanting to use `bind(..)`, they are essentially fleeing from `this` instead of understanding and embracing it.

If you find yourself writing `this`-style code, but most or all the time, you defeat the `this` mechanism with lexical `self = this` or arrow-function "tricks", perhaps you should either:

1. Use only lexical scope and forget the false pretense of `this`-style code.
2. Embrace `this`-style mechanisms completely, including using `bind(..)` where necessary, and try to avoid `self = this` and arrow-function "lexical this" tricks.

A program can effectively use both styles of code (lexical and `this`), but inside of the same function, and indeed for the same sorts of look-ups, mixing the two mechanisms is usually asking for harder-to-maintain code, and probably working too hard to be clever.

## Review (TL;DR)

Determining the `this` binding for an executing function requires finding the direct call-site of that function. Once examined, four rules can be applied to the call-site, in *this* order of precedence:

1. Called with `new`? Use the newly constructed object.
2. Called with `call` or `apply` (or `bind`)? Use the specified object.
3. Called with a context object owning the call? Use that context object.
4. Default: `undefined` in `strict mode`, global object otherwise.

Be careful of accidental/unintentional invoking of the *default binding* rule. In cases where you want to "safely" ignore a `this` binding, a "DMZ" object like `o = Object.create(null)` is a good placeholder value that protects the `global` object from unintended side-effects.

Instead of the four standard binding rules, ES6 arrow-functions use lexical scoping for `this` binding, which means they adopt the `this` binding (whatever it is) from its enclosing function call. They are essentially a syntactic replacement of `self = this` in pre-ES6 coding.

# You Don't Know JS Yet: Objects & Classes - 2nd Edition

## Chapter 3: Objects

**NOTE:**

Work in progress

**NOTE:**

Everything below here is previous text from 1st edition, and is only here for reference while 2nd edition work is underway.  
**Please ignore this stuff.**

In Chapters 1 and 2, we explained how the `this` binding points to various objects depending on the call-site of the function invocation. But what exactly are objects, and why do we need to point to them? We will explore objects in detail in this chapter.

## Syntax

Objects come in two forms: the declarative (literal) form, and the constructed form.

The literal syntax for an object looks like this:

```
var myObj = {
  key: value
  // ...
};
```

The constructed form looks like this:

```
var myObj = new Object();
myObj.key = value;
```

The constructed form and the literal form result in exactly the same sort of object. The only difference really is that you can add one or more key/value pairs to the literal declaration, whereas with constructed-form objects, you must add the properties one-by-one.

**Note:** It's extremely uncommon to use the "constructed form" for creating objects as just shown. You would pretty much always want to use the literal syntax form. The same will be true of most of the built-in objects (see below).

# Type

Objects are the general building block upon which much of JS is built. They are one of the 6 primary types (called "language types" in the specification) in JS:

- `string`
- `number`
- `boolean`
- `null`
- `undefined`
- `object`

Note that the *simple primitives* (`string`, `number`, `boolean`, `null`, and `undefined`) are **not** themselves `objects`. `null` is sometimes referred to as an object type, but this misconception stems from a bug in the language which causes `typeof null` to return the string `"object"` incorrectly (and confusingly). In fact, `null` is its own primitive type.

**It's a common mis-statement that "everything in JavaScript is an object". This is clearly not true.**

By contrast, there *are* a few special object sub-types, which we can refer to as *complex primitives*.

`function` is a sub-type of object (technically, a "callable object"). Functions in JS are said to be "first class" in that they are basically just normal objects (with callable behavior semantics bolted on), and so they can be handled like any other plain object.

Arrays are also a form of objects, with extra behavior. The organization of contents in arrays is slightly more structured than for general objects.

## Built-in Objects

There are several other object sub-types, usually referred to as built-in objects. For some of them, their names seem to imply they are directly related to their simple primitives counter-parts, but in fact, their relationship is more complicated, which we'll explore shortly.

- `String`
- `Number`
- `Boolean`
- `Object`
- `Function`
- `Array`
- `Date`
- `RegExp`
- `Error`

These built-ins have the appearance of being actual types, even classes, if you rely on the similarity to other languages such as Java's `String` class.

But in JS, these are actually just built-in functions. Each of these built-in functions can be used as a constructor (that is, a function call with the `new` operator -- see Chapter 2), with the result being a newly *constructed* object of the sub-type in question. For instance:

```
var strPrimitive = "I am a string";
typeof strPrimitive;                                // "string"
strPrimitive instanceof String;                     // false

var strObject = new String( "I am a string" );
typeof strObject;                                  // "object"
strObject instanceof String;                      // true
```

```
// inspect the object sub-type
Object.prototype.toString.call( strObject ); // [object String]
```

We'll see in detail in a later chapter exactly how the `Object.prototype.toString...` bit works, but briefly, we can inspect the internal sub-type by borrowing the base default `toString()` method, and you can see it reveals that `strObject` is an object that was in fact created by the `String` constructor.

The primitive value `"I am a string"` is not an object, it's a primitive literal and immutable value. To perform operations on it, such as checking its length, accessing its individual character contents, etc, a `String` object is required.

Luckily, the language automatically coerces a `"string"` primitive to a `String` object when necessary, which means you almost never need to explicitly create the Object form. It is **strongly preferred** by the majority of the JS community to use the literal form for a value, where possible, rather than the constructed object form.

Consider:

```
var strPrimitive = "I am a string";

console.log( strPrimitive.length ); // 13

console.log( strPrimitive.charAt( 3 ) ); // "m"
```

In both cases, we call a property or method on a string primitive, and the engine automatically coerces it to a `String` object, so that the property/method access works.

The same sort of coercion happens between the number literal primitive `42` and the `new Number(42)` object wrapper, when using methods like `42.359.toFixed(2)`. Likewise for `Boolean` objects from `"boolean"` primitives.

`null` and `undefined` have no object wrapper form, only their primitive values. By contrast, `Date` values can *only* be created with their constructed object form, as they have no literal form counter-part.

`Object`, `Array`, `Function`, and `RegExp` (regular expressions) are all objects regardless of whether the literal or constructed form is used. The constructed form does offer, in some cases, more options in creation than the literal form counterpart. Since objects are created either way, the simpler literal form is almost universally preferred. **Only use the constructed form if you need the extra options.**

`Error` objects are rarely created explicitly in code, but usually created automatically when exceptions are thrown. They can be created with the constructed form `new Error(..)`, but it's often unnecessary.

## Contents

As mentioned earlier, the contents of an object consist of values (any type) stored at specifically named *locations*, which we call properties.

It's important to note that while we say "contents" which implies that these values are *actually* stored inside the object, that's merely an appearance. The engine stores values in implementation-dependent ways, and may very well not store them *in* some object container. What *is* stored in the container are these property names, which act as pointers (technically, *references*) to where the values are stored.

Consider:

```
var myObject = {
  a: 2
};

myObject.a; // 2
```

```
myObject["a"]; // 2
```

To access the value at the *location* `a` in `myObject`, we need to use either the `.` operator or the `[ ]` operator. The `.a` syntax is usually referred to as "property" access, whereas the `["a"]` syntax is usually referred to as "key" access. In reality, they both access the same *location*, and will pull out the same value, `2`, so the terms can be used interchangeably. We will use the most common term, "property access" from here on.

The main difference between the two syntaxes is that the `.` operator requires an `Identifier` compatible property name after it, whereas the `[..]` syntax can take basically any UTF-8/unicode compatible string as the name for the property. To reference a property of the name "Super-Fun!", for instance, you would have to use the `["Super-Fun!"]` access syntax, as `Super-Fun!` is not a valid `Identifier` property name.

Also, since the `[..]` syntax uses a string's **value** to specify the location, this means the program can programmatically build up the value of the string, such as:

```
var wantA = true;
var myObject = {
  a: 2
};

var idx;

if (wantA) {
  idx = "a";
}

// later

console.log( myObject[idx] ); // 2
```

In objects, property names are **always** strings. If you use any other value besides a `string` (primitive) as the property, it will first be converted to a string. This even includes numbers, which are commonly used as array indexes, so be careful not to confuse the use of numbers between objects and arrays.

```
var myObject = { };

myObject[true] = "foo";
myObject[3] = "bar";
myObject[myObject] = "baz";

myObject["true"];           // "foo"
myObject["3"];             // "bar"
myObject["[object Object]"]; // "baz"
```

## Computed Property Names

The `myObject[...]` property access syntax we just described is useful if you need to use a computed expression value *as* the key name, like `myObject[prefix + name]`. But that's not really helpful when declaring objects using the object-literal syntax.

ES6 adds *computed property names*, where you can specify an expression, surrounded by a `[ ]` pair, in the key-name position of an object-literal declaration:

```
var prefix = "foo";

var myObject = {
  [prefix + "bar"]: "hello",
  [prefix + "baz"]: "world"
};
```

```
myObject["foobar"]; // hello
myObject["foobaz"]; // world
```

The most common usage of *computed property names* will probably be for ES6 `Symbol`s, which we will not be covering in detail in this book. In short, they're a new primitive data type which has an opaque unguessable value (technically a `string` value). You will be strongly discouraged from working with the *actual value* of a `Symbol` (which can theoretically be different between different JS engines), so the name of the `Symbol`, like `Symbol.Something` (just a made up name!), will be what you use:

```
var myObject = {
  [Symbol.Something]: "hello world"
};
```

## Property vs. Method

Some developers like to make a distinction when talking about a property access on an object, if the value being accessed happens to be a function. Because it's tempting to think of the function as *belonging* to the object, and in other languages, functions which belong to objects (aka, "classes") are referred to as "methods", it's not uncommon to hear, "method access" as opposed to "property access".

**The specification makes this same distinction**, interestingly.

Technically, functions never "belong" to objects, so saying that a function that just happens to be accessed on an object reference is automatically a "method" seems a bit of a stretch of semantics.

It is true that some functions have `this` references in them, and that *sometimes* these `this` references refer to the object reference at the call-site. But this usage really does not make that function any more a "method" than any other function, as `this` is dynamically bound at run-time, at the call-site, and thus its relationship to the object is indirect, at best.

Every time you access a property on an object, that is a **property access**, regardless of the type of value you get back. If you *happen* to get a function from that property access, it's not magically a "method" at that point. There's nothing special (outside of possible implicit `this` binding as explained earlier) about a function that comes from a property access.

For instance:

```
function foo() {
  console.log( "foo" );
}

var someFoo = foo;    // variable reference to `foo`

var myObject = {
  someFoo: foo
};

foo;                  // function foo(){...}

someFoo;              // function foo(){...}

myObject.someFoo;    // function foo(){...}
```

`someFoo` and `myObject.someFoo` are just two separate references to the same function, and neither implies anything about the function being special or "owned" by any other object. If `foo()` above was defined to have a `this` reference inside it, that `myObject.someFoo` *implicit binding* would be the **only** observable difference between the two references. Neither reference really makes sense to be called a "method".

**Perhaps one could argue** that a function *becomes a method*, not at definition time, but during run-time just for that invocation, depending on how it's called at its call-site (with an object reference context or not -- see Chapter 2 for more details). Even this interpretation is a bit of a stretch.

The safest conclusion is probably that "function" and "method" are interchangeable in JavaScript.

**Note:** ES6 adds a `super` reference, which is typically going to be used with `class` (see Appendix A). The way `super` behaves (static binding rather than late binding as `this`) gives further weight to the idea that a function which is `super` bound somewhere is more a "method" than "function". But again, these are just subtle semantic (and mechanical) nuances.

Even when you declare a function expression as part of the object-literal, that function doesn't magically *belong* more to the object -- still just multiple references to the same function object:

```
var myObject = {
  foo: function foo() {
    console.log( "foo" );
  }
};

var someFoo = myObject.foo;

someFoo;      // function foo(){...}

myObject.foo; // function foo(){...}
```

**Note:** In Chapter 6, we will cover an ES6 short-hand for that `foo: function foo(){ ... }` declaration syntax in our object-literal.

## Arrays

Arrays also use the `[ ]` access form, but as mentioned above, they have slightly more structured organization for how and where values are stored (though still no restriction on what *type* of values are stored). Arrays assume *numeric indexing*, which means that values are stored in locations, usually called *indices*, at non-negative integers, such as `0` and `42`.

```
var myArray = [ "foo", 42, "bar" ];

myArray.length;      // 3

myArray[0];          // "foo"

myArray[2];          // "bar"
```

Arrays *are* objects, so even though each index is a positive integer, you can *also* add properties onto the array:

```
var myArray = [ "foo", 42, "bar" ];

myArray.baz = "baz";

myArray.length;      // 3

myArray.baz;          // "baz"
```

Notice that adding named properties (regardless of `.` or `[ ]` operator syntax) does not change the reported `length` of the array.

You *could* use an array as a plain key/value object, and never add any numeric indices, but this is a bad idea because arrays have behavior and optimizations specific to their intended use, and likewise with plain objects. Use objects to store key/value pairs, and arrays to store values at numeric indices.

**Be careful:** If you try to add a property to an array, but the property name *looks* like a number, it will end up instead as a numeric index (thus modifying the array contents):

```
var myArray = [ "foo", 42, "bar" ];
```

```
myArray["3"] = "baz";
myArray.length; // 4
myArray[3]; // "baz"
```

## Duplicating Objects

One of the most commonly requested features when developers newly take up the JavaScript language is how to duplicate an object. It would seem like there should just be a built-in `copy()` method, right? It turns out that it's a little more complicated than that, because it's not fully clear what, by default, should be the algorithm for the duplication.

For example, consider this object:

```
function anotherFunction() { /*...*/ }

var anotherObject = {
  c: true
};

var anotherArray = [];

var myObject = {
  a: 2,
  b: anotherObject, // reference, not a copy!
  c: anotherArray, // another reference!
  d: anotherFunction
};

anotherArray.push( anotherObject, myObject );
```

What exactly should be the representation of a *copy* of `myObject`?

Firstly, we should answer if it should be a *shallow* or *deep* copy. A *shallow copy* would end up with `a` on the new object as a copy of the value `2`, but `b`, `c`, and `d` properties as just references to the same places as the references in the original object. A *deep copy* would duplicate not only `myObject`, but `anotherObject` and `anotherArray`. But then we have issues that `anotherArray` has references to `anotherObject` and `myObject` in it, so *those* should also be duplicated rather than reference-preserved. Now we have an infinite circular duplication problem because of the circular reference.

Should we detect a circular reference and just break the circular traversal (leaving the deep element not fully duplicated)? Should we error out completely? Something in between?

Moreover, it's not really clear what "duplicating" a function would mean? There are some hacks like pulling out the `toString()` serialization of a function's source code (which varies across implementations and is not even reliable in all engines depending on the type of function being inspected).

So how do we resolve all these tricky questions? Various JS frameworks have each picked their own interpretations and made their own decisions. But which of these (if any) should JS adopt as *the standard*? For a long time, there was no clear answer.

One subset solution is that objects which are JSON-safe (that is, can be serialized to a JSON string and then re-parsed to an object with the same structure and values) can easily be *duplicated* with:

```
var newObj = JSON.parse( JSON.stringify( someObj ) );
```

Of course, that requires you to ensure your object is JSON safe. For some situations, that's trivial. For others, it's insufficient.

At the same time, a shallow copy is fairly understandable and has far less issues, so ES6 has now defined `Object.assign(..)` for this task. `Object.assign(..)` takes a *target* object as its first parameter, and one or more *source* objects as its subsequent parameters. It iterates over all the *enumerable* (see below), *owned keys (immediately present)* on the *source* object(s) and copies them (via `=` assignment only) to *target*. It also, helpfully, returns *target*, as you can see below:

```
var newObj = Object.assign( {}, myObject );

newObj.a; // 2
newObj.b === anotherObject; // true
newObj.c === anotherArray; // true
newObj.d === anotherFunction; // true
```

**Note:** In the next section, we describe "property descriptors" (property characteristics) and show the use of `Object.defineProperty(..)`. The duplication that occurs for `Object.assign(..)` however is purely `=` style assignment, so any special characteristics of a property (like `writable`) on a source object **are not preserved** on the target object.

## Property Descriptors

Prior to ES5, the JavaScript language gave no direct way for your code to inspect or draw any distinction between the characteristics of properties, such as whether the property was read-only or not.

But as of ES5, all properties are described in terms of a **property descriptor**.

Consider this code:

```
var myObject = {
  a: 2
};

Object.getOwnPropertyDescriptor( myObject, "a" );
// {
//   value: 2,
//   writable: true,
//   enumerable: true,
//   configurable: true
// }
```

As you can see, the property descriptor (called a "data descriptor" since it's only for holding a data value) for our normal object property `a` is much more than just its `value` of `2`. It includes 3 other characteristics: `writable`, `enumerable`, and `configurable`.

While we can see what the default values for the property descriptor characteristics are when we create a normal property, we can use `Object.defineProperty(..)` to add a new property, or modify an existing one (if it's `configurable`!), with the desired characteristics.

For example:

```
var myObject = {};

Object.defineProperty( myObject, "a", {
  value: 2,
  writable: true,
  configurable: true,
  enumerable: true
} );

myObject.a; // 2
```

Using `defineProperty(..)`, we added the plain, normal `a` property to `myObject` in a manually explicit way. However, you generally wouldn't use this manual approach unless you wanted to modify one of the descriptor characteristics from its normal behavior.

## Writable

The ability for you to change the value of a property is controlled by `writable`.

Consider:

```
var myObject = {};

Object.defineProperty( myObject, "a", {
    value: 2,
    writable: false, // not writable!
    configurable: true,
    enumerable: true
} );

myObject.a = 3;

myObject.a; // 2
```

As you can see, our modification of the `value` silently failed. If we try in `strict mode`, we get an error:

```
"use strict";

var myObject = {};

Object.defineProperty( myObject, "a", {
    value: 2,
    writable: false, // not writable!
    configurable: true,
    enumerable: true
} );

myObject.a = 3; // TypeError
```

The `TypeError` tells us we cannot change a non-writable property.

**Note:** We will discuss getters/setters shortly, but briefly, you can observe that `writable:false` means a value cannot be changed, which is somewhat equivalent to if you defined a no-op setter. Actually, your no-op setter would need to throw a `TypeError` when called, to be truly conformant to `writable:false`.

## Configurable

As long as a property is currently configurable, we can modify its descriptor definition, using the same `defineProperty(..)` utility.

```
var myObject = {
    a: 2
};

myObject.a = 3;
myObject.a; // 3

Object.defineProperty( myObject, "a", {
    value: 4,
    writable: true,
    configurable: false, // not configurable!
    enumerable: true
});
```

```

} );

myObject.a;                      // 4
myObject.a = 5;
myObject.a;                      // 5

Object.defineProperty( myObject, "a", {
  value: 6,
  writable: true,
  configurable: true,
  enumerable: true
} ); // TypeError

```

The final `defineProperty(..)` call results in a `TypeError`, regardless of `strict mode`, if you attempt to change the descriptor definition of a non-configurable property. Be careful: as you can see, changing `configurable` to `false` is a **one-way action, and cannot be undone!**

**Note:** There's a nuanced exception to be aware of: even if the property is already `configurable:false`, `writable` can always be changed from `true` to `false` without error, but not back to `true` if already `false`.

Another thing `configurable:false` prevents is the ability to use the `delete` operator to remove an existing property.

```

var myObject = {
  a: 2
};

myObject.a;                      // 2
delete myObject.a;
myObject.a;                      // undefined

Object.defineProperty( myObject, "a", {
  value: 2,
  writable: true,
  configurable: false,
  enumerable: true
} );

myObject.a;                      // 2
delete myObject.a;
myObject.a;                      // 2

```

As you can see, the last `delete` call failed (silently) because we made the `a` property non-configurable.

`delete` is only used to remove object properties (which can be removed) directly from the object in question. If an object property is the last remaining *reference* to some object/function, and you `delete` it, that removes the reference and now that unreferenced object/function can be garbage collected. But, it is **not** proper to think of `delete` as a tool to free up allocated memory as it does in other languages (like C/C++). `delete` is just an object property removal operation -- nothing more.

## Enumerable

The final descriptor characteristic we will mention here (there are two others, which we deal with shortly when we discuss getter/setters) is `enumerable`.

The name probably makes it obvious, but this characteristic controls if a property will show up in certain object-property enumerations, such as the `for..in` loop. Set to `false` to keep it from showing up in such enumerations, even though it's still completely accessible. Set to `true` to keep it present.

All normal user-defined properties are defaulted to `enumerable`, as this is most commonly what you want. But if you have a special property you want to hide from enumeration, set it to `enumerable:false`.

We'll demonstrate enumerability in much more detail shortly, so keep a mental bookmark on this topic.

## Immutability

It is sometimes desired to make properties or objects that cannot be changed (either by accident or intentionally). ES5 adds support for handling that in a variety of different nuanced ways.

It's important to note that **all** of these approaches create shallow immutability. That is, they affect only the object and its direct property characteristics. If an object has a reference to another object (array, object, function, etc), the *contents* of that object are not affected, and remain mutable.

```
myImmutableObject.foo; // [1,2,3]
myImmutableObject.foo.push( 4 );
myImmutableObject.foo; // [1,2,3,4]
```

We assume in this snippet that `myImmutableObject` is already created and protected as immutable. But, to also protect the contents of `myImmutableObject.foo` (which is its own object -- array), you would also need to make `foo` immutable, using one or more of the following functionalities.

**Note:** It is not terribly common to create deeply entrenched immutable objects in JS programs. Special cases can certainly call for it, but as a general design pattern, if you find yourself wanting to *seal* or *freeze* all your objects, you may want to take a step back and reconsider your program design to be more robust to potential changes in objects' values.

## Object Constant

By combining `writable:false` and `configurable:false`, you can essentially create a *constant* (cannot be changed, redefined or deleted) as an object property, like:

```
var myObject = {};

Object.defineProperty( myObject, "FAVORITE_NUMBER", {
  value: 42,
  writable: false,
  configurable: false
});
```

## Prevent Extensions

If you want to prevent an object from having new properties added to it, but otherwise leave the rest of the object's properties alone, call `Object.preventExtensions(..)`:

```
var myObject = {
  a: 2
};

Object.preventExtensions( myObject );

myObject.b = 3;
myObject.b; // undefined
```

In `non-strict mode`, the creation of `b` fails silently. In `strict mode`, it throws a `TypeError`.

## Seal

`Object.seal(..)` creates a "sealed" object, which means it takes an existing object and essentially calls `Object.preventExtensions(..)` on it, but also marks all its existing properties as `configurable:false`.

So, not only can you not add any more properties, but you also cannot reconfigure or delete any existing properties (though you *can* still modify their values).

## Freeze

`Object.freeze(..)` creates a frozen object, which means it takes an existing object and essentially calls `Object.seal(..)` on it, but it also marks all "data accessor" properties as `writable:false`, so that their values cannot be changed.

This approach is the highest level of immutability that you can attain for an object itself, as it prevents any changes to the object or to any of its direct properties (though, as mentioned above, the contents of any referenced other objects are unaffected).

You could "deep freeze" an object by calling `Object.freeze(..)` on the object, and then recursively iterating over all objects it references (which would have been unaffected thus far), and calling `Object.freeze(..)` on them as well. Be careful, though, as that could affect other (shared) objects you're not intending to affect.

## [[Get]]

There's a subtle, but important, detail about how property accesses are performed.

Consider:

```
var myObject = {
  a: 2
};

myObject.a; // 2
```

The `myObject.a` is a property access, but it doesn't *just* look in `myObject` for a property of the name `a`, as it might seem.

According to the spec, the code above actually performs a `[[Get]]` operation (kinda like a function call: `[[Get]]()`) on the `myObject`. The default built-in `[[Get]]` operation for an object *first* inspects the object for a property of the requested name, and if it finds it, it will return the value accordingly.

However, the `[[Get]]` algorithm defines other important behavior if it does *not* find a property of the requested name. We will examine in Chapter 5 what happens *next* (traversal of the `[[Prototype]]` chain, if any).

But one important result of this `[[Get]]` operation is that if it cannot through any means come up with a value for the requested property, it instead returns the value `undefined`.

```
var myObject = {
  a: 2
};

myObject.b; // undefined
```

This behavior is different from when you reference *variables* by their identifier names. If you reference a variable that cannot be resolved within the applicable lexical scope look-up, the result is not `undefined` as it is for object properties, but instead a `ReferenceError` is thrown.

```
var myObject = {
  a: undefined
};

myObject.a; // undefined

myObject.b; // undefined
```

From a *value* perspective, there is no difference between these two references -- they both result in `undefined`. However, the `[[Get]]` operation underneath, though subtle at a glance, potentially performed a bit more "work" for the reference `myObject.b` than for the reference `myObject.a`.

Inspecting only the value results, you cannot distinguish whether a property exists and holds the explicit value `undefined`, or whether the property does *not* exist and `undefined` was the default return value after `[[Get]]` failed to return something explicitly. However, we will see shortly how you *can* distinguish these two scenarios.

## [[Put]]

Since there's an internally defined `[[Get]]` operation for getting a value from a property, it should be obvious there's also a default `[[Put]]` operation.

It may be tempting to think that an assignment to a property on an object would just invoke `[[Put]]` to set or create that property on the object in question. But the situation is more nuanced than that.

When invoking `[[Put]]`, how it behaves differs based on a number of factors, including (most impactfully) whether the property is already present on the object or not.

If the property is present, the `[[Put]]` algorithm will roughly check:

1. Is the property an accessor descriptor (see "Getters & Setters" section below)? **If so, call the setter, if any.**
2. Is the property a data descriptor with `writable` of `false`? **If so, silently fail in non-strict mode, or throw `TypeError` in strict mode.**
3. Otherwise, set the value to the existing property as normal.

If the property is not yet present on the object in question, the `[[Put]]` operation is even more nuanced and complex. We will revisit this scenario in Chapter 5 when we discuss `[[Prototype]]` to give it more clarity.

## Getters & Setters

The default `[[Put]]` and `[[Get]]` operations for objects completely control how values are set to existing or new properties, or retrieved from existing properties, respectively.

**Note:** Using future/advanced capabilities of the language, it may be possible to override the default `[[Get]]` or `[[Put]]` operations for an entire object (not just per property). This is beyond the scope of our discussion in this book, but will be covered later in the "You Don't Know JS" series.

ES5 introduced a way to override part of these default operations, not on an object level but a per-property level, through the use of getters and setters. Getters are properties which actually call a hidden function to retrieve a value. Setters are properties which actually call a hidden function to set a value.

When you define a property to have either a getter or a setter or both, its definition becomes an "accessor descriptor" (as opposed to a "data descriptor"). For accessor-descriptors, the `value` and `writable` characteristics of the descriptor are moot and ignored, and instead JS considers the `set` and `get` characteristics of the property (as well as `configurable` and `enumerable`).

Consider:

```
var myObject = {
  // define a getter for `a`
  get a() {
    return 2;
  }
};

Object.defineProperty(
  myObject, // target
```

```

    "b",           // property name
    {             // descriptor
        // define a getter for `b`
        get: function(){ return this.a * 2 },
        // make sure `b` shows up as an object property
        enumerable: true
    }
);

myObject.a; // 2

myObject.b; // 4

```

Either through object-literal syntax with `get a() { .. }` or through explicit definition with `defineProperty(..)`, in both cases we created a property on the object that actually doesn't hold a value, but whose access automatically results in a hidden function call to the getter function, with whatever value it returns being the result of the property access.

```

var myObject = {
    // define a getter for `a`
    get a() {
        return 2;
    }
};

myObject.a = 3;

myObject.a; // 2

```

Since we only defined a getter for `a`, if we try to set the value of `a` later, the set operation won't throw an error but will just silently throw the assignment away. Even if there was a valid setter, our custom getter is hard-coded to return only `2`, so the set operation would be moot.

To make this scenario more sensible, properties should also be defined with setters, which override the default `[[Put]]` operation (aka, assignment), per-property, just as you'd expect. You will almost certainly want to always declare both getter and setter (having only one or the other often leads to unexpected/surprising behavior):

```

var myObject = {
    // define a getter for `a`
    get a() {
        return this._a_;
    },
    // define a setter for `a`
    set a(val) {
        this._a_ = val * 2;
    }
};

myObject.a = 2;

myObject.a; // 4

```

**Note:** In this example, we actually store the specified value `2` of the assignment (`[[Put]]` operation) into another variable `_a_`. The `_a_` name is purely by convention for this example and implies nothing special about its behavior -- it's a normal property like any other.

## Existence

We showed earlier that a property access like `myObject.a` may result in an `undefined` value if either the explicit `undefined` is stored there or the `a` property doesn't exist at all. So, if the value is the same in both cases, how else do we distinguish them?

We can ask an object if it has a certain property *without* asking to get that property's value:

```
var myObject = {
  a: 2
};

("a" in myObject);           // true
("b" in myObject);           // false

myObject.hasOwnProperty("a"); // true
myObject.hasOwnProperty("b"); // false
```

The `in` operator will check to see if the property is *in* the object, or if it exists at any higher level of the `[[Prototype]]` chain object traversal (see Chapter 5). By contrast, `hasOwnProperty(..)` checks to see if *only* `myObject` has the property or not, and will *not* consult the `[[Prototype]]` chain. We'll come back to the important differences between these two operations in Chapter 5 when we explore `[[Prototype]]`s in detail.

`hasOwnProperty(..)` is accessible for all normal objects via delegation to `Object.prototype` (see Chapter 5). But it's possible to create an object that does not link to `Object.prototype` (via `Object.create(null)` -- see Chapter 5). In this case, a method call like `myObject.hasOwnProperty(..)` would fail.

In that scenario, a more robust way of performing such a check is `Object.prototype.hasOwnProperty.call(myObject, "a")`, which borrows the base `hasOwnProperty(..)` method and uses *explicit this binding* (see Chapter 2) to apply it against our `myObject`.

**Note:** The `in` operator has the appearance that it will check for the existence of a *value* inside a container, but it actually checks for the existence of a property name. This difference is important to note with respect to arrays, as the temptation to try a check like `4 in [2, 4, 6]` is strong, but this will not behave as expected.

## Enumeration

Previously, we explained briefly the idea of "enumerability" when we looked at the `enumerable` property descriptor characteristic. Let's revisit that and examine it in more close detail.

```
var myObject = { };

Object.defineProperty(
  myObject,
  "a",
  // make `a` enumerable, as normal
  { enumerable: true, value: 2 }
);

Object.defineProperty(
  myObject,
  "b",
  // make `b` NON-enumerable
  { enumerable: false, value: 3 }
);

myObject.b; // 3
("b" in myObject); // true
myObject.hasOwnProperty("b"); // true

// .....

for (var k in myObject) {
  console.log( k, myObject[k] );
```

```
}
```

You'll notice that `myObject.b` in fact **exists** and has an accessible value, but it doesn't show up in a `for..in` loop (though, surprisingly, it **is** revealed by the `in` operator existence check). That's because "enumerable" basically means "will be included if the object's properties are iterated through".

**Note:** `for..in` loops applied to arrays can give somewhat unexpected results, in that the enumeration of an array will include not only all the numeric indices, but also any enumerable properties. It's a good idea to use `for..in` loops *only* on objects, and traditional `for` loops with numeric index iteration for the values stored in arrays.

Another way that enumerable and non-enumerable properties can be distinguished:

```
var myObject = { };

Object.defineProperty(
  myObject,
  "a",
  // make `a` enumerable, as normal
  { enumerable: true, value: 2 }
);

Object.defineProperty(
  myObject,
  "b",
  // make `b` non-enumerable
  { enumerable: false, value: 3 }
);

myObject.propertyIsEnumerable( "a" ); // true
myObject.propertyIsEnumerable( "b" ); // false

Object.keys( myObject ); // ["a"]
Object.getOwnPropertyNames( myObject ); // ["a", "b"]
```

`propertyIsEnumerable(..)` tests whether the given property name exists *directly* on the object and is also `enumerable:true`.

`Object.keys(..)` returns an array of all enumerable properties, whereas `Object.getOwnPropertyNames(..)` returns an array of *all* properties, enumerable or not.

Whereas `in` vs. `hasOwnProperty(..)` differ in whether they consult the `[[Prototype]]` chain or not, `Object.keys(..)` and `Object.getOwnPropertyNames(..)` both inspect *only* the direct object specified.

There's (currently) no built-in way to get a list of **all properties** which is equivalent to what the `in` operator test would consult (traversing all properties on the entire `[[Prototype]]` chain, as explained in Chapter 5). You could approximate such a utility by recursively traversing the `[[Prototype]]` chain of an object, and for each level, capturing the list from `Object.keys(..)` -- only enumerable properties.

## Iteration

The `for..in` loop iterates over the list of enumerable properties on an object (including its `[[Prototype]]` chain). But what if you instead want to iterate over the values?

With numerically-indexed arrays, iterating over the values is typically done with a standard `for` loop, like:

```
var myArray = [1, 2, 3];

for (var i = 0; i < myArray.length; i++) {
  console.log( myArray[i] );
}
```

```
// 1 2 3
```

This isn't iterating over the values, though, but iterating over the indices, where you then use the index to reference the value, as `myArray[i]`.

ES5 also added several iteration helpers for arrays, including `forEach(..)`, `every(..)`, and `some(..)`. Each of these helpers accepts a function callback to apply to each element in the array, differing only in how they respectively respond to a return value from the callback.

`forEach(..)` will iterate over all values in the array, and ignores any callback return values. `every(..)` keeps going until the end or the callback returns a `false` (or "falsy") value, whereas `some(..)` keeps going until the end or the callback returns a `true` (or "truthy") value.

These special return values inside `every(..)` and `some(..)` act somewhat like a `break` statement inside a normal `for` loop, in that they stop the iteration early before it reaches the end.

If you iterate on an object with a `for..in` loop, you're also only getting at the values indirectly, because it's actually iterating only over the enumerable properties of the object, leaving you to access the properties manually to get the values.

**Note:** As contrasted with iterating over an array's indices in a numerically ordered way (`for` loop or other iterators), the order of iteration over an object's properties is **not guaranteed** and may vary between different JS engines. **Do not rely** on any observed ordering for anything that requires consistency among environments, as any observed agreement is unreliable.

But what if you want to iterate over the values directly instead of the array indices (or object properties)? Helpfully, ES6 adds a `for..of` loop syntax for iterating over arrays (and objects, if the object defines its own custom iterator):

```
var myArray = [ 1, 2, 3 ];

for (var v of myArray) {
  console.log( v );
}

// 1
// 2
// 3
```

The `for..of` loop asks for an iterator object (from a default internal function known as `@@iterator` in spec-speak) of the *thing* to be iterated, and the loop then iterates over the successive return values from calling that iterator object's `next()` method, once for each loop iteration.

Arrays have a built-in `@@iterator`, so `for..of` works easily on them, as shown. But let's manually iterate the array, using the built-in `@@iterator`, to see how it works:

```
var myArray = [ 1, 2, 3 ];
var it = myArray[Symbol.iterator]();

it.next(); // { value:1, done:false }
it.next(); // { value:2, done:false }
it.next(); // { value:3, done:false }
it.next(); // { done:true }
```

**Note:** We get at the `@@iterator` *internal property* of an object using an ES6 `Symbol : Symbol.iterator`. We briefly mentioned `Symbol` semantics earlier in the chapter (see "Computed Property Names"), so the same reasoning applies here. You'll always want to reference such special properties by `Symbol` name reference instead of by the special value it may hold. Also, despite the name's implications, `@@iterator` is **not the iterator object** itself, but a **function that returns** the iterator object -- a subtle but important detail!

As the above snippet reveals, the return value from an iterator's `next()` call is an object of the form `{ value: .. , done: .. }`, where `value` is the current iteration value, and `done` is a `boolean` that indicates if there's more to iterate.

Notice the value `3` was returned with a `done:false`, which seems strange at first glance. You have to call the `next()` a fourth time (which the `for..of` loop in the previous snippet automatically does) to get `done:true` and know you're truly done iterating. The reason for this quirk is beyond the scope of what we'll discuss here, but it comes from the semantics of ES6 generator functions.

While arrays do automatically iterate in `for..of` loops, regular objects **do not have a built-in `@@iterator`**. The reasons for this intentional omission are more complex than we will examine here, but in general it was better to not include some implementation that could prove troublesome for future types of objects.

It is possible to define your own default `@@iterator` for any object that you care to iterate over. For example:

```
var myObject = {
  a: 2,
  b: 3
};

Object.defineProperty( myObject, Symbol.iterator, {
  enumerable: false,
  writable: false,
  configurable: true,
  value: function() {
    var o = this;
    var idx = 0;
    var ks = Object.keys( o );
    return {
      next: function() {
        return {
          value: o[ks[idx++]],
          done: (idx > ks.length)
        };
      }
    };
  }
} );

// iterate `myObject` manually
var it = myObject[Symbol.iterator]();
it.next(); // { value:2, done:false }
it.next(); // { value:3, done:false }
it.next(); // { value:undefined, done:true }

// iterate `myObject` with `for..of`
for (var v of myObject) {
  console.log( v );
}
// 2
// 3
```

**Note:** We used `Object.defineProperty(..)` to define our custom `@@iterator` (mostly so we could make it non-enumerable), but using the `Symbol` as a *computed property name* (covered earlier in this chapter), we could have declared it directly, like `var myObject = { a:2, b:3, [Symbol.iterator]: function(){ /* .. */ } }`.

Each time the `for..of` loop calls `next()` on `myObject`'s iterator object, the internal pointer will advance and return back the next value from the object's properties list (see a previous note about iteration ordering on object properties/values).

The iteration we just demonstrated is a simple value-by-value iteration, but you can of course define arbitrarily complex iterations for your custom data structures, as you see fit. Custom iterators combined with ES6's `for..of` loop are a powerful new syntactic tool for manipulating user-defined objects.

For example, a list of `Pixel` objects (with `x` and `y` coordinate values) could decide to order its iteration based on the linear distance from the `(0,0)` origin, or filter out points that are "too far away", etc. As long as your iterator returns the expected `{ value: ... }` return values from `next()` calls, and a `{ done: true }` after the iteration is complete, ES6's `for..of` can

iterate over it.

In fact, you can even generate "infinite" iterators which never "finish" and always return a new value (such as a random number, an incremented value, a unique identifier, etc), though you probably will not use such iterators with an unbounded `for..of` loop, as it would never end and would hang your program.

```
var randoms = {
  [Symbol.iterator]: function() {
    return {
      next: function() {
        return { value: Math.random() };
      }
    };
  }
};

var randoms_pool = [];
for (var n of randoms) {
  randoms_pool.push( n );

  // don't proceed unbounded!
  if (randoms_pool.length === 100) break;
}
```

This iterator will generate random numbers "forever", so we're careful to only pull out 100 values so our program doesn't hang.

## Review (TL;DR)

Objects in JS have both a literal form (such as `var a = { .. }`) and a constructed form (such as `var a = new Array(..)`). The literal form is almost always preferred, but the constructed form offers, in some cases, more creation options.

Many people mistakenly claim "everything in JavaScript is an object", but this is incorrect. Objects are one of the 6 (or 7, depending on your perspective) primitive types. Objects have sub-types, including `function`, and also can be behavior-specialized, like `[object Array]` as the internal label representing the array object sub-type.

Objects are collections of key/value pairs. The values can be accessed as properties, via `.propName` or `["propName"]` syntax. Whenever a property is accessed, the engine actually invokes the internal default `[[Get]]` operation (and `[[Put]]` for setting values), which not only looks for the property directly on the object, but which will traverse the `[[Prototype]]` chain (see Chapter 5) if not found.

Properties have certain characteristics that can be controlled through property descriptors, such as `writable` and `configurable`. In addition, objects can have their mutability (and that of their properties) controlled to various levels of immutability using `Object.preventExtensions(..)`, `Object.seal(..)`, and `Object.freeze(..)`.

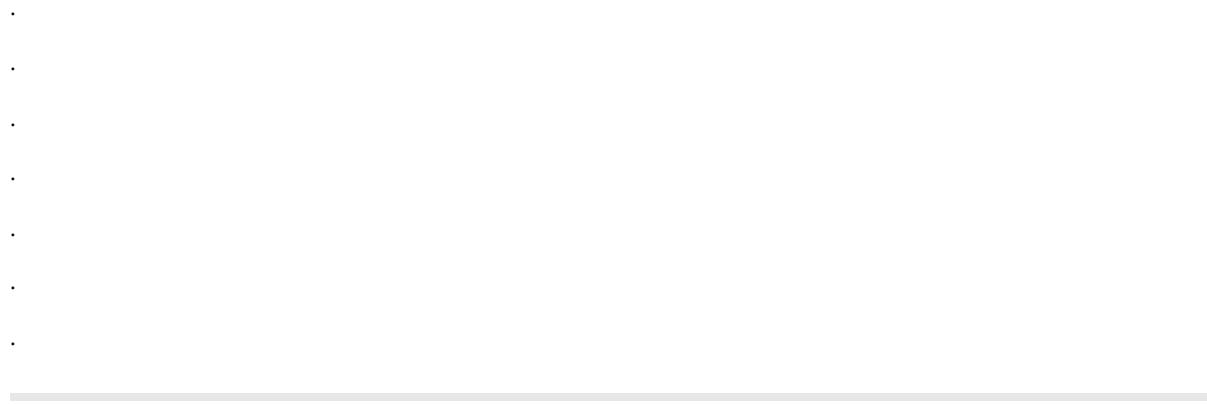
Properties don't have to contain values -- they can be "accessor properties" as well, with getters/setters. They can also be either `enumerable` or not, which controls if they show up in `for..in` loop iterations, for instance.

You can also iterate over **the values** in data structures (arrays, objects, etc) using the ES6 `for..of` syntax, which looks for either a built-in or custom `@@iterator` object consisting of a `next()` method to advance through the data values one at a time.

# You Don't Know JS yet: Objects & Classes - 2nd Edition

## Chapter 4: Mixing (Up) "Class" Objects

<b>NOTE:</b>
Work in progress



<b>NOTE:</b>
Everything below here is previous text from 1st edition, and is only here for reference while 2nd edition work is underway. <b>Please ignore this stuff.</b>

Following our exploration of objects from the previous chapter, it's natural that we now turn our attention to "object oriented (OO) programming", with "classes". We'll first look at "class orientation" as a design pattern, before examining the mechanics of "classes": "instantiation", "inheritance" and "(relative) polymorphism".

We'll see that these concepts don't really map very naturally to the object mechanism in JS, and the lengths (mixins, etc.) many JavaScript developers go to overcome such challenges.

**Note:** This chapter spends quite a bit of time (the first half!) on heavy "object oriented programming" theory. We eventually relate these ideas to real concrete JavaScript code in the second half, when we talk about "Mixins". But there's a lot of concept and pseudo-code to wade through first, so don't get lost -- just stick with it!

## Class Theory

"Class/Inheritance" describes a certain form of code organization and architecture -- a way of modeling real world problem domains in our software.

OO or class oriented programming stresses that data intrinsically has associated behavior (of course, different depending on the type and nature of the data!) that operates on it, so proper design is to package up (aka, encapsulate) the data and the behavior together. This is sometimes called "data structures" in formal computer science.

For example, a series of characters that represents a word or phrase is usually called a "string". The characters are the data. But you almost never just care about the data, you usually want to *do things* with the data, so the behaviors that can apply to that data (calculating its length, appending data, searching, etc.) are all designed as methods of a `String` class.

Any given string is just an instance of this class, which means that it's a neatly collected packaging of both the character data and the functionality we can perform on it.

Classes also imply a way of *classifying* a certain data structure. The way we do this is to think about any given structure as a specific variation of a more general base definition.

Let's explore this classification process by looking at a commonly cited example. A *car* can be described as a specific implementation of a more general "class" of thing, called a *vehicle*.

We model this relationship in software with classes by defining a `Vehicle` class and a `Car` class.

The definition of `Vehicle` might include things like propulsion (engines, etc.), the ability to carry people, etc., which would all be the behaviors. What we define in `Vehicle` is all the stuff that is common to all (or most of) the different types of vehicles (the "planes, trains, and automobiles").

It might not make sense in our software to re-define the basic essence of "ability to carry people" over and over again for each different type of vehicle. Instead, we define that capability once in `Vehicle`, and then when we define `Car`, we simply indicate that it "inherits" (or "extends") the base definition from `Vehicle`. The definition of `Car` is said to specialize the general `Vehicle` definition.

While `Vehicle` and `Car` collectively define the behavior by way of methods, the data in an instance would be things like the unique VIN of a specific car, etc.

#### **And thus, classes, inheritance, and instantiation emerge.**

Another key concept with classes is "polymorphism", which describes the idea that a general behavior from a parent class can be overridden in a child class to give it more specifics. In fact, relative polymorphism lets us reference the base behavior from the overridden behavior.

Class theory strongly suggests that a parent class and a child class share the same method name for a certain behavior, so that the child overrides the parent (differentially). As we'll see later, doing so in your JavaScript code is opting into frustration and code brittleness.

## **"Class" Design Pattern**

You may never have thought about classes as a "design pattern", since it's most common to see discussion of popular "OO Design Patterns", like "Iterator", "Observer", "Factory", "Singleton", etc. As presented this way, it's almost an assumption that OO classes are the lower-level mechanics by which we implement all (higher level) design patterns, as if OO is a given foundation for *all* (proper) code.

Depending on your level of formal education in programming, you may have heard of "procedural programming" as a way of describing code which only consists of procedures (aka, functions) calling other functions, without any higher abstractions. You may have been taught that classes were the *proper* way to transform procedural-style "spaghetti code" into well-formed, well-organized code.

Of course, if you have experience with "functional programming" (Monads, etc.), you know very well that classes are just one of several common design patterns. But for others, this may be the first time you've asked yourself if classes really are a fundamental foundation for code, or if they are an optional abstraction on top of code.

Some languages (like Java) don't give you the choice, so it's not very *optional* at all -- everything's a class. Other languages like C/C++ or PHP give you both procedural and class-oriented syntaxes, and it's left more to the developer's choice which style or mixture of styles is appropriate.

## **JavaScript "Classes"**

Where does JavaScript fall in this regard? JS has had *some* class-like syntactic elements (like `new` and `instanceof`) for quite awhile, and more recently in ES6, some additions, like the `class` keyword (see Appendix A).

But does that mean JavaScript actually *has* classes? Plain and simple: **No.**

Since classes are a design pattern, you *can*, with quite a bit of effort (as we'll see throughout the rest of this chapter), implement approximations for much of classical class functionality. JS tries to satisfy the extremely pervasive *desire* to design with classes by providing seemingly class-like syntax.

While we may have a syntax that looks like classes, it's as if JavaScript mechanics are fighting against you using the *class design pattern*, because behind the curtain, the mechanisms that you build on are operating quite differently. Syntactic sugar and (extremely widely used) JS "Class" libraries go a long way toward hiding this reality from you, but sooner or later you will face the fact that the *classes* you have in other languages are not like the "classes" you're faking in JS.

What this boils down to is that classes are an optional pattern in software design, and you have the choice to use them in JavaScript or not. Since many developers have a strong affinity to class oriented software design, we'll spend the rest of this chapter exploring what it takes to maintain the illusion of classes with what JS provides, and the pain points we experience.

## Class Mechanics

In many class-oriented languages, the "standard library" provides a "stack" data structure (push, pop, etc.) as a `Stack` class. This class would have an internal set of variables that stores the data, and it would have a set of publicly accessible behaviors ("methods") provided by the class, which gives your code the ability to interact with the (hidden) data (adding & removing data, etc.).

But in such languages, you don't really operate directly on `Stack` (unless making a `Static` class member reference, which is outside the scope of our discussion). The `Stack` class is merely an abstract explanation of what *any* "stack" should do, but it's not itself a "stack". You must **instantiate** the `Stack` class before you have a concrete data structure *thing* to operate against.

## Building

The traditional metaphor for "class" and "instance" based thinking comes from a building construction.

An architect plans out all the characteristics of a building: how wide, how tall, how many windows and in what locations, even what type of material to use for the walls and roof. She doesn't necessarily care, at this point, *where* the building will be built, nor does she care *how many* copies of that building will be built.

She also doesn't care very much about the contents of the building -- the furniture, wall paper, ceiling fans, etc. -- only what type of structure they will be contained by.

The architectural blue-prints she produces are only *plans* for a building. They don't actually constitute a building we can walk into and sit down. We need a builder for that task. A builder will take those plans and follow them, exactly, as he *builds* the building. In a very real sense, he is *copying* the intended characteristics from the plans to the physical building.

Once complete, the building is a physical instantiation of the blue-print plans, hopefully an essentially perfect *copy*. And then the builder can move to the open lot next door and do it all over again, creating yet another *copy*.

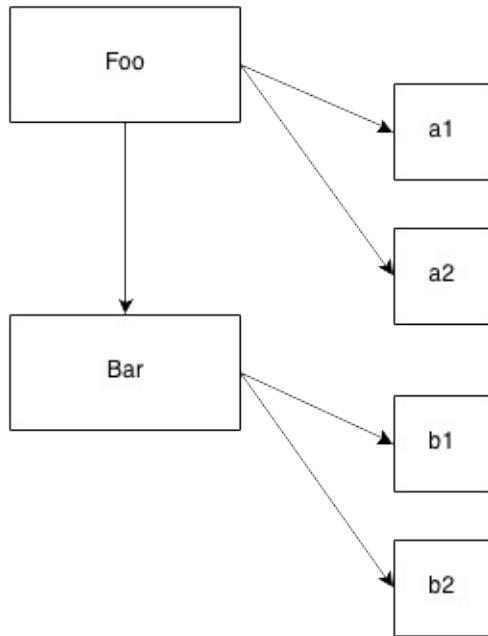
The relationship between building and blue-print is indirect. You can examine a blue-print to understand how the building was structured, for any parts where direct inspection of the building itself was insufficient. But if you want to open a door, you have to go to the building itself -- the blue-print merely has lines drawn on a page that *represent* where the door should be.

A class is a blue-print. To actually *get* an object we can interact with, we must build (aka, "instantiate") something from the class. The end result of such "construction" is an object, typically called an "instance", which we can directly call methods on and access any public data properties from, as necessary.

**This object is a copy** of all the characteristics described by the class.

You likely wouldn't expect to walk into a building and find, framed and hanging on the wall, a copy of the blue-prints used to plan the building, though the blue-prints are probably on file with a public records office. Similarly, you don't generally use an object instance to directly access and manipulate its class, but it is usually possible to at least determine *which class* an object instance comes from.

It's more useful to consider the direct relationship of a class to an object instance, rather than any indirect relationship between an object instance and the class it came from. **A class is instantiated into object form by a copy operation.**



As you can see, the arrows move from left to right, and from top to bottom, which indicates the copy operations that occur, both conceptually and physically.

## Constructor

Instances of classes are constructed by a special method of the class, usually of the same name as the class, called a *constructor*. This method's explicit job is to initialize any information (state) the instance will need.

For example, consider this loose pseudo-code (invented syntax) for classes:

```

class CoolGuy {
    specialTrick = nothing

    CoolGuy( trick ) {
        specialTrick = trick
    }

    showOff() {
        output( "Here's my trick: ", specialTrick )
    }
}
  
```

To make a `CoolGuy` instance, we would call the class constructor:

```

Joe = new CoolGuy( "jumping rope" )

Joe.showOff() // Here's my trick: jumping rope
  
```

Notice that the `CoolGuy` class has a constructor `CoolGuy()`, which is actually what we call when we say `new CoolGuy(..)`. We get an object back (an instance of our class) from the constructor, and we can call the method `showOff()`, which prints out that particular `CoolGuy`'s special trick.

*Obviously, jumping rope makes Joe a pretty cool guy.*

The constructor of a class *belongs* to the class, almost universally with the same name as the class. Also, constructors pretty much always need to be called with `new` to let the language engine know you want to construct a *new* class instance.

## Class Inheritance

In class-oriented languages, not only can you define a class which can be instantiated itself, but you can define another class that **inherits** from the first class.

The second class is often said to be a "child class" whereas the first is the "parent class". These terms obviously come from the metaphor of parents and children, though the metaphors here are a bit stretched, as you'll see shortly.

When a parent has a biological child, the genetic characteristics of the parent are copied into the child. Obviously, in most biological reproduction systems, there are two parents who co-equally contribute genes to the mix. But for the purposes of the metaphor, we'll assume just one parent.

Once the child exists, he or she is separate from the parent. The child was heavily influenced by the inheritance from his or her parent, but is unique and distinct. If a child ends up with red hair, that doesn't mean the parent's hair *was* or automatically *becomes* red.

In a similar way, once a child class is defined, it's separate and distinct from the parent class. The child class contains an initial copy of the behavior from the parent, but can then override any inherited behavior and even define new behavior.

It's important to remember that we're talking about parent and child **classes**, which aren't physical things. This is where the metaphor of parent and child gets a little confusing, because we actually should say that a parent class is like a parent's DNA and a child class is like a child's DNA. We have to make (aka "instantiate") a person out of each set of DNA to actually have a physical person to have a conversation with.

Let's set aside biological parents and children, and look at inheritance through a slightly different lens: different types of vehicles. That's one of the most canonical (and often groan-worthy) metaphors to understand inheritance.

Let's revisit the `Vehicle` and `car` discussion from earlier in this chapter. Consider this loose pseudo-code (invented syntax) for inherited classes:

```
class Vehicle {
    engines = 1

    ignition() {
        output( "Turning on my engine." )
    }

    drive() {
        ignition()
        output( "Steering and moving forward!" )
    }
}

class Car inherits Vehicle {
    wheels = 4

    drive() {
        inherited:drive()
        output( "Rolling on all ", wheels, " wheels!" )
    }
}

class SpeedBoat inherits Vehicle {
    engines = 2

    ignition() {
        output( "Turning on my ", engines, " engines." )
    }
}
```

```

pilot() {
    inherited:drive()
    output( "Speeding through the water with ease!" )
}
}

```

**Note:** For clarity and brevity, constructors for these classes have been omitted.

We define the `Vehicle` class to assume an engine, a way to turn on the ignition, and a way to drive around. But you wouldn't ever manufacture just a generic "vehicle", so it's really just an abstract concept at this point.

So then we define two specific kinds of vehicle: `car` and `SpeedBoat`. They each inherit the general characteristics of `Vehicle`, but then they specialize the characteristics appropriately for each kind. A car needs 4 wheels, and a speed boat needs 2 engines, which means it needs extra attention to turn on the ignition of both engines.

## Polymorphism

`car` defines its own `drive()` method, which overrides the method of the same name it inherited from `Vehicle`. But then, `Car`'s `drive()` method calls `inherited:drive()`, which indicates that `car` can reference the original pre-overridden `drive()` it inherited. `SpeedBoat`'s `pilot()` method also makes a reference to its inherited copy of `drive()`.

This technique is called "polymorphism", or "virtual polymorphism". More specifically to our current point, we'll call it "relative polymorphism".

Polymorphism is a much broader topic than we will exhaust here, but our current "relative" semantics refers to one particular aspect: the idea that any method can reference another method (of the same or different name) at a higher level of the inheritance hierarchy. We say "relative" because we don't absolutely define which inheritance level (aka, class) we want to access, but rather relatively reference it by essentially saying "look one level up".

In many languages, the keyword `super` is used, in place of this example's `inherited:`, which leans on the idea that a "super class" is the parent/ancestor of the current class.

Another aspect of polymorphism is that a method name can have multiple definitions at different levels of the inheritance chain, and these definitions are automatically selected as appropriate when resolving which methods are being called.

We see two occurrences of that behavior in our example above: `drive()` is defined in both `Vehicle` and `car`, and `ignition()` is defined in both `Vehicle` and `SpeedBoat`.

**Note:** Another thing that traditional class-oriented languages give you via `super` is a direct way for the constructor of a child class to reference the constructor of its parent class. This is largely true because with real classes, the constructor belongs to the class. However, in JS, it's the reverse -- it's actually more appropriate to think of the "class" belonging to the constructor (the `Foo.prototype...` type references). Since in JS the relationship between child and parent exists only between the two `.prototype` objects of the respective constructors, the constructors themselves are not directly related, and thus there's no simple way to relatively reference one from the other (see Appendix A for ES6 `class` which "solves" this with `super`).

An interesting implication of polymorphism can be seen specifically with `ignition()`. Inside `pilot()`, a relative-polymorphic reference is made to (the inherited) `Vehicle`'s version of `drive()`. But that `drive()` references an `ignition()` method just by name (no relative reference).

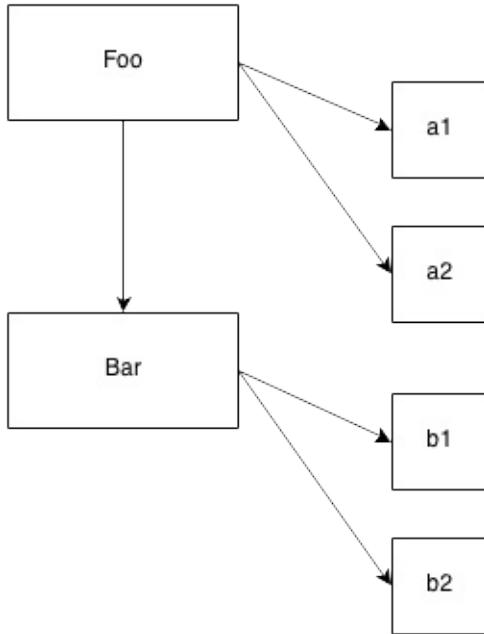
Which version of `ignition()` will the language engine use, the one from `Vehicle` or the one from `SpeedBoat`? **It uses the `SpeedBoat` version of `ignition()`**. If you were to instantiate `Vehicle` class itself, and then call its `drive()`, the language engine would instead just use `Vehicle`'s `ignition()` method definition.

Put another way, the definition for the method `ignition()` *polymorphs* (changes) depending on which class (level of inheritance) you are referencing an instance of.

This may seem like overly deep academic detail. But understanding these details is necessary to properly contrast similar (but distinct) behaviors in JavaScript's `[[Prototype]]` mechanism.

When classes are inherited, there is a way **for the classes themselves** (not the object instances created from them!) to *relatively* reference the class inherited from, and this relative reference is usually called `super`.

Remember this figure from earlier:



Notice how for both instantiation (`a1`, `a2`, `b1`, and `b2`) *and* inheritance (`Bar`), the arrows indicate a copy operation.

Conceptually, it would seem a child class `Bar` can access behavior in its parent class `Foo` using a relative polymorphic reference (aka, `super`). However, in reality, the child class is merely given a copy of the inherited behavior from its parent class. If the child "overrides" a method it inherits, both the original and overridden versions of the method are actually maintained, so that they are both accessible.

Don't let polymorphism confuse you into thinking a child class is linked to its parent class. A child class instead gets a copy of what it needs from the parent class. **Class inheritance implies copies**.

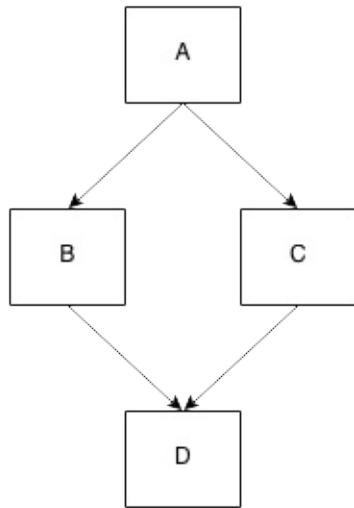
## Multiple Inheritance

Recall our earlier discussion of parent(s) and children and DNA? We said that the metaphor was a bit weird because biologically most offspring come from two parents. If a class could inherit from two other classes, it would more closely fit the parent/child metaphor.

Some class-oriented languages allow you to specify more than one "parent" class to "inherit" from. Multiple-inheritance means that each parent class definition is copied into the child class.

On the surface, this seems like a powerful addition to class-orientation, giving us the ability to compose more functionality together. However, there are certainly some complicating questions that arise. If both parent classes provide a method called `drive()`, which version would a `drive()` reference in the child resolve to? Would you always have to manually specify which parent's `drive()` you meant, thus losing some of the gracefulness of polymorphic inheritance?

There's another variation, the so called "Diamond Problem", which refers to the scenario where a child class "D" inherits from two parent classes ("B" and "C"), and each of those in turn inherits from a common "A" parent. If "A" provides a method `drive()`, and both "B" and "C" override (polymorph) that method, when `D` references `drive()`, which version should it use (`B:drive()` or `C:drive()`)?



These complications go even much deeper than this quick glance. We address them here only so we can contrast to how JavaScript's mechanisms work.

JavaScript is simpler: it does not provide a native mechanism for "multiple inheritance". Many see this as a good thing, because the complexity savings more than make up for the "reduced" functionality. But this doesn't stop developers from trying to fake it in various ways, as we'll see next.

## Mixins

JavaScript's object mechanism does not *automatically* perform copy behavior when you "inherit" or "instantiate". Plainly, there are no "classes" in JavaScript to instantiate, only objects. And objects don't get copied to other objects, they get *linked together* (more on that in Chapter 5).

Since observed class behaviors in other languages imply copies, let's examine how JS developers **fake** the *missing* copy behavior of classes in JavaScript: mixins. We'll look at two types of "mixin": **explicit** and **implicit**.

### Explicit Mixins

Let's again revisit our `Vehicle` and `car` example from before. Since JavaScript will not automatically copy behavior from `Vehicle` to `car`, we can instead create a utility that manually copies. Such a utility is often called `extend(..)` by many libraries/frameworks, but we will call it `Mixin(..)` here for illustrative purposes.

```

// vastly simplified `Mixin(..)` example:
function mixin( sourceObj, targetObj ) {
    for ( var key in sourceObj ) {
        // only copy if not already present
        if (!(key in targetObj)) {
            targetObj[key] = sourceObj[key];
        }
    }
    return targetObj;
}

var Vehicle = {
    engines: 1,

    ignition: function() {
        console.log( "Turning on my engine." );
    },

    drive: function() {
  
```

```

        this.ignition();
        console.log( "Steering and moving forward!" );
    }
};

var Car = mixin( Vehicle, {
    wheels: 4,

    drive: function() {
        Vehicle.drive.call( this );
        console.log( "Rolling on all " + this.wheels + " wheels!" );
    }
} );

```

**Note:** Subtly but importantly, we're not dealing with classes anymore, because there are no classes in JavaScript. `Vehicle` and `car` are just objects that we make copies from and to, respectively.

`Car` now has a copy of the properties and functions from `Vehicle`. Technically, functions are not actually duplicated, but rather *references* to the functions are copied. So, `car` now has a property called `ignition`, which is a copied reference to the `ignition()` function, as well as a property called `engines` with the copied value of `1` from `Vehicle`.

`Car` already had a `drive` property (function), so that property reference was not overridden (see the `if` statement in `mixin(...)` above).

## "Polymorphism" Revisited

Let's examine this statement: `Vehicle.drive.call( this )`. This is what I call "explicit pseudo-polymorphism". Recall in our previous pseudo-code this line was `inherited:drive()`, which we called "relative polymorphism".

JavaScript does not have (prior to ES6; see Appendix A) a facility for relative polymorphism. So, **because both `car` and `Vehicle` had a function of the same name: `drive()`**, to distinguish a call to one or the other, we must make an absolute (not relative) reference. We explicitly specify the `Vehicle` object by name, and call the `drive()` function on it.

But if we said `Vehicle.drive()`, the `this` binding for that function call would be the `Vehicle` object instead of the `Car` object (see Chapter 2), which is not what we want. So, instead we use `.call( this )` (Chapter 2) to ensure that `drive()` is executed in the context of the `Car` object.

**Note:** If the function name identifier for `Car.drive()` hadn't overlapped with (aka, "shadowed"; see Chapter 5) `Vehicle.drive()`, we wouldn't have been exercising "method polymorphism". So, a reference to `Vehicle.drive()` would have been copied over by the `mixin(...)` call, and we could have accessed directly with `this.drive()`. The chosen identifier overlap **shadowing** is why we have to use the more complex *explicit pseudo-polymorphism* approach.

In class-oriented languages, which have relative polymorphism, the linkage between `Car` and `Vehicle` is established once, at the top of the class definition, which makes for only one place to maintain such relationships.

But because of JavaScript's peculiarities, explicit pseudo-polymorphism (because of shadowing!) creates brittle manual/explicit linkage **in every single function where you need such a (pseudo-)polymorphic reference**. This can significantly increase the maintenance cost. Moreover, while explicit pseudo-polymorphism can emulate the behavior of "multiple inheritance", it only increases the complexity and brittleness.

The result of such approaches is usually more complex, harder-to-read, *and* harder-to-maintain code. **Explicit pseudo-polymorphism should be avoided wherever possible**, because the cost outweighs the benefit in most respects.

## Mixing Copies

Recall the `mixin(...)` utility from above:

```
// vastly simplified `mixin()` example:
```

```

function mixin( sourceObj, targetObj ) {
  for (var key in sourceObj) {
    // only copy if not already present
    if (!(key in targetObj)) {
      targetObj[key] = sourceObj[key];
    }
  }

  return targetObj;
}

```

Now, let's examine how `mixin(..)` works. It iterates over the properties of `sourceObj` (`Vehicle` in our example) and if there's no matching property of that name in `targetObj` (`car` in our example), it makes a copy. Since we're making the copy after the initial object exists, we are careful to not copy over a target property.

If we made the copies first, before specifying the `car` specific contents, we could omit this check against `targetObj`, but that's a little more clunky and less efficient, so it's generally less preferred:

```

// alternate mixin, less "safe" to overwrites
function mixin( sourceObj, targetObj ) {
  for (var key in sourceObj) {
    targetObj[key] = sourceObj[key];
  }

  return targetObj;
}

var Vehicle = {
  // ...
};

// first, create an empty object with
// Vehicle's stuff copied in
var Car = mixin( Vehicle, { } );

// now copy the intended contents into Car
mixin( {
  wheels: 4,

  drive: function() {
    // ...
  }
}, Car );

```

Either approach, we have explicitly copied the non-overlapping contents of `Vehicle` into `Car`. The name "mixin" comes from an alternate way of explaining the task: `Car` has `Vehicle`'s contents **mixed-in**, just like you mix in chocolate chips into your favorite cookie dough.

As a result of the copy operation, `Car` will operate somewhat separately from `Vehicle`. If you add a property onto `Car`, it will not affect `Vehicle`, and vice versa.

**Note:** A few minor details have been skimmed over here. There are still some subtle ways the two objects can "affect" each other even after copying, such as if they both share a reference to a common object (such as an array).

Since the two objects also share references to their common functions, that means that **even manual copying of functions (aka, mixins) from one object to another doesn't actually emulate the real duplication from class to instance that occurs in class-oriented languages**.

JavaScript functions can't really be duplicated (in a standard, reliable way), so what you end up with instead is a **duplicated reference** to the same shared function object (functions are objects; see Chapter 3). If you modified one of the shared **function objects** (like `ignition()`) by adding properties on top of it, for instance, both `Vehicle` and `Car` would be "affected" via the shared reference.

Explicit mixins are a fine mechanism in JavaScript. But they appear more powerful than they really are. Not much benefit is *actually* derived from copying a property from one object to another, **as opposed to just defining the properties twice**, once on each object. And that's especially true given the function-object reference nuance we just mentioned.

If you explicitly mix-in two or more objects into your target object, you can **partially emulate** the behavior of "multiple inheritance", but there's no direct way to handle collisions if the same method or property is being copied from more than one source. Some developers/libraries have come up with "late binding" techniques and other exotic work-arounds, but fundamentally these "tricks" are *usually* more effort (and lesser performance!) than the pay-off.

Take care only to use explicit mixins where it actually helps make more readable code, and avoid the pattern if you find it making code that's harder to trace, or if you find it creates unnecessary or unwieldy dependencies between objects.

**If it starts to get harder to properly use mixins than before you used them**, you should probably stop using mixins. In fact, if you have to use a complex library/utility to work out all these details, it might be a sign that you're going about it the harder way, perhaps unnecessarily. In Chapter 6, we'll try to distill a simpler way that accomplishes the desired outcomes without all the fuss.

## Parasitic Inheritance

A variation on this explicit mixin pattern, which is both in some ways explicit and in other ways implicit, is called "parasitic inheritance", popularized mainly by Douglas Crockford.

Here's how it can work:

```
// "Traditional JS Class" `Vehicle`
function Vehicle() {
    this.engines = 1;
}
Vehicle.prototype.ignition = function() {
    console.log( "Turning on my engine." );
};
Vehicle.prototype.drive = function() {
    this.ignition();
    console.log( "Steering and moving forward!" );
};

// "Parasitic Class" `Car`
function Car() {
    // first, `car` is a `Vehicle`
    var car = new Vehicle();

    // now, let's modify our `car` to specialize it
    car.wheels = 4;

    // save a privileged reference to `Vehicle::drive()`
    var vehDrive = car.drive;

    // override `Vehicle::drive()`
    car.drive = function() {
        vehDrive.call( this );
        console.log( "Rolling on all " + this.wheels + " wheels!" );
    };
}

return car;
}

var myCar = new Car();

myCar.drive();
// Turning on my engine.
// Steering and moving forward!
// Rolling on all 4 wheels!
```

As you can see, we initially make a copy of the definition from the `Vehicle` "parent class" (object), then mixin our "child class" (object) definition (preserving privileged parent-class references as needed), and pass off this composed object `car` as our child instance.

**Note:** when we call `new car()`, a new object is created and referenced by `Car`'s `this` reference (see Chapter 2). But since we don't use that object, and instead return our own `car` object, the initially created object is just discarded. So, `car()` could be called without the `new` keyword, and the functionality above would be identical, but without the wasted object creation/garbage-collection.

## Implicit Mixins

Implicit mixins are closely related to *explicit pseudo-polymorphism* as explained previously. As such, they come with the same caveats and warnings.

Consider this code:

```
var Something = {
  cool: function() {
    this.greeting = "Hello World";
    this.count = this.count ? this.count + 1 : 1;
  }
};

Something.cool();
Something.greeting; // "Hello World"
Something.count; // 1

var Another = {
  cool: function() {
    // implicit mixin of `Something` to `Another`
    Something.cool.call( this );
  }
};

Another.cool();
Another.greeting; // "Hello World"
Another.count; // 1 (not shared state with `Something`)
```

With `Something.cool.call( this )`, which can happen either in a "constructor" call (most common) or in a method call (shown here), we essentially "borrow" the function `Something.cool()` and call it in the context of `Another` (via its `this` binding; see Chapter 2) instead of `Something`. The end result is that the assignments that `Something.cool()` makes are applied against the `Another` object rather than the `Something` object.

So, it is said that we "mixed in" `Something`'s behavior with (or into) `Another`.

While this sort of technique seems to take useful advantage of `this` rebinding functionality, it is the brittle `Something.cool.call( this )` call, which cannot be made into a relative (and thus more flexible) reference, that you should **heed with caution**. Generally, **avoid such constructs where possible** to keep cleaner and more maintainable code.

## Review (TL;DR)

Classes are a design pattern. Many languages provide syntax which enables natural class-oriented software design. JS also has a similar syntax, but it behaves **very differently** from what you're used to with classes in those other languages.

**Classes mean copies.**

When traditional classes are instantiated, a copy of behavior from class to instance occurs. When classes are inherited, a copy of behavior from parent to child also occurs.

Polymorphism (having different functions at multiple levels of an inheritance chain with the same name) may seem like it implies a referential relative link from child back to parent, but it's still just a result of copy behavior.

JavaScript **does not automatically** create copies (as classes imply) between objects.

The mixin pattern (both explicit and implicit) is often used to *sort of* emulate class copy behavior, but this usually leads to ugly and brittle syntax like explicit pseudo-polymorphism (`otherObj.methodName.call(this, ...)`), which often results in harder to understand and maintain code.

Explicit mixins are also not exactly the same as class *copy*, since objects (and functions!) only have shared references duplicated, not the objects/functions duplicated themselves. Not paying attention to such nuance is the source of a variety of gotchas.

In general, faking classes in JS often sets more landmines for future coding than solving present *real* problems.

# You Don't Know JS Yet: Objects & Classes - 2nd Edition

## Chapter 5: Prototypes

**NOTE:**

Work in progress

**NOTE:**

Everything below here is previous text from 1st edition, and is only here for reference while 2nd edition work is underway.  
**Please ignore this stuff.**

In Chapters 3 and 4, we mentioned the `[[Prototype]]` chain several times, but haven't said what exactly it is. We will now examine prototypes in detail.

**Note:** All of the attempts to emulate class-copy behavior, as described previously in Chapter 4, labeled as variations of "mixins", completely circumvent the `[[Prototype]]` chain mechanism we examine here in this chapter.

### [[Prototype]]

Objects in JavaScript have an internal property, denoted in the specification as `[[Prototype]]`, which is simply a reference to another object. Almost all objects are given a non-`null` value for this property, at the time of their creation.

**Note:** We will see shortly that it *is* possible for an object to have an empty `[[Prototype]]` linkage, though this is somewhat less common.

Consider:

```
var myObject = {
  a: 2
};

myObject.a; // 2
```

What is the `[[Prototype]]` reference used for? In Chapter 3, we examined the `[[Get]]` operation that is invoked when you reference a property on an object, such as `myObject.a`. For that default `[[Get]]` operation, the first step is to check if the object itself has a property `a` on it, and if so, it's used.

**Note:** ES6 Proxies are outside of our discussion scope in this book (will be covered in a later book in the series!), but everything we discuss here about normal `[[Get]]` and `[[Put]]` behavior does not apply if a `Proxy` is involved.

But it's what happens if a `isn't` present on `myObject` that brings our attention now to the `[[Prototype]]` link of the object.

The default `[[Get]]` operation proceeds to follow the `[[Prototype]]` link of the object if it cannot find the requested property on the object directly.

```
var anotherObject = {
  a: 2
};

// create an object linked to `anotherObject`
var myObject = Object.create( anotherObject );

myObject.a; // 2
```

**Note:** We will explain what `Object.create(...)` does, and how it operates, shortly. For now, just assume it creates an object with the `[[Prototype]]` linkage we're examining to the object specified.

So, we have `myObject` that is now `[[Prototype]]` linked to `anotherObject`. Clearly `myObject.a` doesn't actually exist, but nevertheless, the property access succeeds (being found on `anotherObject` instead) and indeed finds the value `2`.

But, if a weren't found on `anotherObject` either, its `[[Prototype]]` chain, if non-empty, is again consulted and followed.

This process continues until either a matching property name is found, or the `[[Prototype]]` chain ends. If no matching property is ever found by the end of the chain, the return result from the `[[Get]]` operation is `undefined`.

Similar to this `[[Prototype]]` chain look-up process, if you use a `for..in` loop to iterate over an object, any property that can be reached via its chain (and is also `enumerable` -- see Chapter 3) will be enumerated. If you use the `in` operator to test for the existence of a property on an object, `in` will check the entire chain of the object (regardless of *enumerability*).

```
var anotherObject = {
  a: 2
};

// create an object linked to `anotherObject`
var myObject = Object.create( anotherObject );

for (var k in myObject) {
  console.log("found: " + k);
}
// found: a

("a" in myObject); // true
```

So, the `[[Prototype]]` chain is consulted, one link at a time, when you perform property look-ups in various fashions. The look-up stops once the property is found or the chain ends.

## Object.prototype

But where exactly does the `[[Prototype]]` chain "end"?

The top-end of every *normal* `[[Prototype]]` chain is the built-in `Object.prototype`. This object includes a variety of common utilities used all over JS, because all normal (built-in, not host-specific extension) objects in JavaScript "descend from" (aka, have at the top of their `[[Prototype]]` chain) the `Object.prototype` object.

Some utilities found here you may be familiar with include `.toString()` and `.valueOf()`. In Chapter 3, we introduced another: `.hasOwnProperty(..)`. And yet another function on `Object.prototype` you may not be familiar with, but which we'll address later in this chapter, is `.isPrototypeOf(..)`.

## Setting & Shadowing Properties

Back in Chapter 3, we mentioned that setting properties on an object was more nuanced than just adding a new property to the object or changing an existing property's value. We will now revisit this situation more completely.

```
myObject.foo = "bar";
```

If the `myObject` object already has a normal data accessor property called `foo` directly present on it, the assignment is as simple as changing the value of the existing property.

If `foo` is not already present directly on `myObject`, the `[[Prototype]]` chain is traversed, just like for the `[[Get]]` operation. If `foo` is not found anywhere in the chain, the property `foo` is added directly to `myObject` with the specified value, as expected.

However, if `foo` is already present somewhere higher in the chain, nuanced (and perhaps surprising) behavior can occur with the `myObject.foo = "bar"` assignment. We'll examine that more in just a moment.

If the property name `foo` ends up both on `myObject` itself and at a higher level of the `[[Prototype]]` chain that starts at `myObject`, this is called *shadowing*. The `foo` property directly on `myObject` *shadows* any `foo` property which appears higher in the chain, because the `myObject.foo` look-up would always find the `foo` property that's lowest in the chain.

As we just hinted, shadowing `foo` on `myObject` is not as simple as it may seem. We will now examine three scenarios for the `myObject.foo = "bar"` assignment when `foo` is **not** already on `myObject` directly, but **is** at a higher level of `myObject`'s `[[Prototype]]` chain:

1. If a normal data accessor (see Chapter 3) property named `foo` is found anywhere higher on the `[[Prototype]]` chain, **and it's not marked as read-only (`writable:false`)** then a new property called `foo` is added directly to `myObject`, resulting in a **shadowed property**.
2. If a `foo` is found higher on the `[[Prototype]]` chain, but it's marked as **read-only (`writable:false`)**, then both the setting of that existing property as well as the creation of the shadowed property on `myObject` **are disallowed**. If the code is running in `strict mode`, an error will be thrown. Otherwise, the setting of the property value will silently be ignored. Either way, **no shadowing occurs**.
3. If a `foo` is found higher on the `[[Prototype]]` chain and it's a setter (see Chapter 3), then the setter will always be called. No `foo` will be added to (aka, shadowed on) `myObject`, nor will the `foo` setter be redefined.

Most developers assume that assignment of a property (`[[Put]]`) will always result in shadowing if the property already exists higher on the `[[Prototype]]` chain, but as you can see, that's only true in one (#1) of the three situations just described.

If you want to shadow `foo` in cases #2 and #3, you cannot use `=` assignment, but must instead use `Object.defineProperty(...)` (see Chapter 3) to add `foo` to `myObject`.

**Note:** Case #2 may be the most surprising of the three. The presence of a *read-only* property prevents a property of the same name being implicitly created (shadowed) at a lower level of a `[[Prototype]]` chain. The reason for this restriction is primarily to reinforce the illusion of class-inherited properties. If you think of the `foo` at a higher level of the chain as having been inherited (copied down) to `myObject`, then it makes sense to enforce the non-writable nature of that `foo` property on `myObject`. If you however separate the illusion from the fact, and recognize that no such inheritance copying *actually* occurred (see Chapters 4 and 5), it's a little unnatural that `myObject` would be prevented from having a `foo` property just because some other object had a non-writable `foo` on it. It's even stranger that this restriction only applies to `=` assignment, but is not enforced when using `Object.defineProperty(...)`.

Shadowing with **methods** leads to ugly *explicit pseudo-polymorphism* (see Chapter 4) if you need to delegate between them. Usually, shadowing is more complicated and nuanced than it's worth, so **you should try to avoid it if possible**. See Chapter 6 for an alternative design pattern, which among other things discourages shadowing in favor of cleaner alternatives.

Shadowing can even occur implicitly in subtle ways, so care must be taken if trying to avoid it. Consider:

```

var anotherObject = {
  a: 2
};

var myObject = Object.create( anotherObject );

anotherObject.a; // 2
myObject.a; // 2

anotherObject.hasOwnProperty( "a" ); // true
myObject.hasOwnProperty( "a" ); // false

myObject.a++; // oops, implicit shadowing!

anotherObject.a; // 2
myObject.a; // 3

myObject.hasOwnProperty( "a" ); // true

```

Though it may appear that `myObject.a++` should (via delegation) look-up and just increment the `anotherObject.a` property itself *in place*, instead the `++` operation corresponds to `myObject.a = myObject.a + 1`. The result is `[[Get]]` looking up a property via `[[Prototype]]` to get the current value `2` from `anotherObject.a`, incrementing the value by one, then `[[Put]]` assigning the `3` value to a new shadowed property `a` on `myObject`. Oops!

Be very careful when dealing with delegated properties that you modify. If you wanted to increment `anotherObject.a`, the only proper way is `anotherObject.a++`.

## "Class"

At this point, you might be wondering: "Why does one object need to link to another object?" What's the real benefit? That is a very appropriate question to ask, but we must first understand what `[[Prototype]]` is **not** before we can fully understand and appreciate what it *is* and how it's useful.

As we explained in Chapter 4, in JavaScript, there are no abstract patterns/blueprints for objects called "classes" as there are in class-oriented languages. JavaScript **just** has objects.

In fact, JavaScript is **almost unique** among languages as perhaps the only language with the right to use the label "object oriented", because it's one of a very short list of languages where an object can be created directly, without a class at all.

In JavaScript, classes can't (being that they don't exist!) describe what an object can do. The object defines its own behavior directly. **There's just the object.**

## "Class" Functions

There's a peculiar kind of behavior in JavaScript that has been shamelessly abused for years to *hack* something that *looks* like "classes". We'll examine this approach in detail.

The peculiar "sort-of class" behavior hinges on a strange characteristic of functions: all functions by default get a public, non-enumerable (see Chapter 3) property on them called `prototype`, which points at an otherwise arbitrary object.

```

function Foo() {
  // ...
}

Foo.prototype; // { }

```

This object is often called "Foo's prototype", because we access it via an unfortunately-named `Foo.prototype` property reference. However, that terminology is hopelessly destined to lead us into confusion, as we'll see shortly. Instead, I will call it "the object formerly known as Foo's prototype". Just kidding. How about: "object arbitrarily labeled 'Foo dot prototype'"?

Whatever we call it, what exactly is this object?

The most direct way to explain it is that each object created from calling `new Foo()` (see Chapter 2) will end up (somewhat arbitrarily) `[[Prototype]]`-linked to this "Foo dot prototype" object.

Let's illustrate:

```
function Foo() {
    // ...
}

var a = new Foo();

Object.getPrototypeOf(a) === Foo.prototype; // true
```

When `a` is created by calling `new Foo()`, one of the things (see Chapter 2 for all *four* steps) that happens is that `a` gets an internal `[[Prototype]]` link to the object that `Foo.prototype` is pointing at.

Stop for a moment and ponder the implications of that statement.

In class-oriented languages, multiple **copies** (aka, "instances") of a class can be made, like stamping something out from a mold. As we saw in Chapter 4, this happens because the process of instantiating (or inheriting from) a class means, "copy the behavior plan from that class into a physical object", and this is done again for each new instance.

But in JavaScript, there are no such copy-actions performed. You don't create multiple instances of a class. You can create multiple objects that `[[Prototype]]` *link* to a common object. But by default, no copying occurs, and thus these objects don't end up totally separate and disconnected from each other, but rather, quite **linked**.

`new Foo()` results in a new object (we called it `a`), and **that** new object `a` is internally `[[Prototype]]` linked to the `Foo.prototype` object.

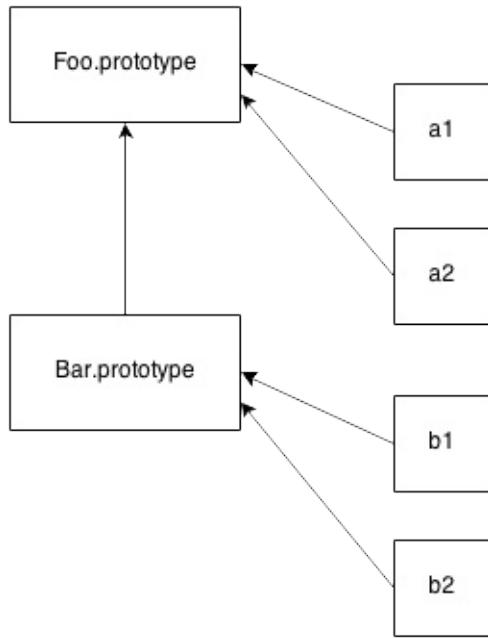
**We end up with two objects, linked to each other.** That's *it*. We didn't instantiate a class. We certainly didn't do any copying of behavior from a "class" into a concrete object. We just caused two objects to be linked to each other.

In fact, the secret, which eludes most JS developers, is that the `new Foo()` function calling had really almost nothing *direct* to do with the process of creating the link. **It was sort of an accidental side-effect.** `new Foo()` is an indirect, round-about way to end up with what we want: **a new object linked to another object**.

Can we get what we want in a more *direct* way? **Yes!** The hero is `Object.create(..)`. But we'll get to that in a little bit.

## What's in a name?

In JavaScript, we don't make *copies* from one object ("class") to another ("instance"). We make *links* between objects. For the `[[Prototype]]` mechanism, visually, the arrows move from right to left, and from bottom to top.



This mechanism is often called "prototypal inheritance" (we'll explore the code in detail shortly), which is commonly said to be the dynamic-language version of "classical inheritance". It's an attempt to piggy-back on the common understanding of what "inheritance" means in the class-oriented world, but *tweak* (**read: pave over**) the understood semantics, to fit dynamic scripting.

The word "inheritance" has a very strong meaning (see Chapter 4), with plenty of mental precedent. Merely adding "prototypal" in front to distinguish the *actually nearly opposite* behavior in JavaScript has left in its wake nearly two decades of miry confusion.

I like to say that sticking "prototypal" in front of "inheritance" to drastically reverse its actual meaning is like holding an orange in one hand, an apple in the other, and insisting on calling the apple a "red orange". No matter what confusing label I put in front of it, that doesn't change the *fact* that one fruit is an apple and the other is an orange.

The better approach is to plainly call an apple an apple -- to use the most accurate and direct terminology. That makes it easier to understand both their similarities and their **many differences**, because we all have a simple, shared understanding of what "apple" means.

Because of the confusion and conflation of terms, I believe the label "prototypal inheritance" itself (and trying to mis-apply all its associated class-orientation terminology, like "class", "constructor", "instance", "polymorphism", etc) has done **more harm than good** in explaining how JavaScript's mechanism *really* works.

"Inheritance" implies a *copy* operation, and JavaScript doesn't copy object properties (natively, by default). Instead, JS creates a link between two objects, where one object can essentially *delegate* property/function access to another object. "Delegation" (see Chapter 6) is a much more accurate term for JavaScript's object-linking mechanism.

Another term which is sometimes thrown around in JavaScript is "differential inheritance". The idea here is that we describe an object's behavior in terms of what is *different* from a more general descriptor. For example, you explain that a car is a kind of vehicle, but one that has exactly 4 wheels, rather than re-describing all the specifics of what makes up a general vehicle (engine, etc).

If you try to think of any given object in JS as the sum total of all behavior that is *available* via delegation, and **in your mind you flatten** all that behavior into one tangible *thing*, then you can (sorta) see how "differential inheritance" might fit.

But just like with "prototypal inheritance", "differential inheritance" pretends that your mental model is more important than what is physically happening in the language. It overlooks the fact that object `B` is not actually differentially constructed, but is instead built with specific characteristics defined, alongside "holes" where nothing is defined. It is in these "holes" (gaps in, or lack of, definition) that delegation *can* take over and, on the fly, "fill them in" with delegated behavior.

The object is not, by native default, flattened into the single differential object, **through copying**, that the mental model of "differential inheritance" implies. As such, "differential inheritance" is just not as natural a fit for describing how JavaScript's `[[Prototype]]` mechanism actually works.

You *can* choose to prefer the "differential inheritance" terminology and mental model, as a matter of taste, but there's no denying the fact that it *only* fits the mental acrobatics in your mind, not the physical behavior in the engine.

## "Constructors"

Let's go back to some earlier code:

```
function Foo() {
  // ...
}

var a = new Foo();
```

What exactly leads us to think `Foo` is a "class"?

For one, we see the use of the `new` keyword, just like class-oriented languages do when they construct class instances. For another, it appears that we are in fact executing a *constructor* method of a class, because `Foo()` is actually a method that gets called, just like how a real class's constructor gets called when you instantiate that class.

To further the confusion of "constructor" semantics, the arbitrarily labeled `Foo.prototype` object has another trick up its sleeve. Consider this code:

```
function Foo() {
  // ...
}

Foo.prototype.constructor === Foo; // true

var a = new Foo();
a.constructor === Foo; // true
```

The `Foo.prototype` object by default (at declaration time on line 1 of the snippet!) gets a public, non-enumerable (see Chapter 3) property called `.constructor`, and this property is a reference back to the function (`Foo` in this case) that the object is associated with. Moreover, we see that object `a` created by the "constructor" call `new Foo()` *seems* to also have a property on it called `.constructor` which similarly points to "the function which created it".

**Note:** This is not actually true. `a` has no `.constructor` property on it, and though `a.constructor` does in fact resolve to the `Foo` function, "constructor" **does not actually mean** "was constructed by", as it appears. We'll explain this strangeness shortly.

Oh, yeah, also... by convention in the JavaScript world, "class"es are named with a capital letter, so the fact that it's `Foo` instead of `foo` is a strong clue that we intend it to be a "class". That's totally obvious to you, right!?

**Note:** This convention is so strong that many JS linters actually *complain* if you call `new` on a method with a lowercase name, or if we don't call `new` on a function that happens to start with a capital letter. That sort of boggles the mind that we struggle so much to get (fake) "class-orientation" *right* in JavaScript that we create linter rules to ensure we use capital letters, even though the capital letter doesn't mean **anything at all** to the JS engine.

## Constructor Or Call?

In the above snippet, it's tempting to think that `Foo` is a "constructor", because we call it with `new` and we observe that it "constructs" an object.

In reality, `Foo` is no more a "constructor" than any other function in your program. Functions themselves are **not** constructors. However, when you put the `new` keyword in front of a normal function call, that makes that function call a "constructor call". In fact, `new` sort of hijacks any normal function and calls it in a fashion that constructs an object, **in addition to whatever else it was going to do**.

For example:

```
function NothingSpecial() {
    console.log( "Don't mind me!" );
}

var a = new NothingSpecial();
// "Don't mind me!"

a; // {}
```

`NothingSpecial` is just a plain old normal function, but when called with `new`, it *constructs* an object, almost as a side-effect, which we happen to assign to `a`. The `call` was a *constructor call*, but `NothingSpecial` is not, in and of itself, a *constructor*.

In other words, in JavaScript, it's most appropriate to say that a "constructor" is **any function called with the `new` keyword in front of it**.

Functions aren't constructors, but function calls are "constructor calls" if and only if `new` is used.

## Mechanics

Are *those* the only common triggers for ill-fated "class" discussions in JavaScript?

**Not quite.** JS developers have strived to simulate as much as they can of class-orientation:

```
function Foo(name) {
    this.name = name;
}

Foo.prototype.myName = function() {
    return this.name;
};

var a = new Foo( "a" );
var b = new Foo( "b" );

a.myName(); // "a"
b.myName(); // "b"
```

This snippet shows two additional "class-orientation" tricks in play:

1. `this.name = name` : adds the `.name` property onto each object (`a` and `b`, respectively; see Chapter 2 about `this` binding), similar to how class instances encapsulate data values.
2. `Foo.prototype.myName = ...` : perhaps the more interesting technique, this adds a property (function) to the `Foo.prototype` object. Now, `a.myName()` works, but perhaps surprisingly. How?

In the above snippet, it's strongly tempting to think that when `a` and `b` are created, the properties/functions on the `Foo.prototype` object are *copied* over to each of `a` and `b` objects. **However, that's not what happens.**

At the beginning of this chapter, we explained the `[[Prototype]]` link, and how it provides the fall-back look-up steps if a property reference isn't found directly on an object, as part of the default `[[Get]]` algorithm.

So, by virtue of how they are created, `a` and `b` each end up with an internal `[[Prototype]]` linkage to `Foo.prototype`. When `myName` is not found on `a` or `b`, respectively, it's instead found (through delegation, see Chapter 6) on `Foo.prototype`.

## "Constructor" Redux

Recall the discussion from earlier about the `.constructor` property, and how it *seems* like `a.constructor === Foo` being true means that `a` has an actual `.constructor` property on it, pointing at `Foo`? **Not correct.**

This is just unfortunate confusion. In actuality, the `.constructor` reference is also *delegated* up to `Foo.prototype`, which **happens to**, by default, have a `.constructor` that points at `Foo`.

It *seems* awfully convenient that an object `a` "constructed by" `Foo` would have access to a `.constructor` property that points to `Foo`. But that's nothing more than a false sense of security. It's a happy accident, almost tangentially, that `a.constructor` *happens to* point at `Foo` via this default `[[Prototype]]` delegation. There are actually several ways that the ill-fated assumption of `.constructor` meaning "was constructed by" can come back to bite you.

For one, the `.constructor` property on `Foo.prototype` is only there by default on the object created when `foo` the function is declared. If you create a new object, and replace a function's default `.prototype` object reference, the new object will not by default magically get a `.constructor` on it.

Consider:

```
function Foo() { /* ... */ }

Foo.prototype = { /* ... */ }; // create a new prototype object

var a1 = new Foo();
a1.constructor === Foo; // false!
a1.constructor === Object; // true!
```

`Object(..)` didn't "construct" `a1` did it? It sure seems like `Foo()` "constructed" it. Many developers think of `Foo()` as doing the construction, but where everything falls apart is when you think "constructor" means "was constructed by", because by that reasoning, `a1.constructor` should be `Foo`, but it isn't!

What's happening? `a1` has no `.constructor` property, so it delegates up the `[[Prototype]]` chain to `Foo.prototype`. But that object doesn't have a `.constructor` either (like the default `Foo.prototype` object would have had!), so it keeps delegating, this time up to `Object.prototype`, the top of the delegation chain. *That* object indeed has a `.constructor` on it, which points to the built-in `Object(..)` function.

### Misconception, busted.

Of course, you can add `.constructor` back to the `Foo.prototype` object, but this takes manual work, especially if you want to match native behavior and have it be non-enumerable (see Chapter 3).

For example:

```
function Foo() { /* ... */ }

Foo.prototype = { /* ... */ }; // create a new prototype object

// Need to properly "fix" the missing `constructor` property on the new object serving as `Foo.prototype`.
// See Chapter 3 for `defineProperty(..)`.

Object.defineProperty(Foo.prototype, "constructor", {
  enumerable: false,
  writable: true,
  configurable: true,
  value: Foo // point `constructor` at `Foo`
});
```

That's a lot of manual work to fix `.constructor`. Moreover, all we're really doing is perpetuating the misconception that "constructor" means "was constructed by". That's an expensive illusion.

The fact is, `.constructor` on an object arbitrarily points, by default, at a function who, reciprocally, has a reference back to the object -- a reference which it calls `.prototype`. The words "constructor" and "prototype" only have a loose default meaning that might or might not hold true later. The best thing to do is remind yourself, "constructor does not mean constructed by".

`.constructor` is not a magic immutable property. It is non-enumerable (see snippet above), but its value is writable (can be changed), and moreover, you can add or overwrite (intentionally or accidentally) a property of the name `constructor` on any object in any `[[Prototype]]` chain, with any value you see fit.

By virtue of how the `[[Get]]` algorithm traverses the `[[Prototype]]` chain, a `.constructor` property reference found anywhere may resolve quite differently than you'd expect.

See how arbitrary its meaning actually is?

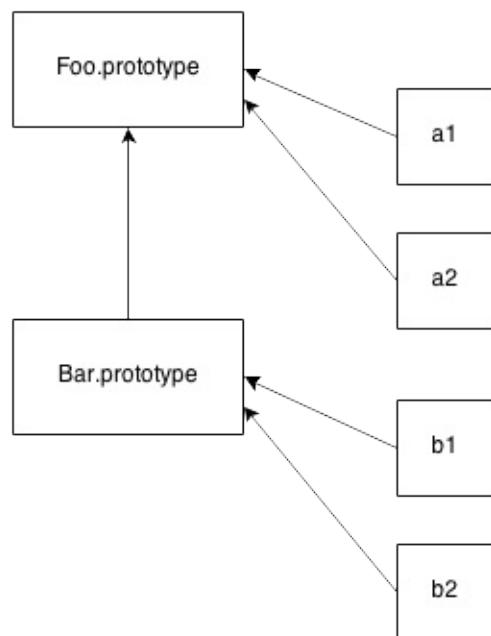
The result? Some arbitrary object-property reference like `a1.constructor` cannot actually be *trusted* to be the assumed default function reference. Moreover, as we'll see shortly, just by simple omission, `a1.constructor` can even end up pointing somewhere quite surprising and insensible.

`.constructor` is extremely unreliable, and an unsafe reference to rely upon in your code. **Generally, such references should be avoided where possible.**

## "(Prototypal) Inheritance"

We've seen some approximations of "class" mechanics as typically hacked into JavaScript programs. But JavaScript "class"es would be rather hollow if we didn't have an approximation of "inheritance".

Actually, we've already seen the mechanism which is commonly called "prototypal inheritance" at work when `a` was able to "inherit from" `Foo.prototype`, and thus get access to the `myName()` function. But we traditionally think of "inheritance" as being a relationship between two "classes", rather than between "class" and "instance".



Recall this figure from earlier, which shows not only delegation from an object (aka, "instance") `a1` to object `Foo.prototype`, but from `Bar.prototype` to `Foo.prototype`, which somewhat resembles the concept of Parent-Child class inheritance.

*Resembles*, except of course for the direction of the arrows, which show these are delegation links rather than copy operations.

And, here's the typical "prototype style" code that creates such links:

```
function Foo(name) {
```

```

    this.name = name;
}

Foo.prototype.myName = function() {
    return this.name;
};

function Bar(name,label) {
    Foo.call( this, name );
    this.label = label;
}

// here, we make a new `Bar.prototype`
// linked to `Foo.prototype`
Bar.prototype = Object.create( Foo.prototype );

// Beware! Now `Bar.prototype.constructor` is gone,
// and might need to be manually "fixed" if you're
// in the habit of relying on such properties!

Bar.prototype.myLabel = function() {
    return this.label;
};

var a = new Bar( "a", "obj a" );

a.myName(); // "a"
a.myLabel(); // "obj a"

```

**Note:** To understand why `this` points to `a` in the above code snippet, see Chapter 2.

The important part is `Bar.prototype = Object.create( Foo.prototype )`. `Object.create(...)` creates a "new" object out of thin air, and links that new object's internal `[[Prototype]]` to the object you specify (`Foo.prototype` in this case).

In other words, that line says: "make a *new* 'Bar dot prototype' object that's linked to 'Foo dot prototype'."

When `function Bar() { ... }` is declared, `Bar`, like any other function, has a `.prototype` link to its default object. But *that* object is not linked to `Foo.prototype` like we want. So, we create a *new* object that *is* linked as we want, effectively throwing away the original incorrectly-linked object.

**Note:** A common mis-conception/confusion here is that either of the following approaches would *also* work, but they do not work as you'd expect:

```

// doesn't work like you want!
Bar.prototype = Foo.prototype;

// works kinda like you want, but with
// side-effects you probably don't want :(
Bar.prototype = new Foo();

```

`Bar.prototype = Foo.prototype` doesn't create a new object for `Bar.prototype` to be linked to. It just makes `Bar.prototype` be another reference to `Foo.prototype`, which effectively links `Bar` directly to **the same object as** `Foo` links to: `Foo.prototype`. This means when you start assigning, like `Bar.prototype.myLabel = ...`, you're modifying **not a separate object** but *the shared* `Foo.prototype` object itself, which would affect any objects linked to `Foo.prototype`. This is almost certainly not what you want. If it *is* what you want, then you likely don't need `Bar` at all, and should just use only `Foo` and make your code simpler.

`Bar.prototype = new Foo()` **does in fact** create a new object which is duly linked to `Foo.prototype` as we'd want. But, it uses the `Foo(...)` "constructor call" to do it. If that function has any side-effects (such as logging, changing state, registering against other objects, **adding data properties to** `this`, etc.), those side-effects happen at the time of this linking (and likely against the wrong object!), rather than only when the eventual `Bar()` "descendants" are created, as would likely be expected.

So, we're left with using `Object.create(..)` to make a new object that's properly linked, but without having the side-effects of calling `Foo(..)`. The slight downside is that we have to create a new object, throwing the old one away, instead of modifying the existing default object we're provided.

It would be *nice* if there was a standard and reliable way to modify the linkage of an existing object. Prior to ES6, there's a non-standard and not fully-cross-browser way, via the `.__proto__` property, which is settable. ES6 adds a `Object.setPrototypeOf(..)` helper utility, which does the trick in a standard and predictable way.

Compare the pre-ES6 and ES6-standardized techniques for linking `Bar.prototype` to `Foo.prototype`, side-by-side:

```
// pre-ES6
// throws away default existing `Bar.prototype`
Bar.prototype = Object.create(Foo.prototype);

// ES6+
// modifies existing `Bar.prototype`
Object.setPrototypeOf(Bar.prototype, Foo.prototype);
```

Ignoring the slight performance disadvantage (throwing away an object that's later garbage collected) of the `Object.create(..)` approach, it's a little bit shorter and may be perhaps a little easier to read than the ES6+ approach. But it's probably a syntactic wash either way.

## Inspecting "Class" Relationships

What if you have an object like `a` and want to find out what object (if any) it delegates to? Inspecting an instance (just an object in JS) for its inheritance ancestry (delegation linkage in JS) is often called *introspection* (or *reflection*) in traditional class-oriented environments.

Consider:

```
function Foo() {
  // ...
}

Foo.prototype.blah = ...;

var a = new Foo();
```

How do we then introspect `a` to find out its "ancestry" (delegation linkage)? The first approach embraces the "class" confusion:

```
a instanceof Foo; // true
```

The `instanceof` operator takes a plain object as its left-hand operand and a **function** as its right-hand operand. The question `instanceof` answers is: **in the entire [[Prototype]] chain of `a`, does the object arbitrarily pointed to by `Foo.prototype` ever appear?**

Unfortunately, this means that you can only inquire about the "ancestry" of some object (`a`) if you have some **function** (`Foo`, with its attached `.prototype` reference) to test with. If you have two arbitrary objects, say `a` and `b`, and want to find out if *the objects* are related to each other through a `[[Prototype]]` chain, `instanceof` alone can't help.

**Note:** If you use the built-in `.bind(..)` utility to make a hard-bound function (see Chapter 2), the function created will not have a `.prototype` property. Using `instanceof` with such a function transparently substitutes the `.prototype` of the *target function* that the hard-bound function was created from.

It's fairly uncommon to use hard-bound functions as "constructor calls", but if you do, it will behave as if the original *target function* was invoked instead, which means that using `instanceof` with a hard-bound function also behaves according to the original function.

This snippet illustrates the ridiculousness of trying to reason about relationships between **two objects** using "class" semantics and `instanceof` :

```
// helper utility to see if `o1` is
// related to (delegates to) `o2`
function isRelatedTo(o1, o2) {
  function F(){}
  F.prototype = o2;
  return o1 instanceof F;
}

var a = {};
var b = Object.create( a );

isRelatedTo( b, a ); // true
```

Inside `isRelatedTo(..)` , we borrow a throw-away function `F` , reassign its `.prototype` to arbitrarily point to some object `o2` , then ask if `o1` is an "instance of" `F` . Obviously `o1` isn't *actually* inherited or descended or even constructed from `F` , so it should be clear why this kind of exercise is silly and confusing. **The problem comes down to the awkwardness of class semantics forced upon JavaScript**, in this case as revealed by the indirect semantics of `instanceof` .

The second, and much cleaner, approach to `[[Prototype]]` reflection is:

```
Foo.prototype.isPrototypeOf( a ); // true
```

Notice that in this case, we don't really care about (or even *need*) `Foo` , we just need an **object** (in our case, arbitrarily labeled `Foo.prototype` ) to test against another **object**. The question `isPrototypeOf(..)` answers is: **in the entire `[[Prototype]]` chain of `a` , does `Foo.prototype` ever appear?**

Same question, and exact same answer. But in this second approach, we don't actually need the indirection of referencing a **function** (`Foo` ) whose `.prototype` property will automatically be consulted.

We *just need two objects* to inspect a relationship between them. For example:

```
// Simply: does `b` appear anywhere in
// `c`'s [[Prototype]] chain?
b.isPrototypeOf( c );
```

Notice, this approach doesn't require a function ("class") at all. It just uses object references directly to `b` and `c` , and inquires about their relationship. In other words, our `isRelatedTo(..)` utility above is built-in to the language, and it's called `isPrototypeOf(..)` .

We can also directly retrieve the `[[Prototype]]` of an object. As of ES5, the standard way to do this is:

```
Object.getPrototypeOf( a );
```

And you'll notice that object reference is what we'd expect:

```
Object.getPrototypeOf( a ) === Foo.prototype; // true
```

Most browsers (not all!) have also long supported a non-standard alternate way of accessing the internal `[[Prototype]]` :

```
a.__proto__ === Foo.prototype; // true
```

The strange `.__proto__` (not standardized until ES6!) property "magically" retrieves the internal `[[Prototype]]` of an object as a reference, which is quite helpful if you want to directly inspect (or even traverse: `.__proto__.__proto__...` ) the chain.

Just as we saw earlier with `.constructor`, `.__proto__` doesn't actually exist on the object you're inspecting (`a` in our running example). In fact, it exists (non-enumerable; see Chapter 2) on the built-in `Object.prototype`, along with the other common utilities (`.toString()`, `.isPrototypeOf(..)`, etc).

Moreover, `.__proto__` looks like a property, but it's actually more appropriate to think of it as a getter/setter (see Chapter 3).

Roughly, we could envision `.__proto__` implemented (see Chapter 3 for object property definitions) like this:

```
Object.defineProperty( Object.prototype, "__proto__", {
  get: function() {
    return Object.getPrototypeOf( this );
  },
  set: function(o) {
    // setPrototypeOf(..) as of ES6
    Object.setPrototypeOf( this, o );
    return o;
  }
} );
```

So, when we access (retrieve the value of) `a.__proto__`, it's like calling `a.__proto__()` (calling the getter function). That function call has `a` as its `this` even though the getter function exists on the `Object.prototype` object (see Chapter 2 for `this` binding rules), so it's just like saying `Object.getPrototypeOf(a)`.

`.__proto__` is also a settable property, just like using ES6's `Object.setPrototypeOf(..)` shown earlier. However, generally you **should not change the `[[Prototype]]` of an existing object**.

There are some very complex, advanced techniques used deep in some frameworks that allow tricks like "subclassing" an `Array`, but this is commonly frowned on in general programming practice, as it usually leads to *much* harder to understand/maintain code.

**Note:** As of ES6, the `class` keyword will allow something that approximates "subclassing" of built-in's like `Array`. See Appendix A for discussion of the `class` syntax added in ES6.

The only other narrow exception (as mentioned earlier) would be setting the `[[Prototype]]` of a default function's `.prototype` object to reference some other object (besides `Object.prototype`). That would avoid replacing that default object entirely with a new linked object. Otherwise, **it's best to treat object `[[Prototype]]` linkage as a read-only characteristic** for ease of reading your code later.

**Note:** The JavaScript community unofficially coined a term for the double-underscore, specifically the leading one in properties like `__proto__ : "dunder"`. So, the "cool kids" in JavaScript would generally pronounce `__proto__` as "dunder proto".

## Object Links

As we've now seen, the `[[Prototype]]` mechanism is an internal link that exists on one object which references some other object.

This linkage is (primarily) exercised when a property/method reference is made against the first object, and no such property/method exists. In that case, the `[[Prototype]]` linkage tells the engine to look for the property/method on the linked-to object. In turn, if that object cannot fulfill the look-up, its `[[Prototype]]` is followed, and so on. This series of links between objects forms what is called the "prototype chain".

### Create( ) ing Links

We've thoroughly debunked why JavaScript's `[[Prototype]]` mechanism is **not** like *classes*, and we've seen how it instead creates **links** between proper objects.

What's the point of the `[[Prototype]]` mechanism? Why is it so common for JS developers to go to so much effort (emulating classes) in their code to wire up these linkages?

Remember we said much earlier in this chapter that `Object.create(..)` would be a hero? Now, we're ready to see how.

```
var foo = {
    something: function() {
        console.log( "Tell me something good..." );
    }
};

var bar = Object.create( foo );

bar.something(); // Tell me something good...
```

`Object.create(..)` creates a new object (`bar`) linked to the object we specified (`foo`), which gives us all the power (delegation) of the `[[Prototype]]` mechanism, but without any of the unnecessary complication of `new` functions acting as classes and constructor calls, confusing `.prototype` and `.constructor` references, or any of that extra stuff.

**Note:** `Object.create(null)` creates an object that has an empty (aka, `null`) `[[Prototype]]` linkage, and thus the object can't delegate anywhere. Since such an object has no prototype chain, the `instanceof` operator (explained earlier) has nothing to check, so it will always return `false`. These special empty- `[[Prototype]]` objects are often called "dictionaries" as they are typically used purely for storing data in properties, mostly because they have no possible surprise effects from any delegated properties/functions on the `[[Prototype]]` chain, and are thus purely flat data storage.

We don't *need* classes to create meaningful relationships between two objects. The only thing we should **really care about** is objects linked together for delegation, and `Object.create(..)` gives us that linkage without all the class cruft.

## Object.create() Polyfilled

`Object.create(..)` was added in ES5. You may need to support pre-ES5 environments (like older IE's), so let's take a look at a simple **partial** polyfill for `Object.create(..)` that gives us the capability that we need even in those older JS environments:

```
if (!Object.create) {
    Object.create = function(o) {
        function F(){}
        F.prototype = o;
        return new F();
    };
}
```

This polyfill works by using a throw-away `F` function and overriding its `.prototype` property to point to the object we want to link to. Then we use `new F()` construction to make a new object that will be linked as we specified.

This usage of `Object.create(..)` is by far the most common usage, because it's the part that *can be* polyfilled. There's an additional set of functionality that the standard ES5 built-in `Object.create(..)` provides, which is **not polyfillable** for pre-ES5. As such, this capability is far-less commonly used. For completeness sake, let's look at that additional functionality:

```
var anotherObject = {
    a: 2
};

var myObject = Object.create( anotherObject, {
    b: {
        enumerable: false,
        writable: true,
        configurable: false,
        value: 3
    },
    c: {
        enumerable: true,
        writable: false,
        configurable: false,
        value: 4
    }
});
```

```

        value: 4
    }
};

myObject.hasOwnProperty( "a" ); // false
myObject.hasOwnProperty( "b" ); // true
myObject.hasOwnProperty( "c" ); // true

myObject.a; // 2
myObject.b; // 3
myObject.c; // 4

```

The second argument to `Object.create(..)` specifies property names to add to the newly created object, via declaring each new property's *property descriptor* (see Chapter 3). Because polyfilling property descriptors into pre-ES5 is not possible, this additional functionality on `Object.create(..)` also cannot be polyfilled.

The vast majority of usage of `Object.create(..)` uses the polyfill-safe subset of functionality, so most developers are fine with using the **partial polyfill** in pre-ES5 environments.

Some developers take a much stricter view, which is that no function should be polyfilled unless it can be *fully* polyfilled. Since `Object.create(..)` is one of those partial-polyfillable utilities, this narrower perspective says that if you need to use any of the functionality of `Object.create(..)` in a pre-ES5 environment, instead of polyfilling, you should use a custom utility, and stay away from using the name `Object.create` entirely. You could instead define your own utility, like:

```

function createAndLinkObject(o) {
    function F(){}
    F.prototype = o;
    return new F();
}

var anotherObject = {
    a: 2
};

var myObject = createAndLinkObject( anotherObject );

myObject.a; // 2

```

I do not share this strict opinion. I fully endorse the common partial-polyfill of `Object.create(..)` as shown above, and using it in your code even in pre-ES5. I'll leave it to you to make your own decision.

## Links As Fallbacks?

It may be tempting to think that these links between objects *primarily* provide a sort of fallback for "missing" properties or methods. While that may be an observed outcome, I don't think it represents the right way of thinking about `[[Prototype]]`.

Consider:

```

var anotherObject = {
    cool: function() {
        console.log( "cool!" );
    }
};

var myObject = Object.create( anotherObject );

myObject.cool(); // "cool!"

```

That code will work by virtue of `[[Prototype]]`, but if you wrote it that way so that `anotherObject` was acting as a fallback **just in case** `myObject` couldn't handle some property/method that some developer may try to call, odds are that your software is going to be a bit more "magical" and harder to understand and maintain.

That's not to say there aren't cases where fallbacks are an appropriate design pattern, but it's not very common or idiomatic in JS, so if you find yourself doing so, you might want to take a step back and reconsider if that's really appropriate and sensible design.

**Note:** In ES6, an advanced functionality called `Proxy` is introduced which can provide something of a "method not found" type of behavior. `Proxy` is beyond the scope of this book, but will be covered in detail in a later book in the "*You Don't Know JS*" series.

#### Don't miss an important but nuanced point here.

Designing software where you intend for a developer to, for instance, call `myObject.cool()` and have that work even though there is no `cool()` method on `myObject` introduces some "magic" into your API design that can be surprising for future developers who maintain your software.

You can however design your API with less "magic" to it, but still take advantage of the power of `[[Prototype]]` linkage.

```
var anotherObject = {
  cool: function() {
    console.log( "cool!" );
  }
};

var myObject = Object.create( anotherObject );

myObject.doCool = function() {
  this.cool(); // internal delegation!
};

myObject.doCool(); // "cool!"
```

Here, we call `myObject.doCool()`, which is a method that *actually exists* on `myObject`, making our API design more explicit (less "magical"). *Internally*, our implementation follows the **delegation design pattern** (see Chapter 6), taking advantage of `[[Prototype]]` delegation to `anotherObject.cool()`.

In other words, delegation will tend to be less surprising/confusing if it's an internal implementation detail rather than plainly exposed in your API design. We will expound on **delegation** in great detail in the next chapter.

## Review (TL;DR)

When attempting a property access on an object that doesn't have that property, the object's internal `[[Prototype]]` linkage defines where the `[[Get]]` operation (see Chapter 3) should look next. This cascading linkage from object to object essentially defines a "prototype chain" (somewhat similar to a nested scope chain) of objects to traverse for property resolution.

All normal objects have the built-in `Object.prototype` as the top of the prototype chain (like the global scope in scope look-up), where property resolution will stop if not found anywhere prior in the chain. `toString()`, `valueOf()`, and several other common utilities exist on this `Object.prototype` object, explaining how all objects in the language are able to access them.

The most common way to get two objects linked to each other is using the `new` keyword with a function call, which among its four steps (see Chapter 2), it creates a new object linked to another object.

The "another object" that the new object is linked to happens to be the object referenced by the arbitrarily named `.prototype` property of the function called with `new`. Functions called with `new` are often called "constructors", despite the fact that they are not actually instantiating a class as *constructors* do in traditional class-oriented languages.

While these JavaScript mechanisms can seem to resemble "class instantiation" and "class inheritance" from traditional class-oriented languages, the key distinction is that in JavaScript, no copies are made. Rather, objects end up linked to each other via an internal `[[Prototype]]` chain.

For a variety of reasons, not the least of which is terminology precedent, "inheritance" (and "prototypal inheritance") and all the other OO terms just do not make sense when considering how JavaScript *actually* works (not just applied to our forced mental models).

Instead, "delegation" is a more appropriate term, because these relationships are not *copies* but delegation **links**.

# You Don't Know JS Yet: Objects & Classes - 2nd Edition

## Chapter 6: Behavior Delegation

**NOTE:**

Work in progress

**NOTE:**

Everything below here is previous text from 1st edition, and is only here for reference while 2nd edition work is underway.  
**Please ignore this stuff.**

In Chapter 5, we addressed the `[[Prototype]]` mechanism in detail, and *why* it's confusing and inappropriate (despite countless attempts for nearly two decades) to describe it as "class" or "inheritance". We trudged through not only the fairly verbose syntax (`.prototype` littering the code), but the various gotchas (like surprising `.constructor` resolution or ugly pseudo-polymorphic syntax). We explored variations of the "mixin" approach, which many people use to attempt to smooth over such rough areas.

It's a common reaction at this point to wonder why it has to be so complex to do something seemingly so simple. Now that we've pulled back the curtain and seen just how dirty it all gets, it's not a surprise that most JS developers never dive this deep, and instead relegate such mess to a "class" library to handle it for them.

I hope by now you're not content to just gloss over and leave such details to a "black box" library. Let's now dig into how we *could and should* be thinking about the object `[[Prototype]]` mechanism in JS, in a **much simpler and more straightforward way** than the confusion of classes.

As a brief review of our conclusions from Chapter 5, the `[[Prototype]]` mechanism is an internal link that exists on one object which references another object.

This linkage is exercised when a property/method reference is made against the first object, and no such property/method exists. In that case, the `[[Prototype]]` linkage tells the engine to look for the property/method on the linked-to object. In turn, if that object cannot fulfill the look-up, its `[[Prototype]]` is followed, and so on. This series of links between objects forms what is called the "prototype chain".

In other words, the actual mechanism, the essence of what's important to the functionality we can leverage in JavaScript, is **all about objects being linked to other objects**.

That single observation is fundamental and critical to understanding the motivations and approaches for the rest of this chapter!

### Towards Delegation-Oriented Design

## Towards Delegation-Oriented Design

To properly focus our thoughts on how to use `[[Prototype]]` in the most straightforward way, we must recognize that it represents a fundamentally different design pattern from classes (see Chapter 4).

**Note:** Some principles of class-oriented design are still very valid, so don't toss out everything you know (just most of it!). For example, *encapsulation* is quite powerful, and is compatible (though not as common) with delegation.

We need to try to change our thinking from the class/inheritance design pattern to the behavior delegation design pattern. If you have done most or all of your programming in your education/career thinking in classes, this may be uncomfortable or feel unnatural. You may need to try this mental exercise quite a few times to get the hang of this very different way of thinking.

I'm going to walk you through some theoretical exercises first, then we'll look side-by-side at a more concrete example to give you practical context for your own code.

### Class Theory

Let's say we have several similar tasks ("XYZ", "ABC", etc) that we need to model in our software.

With classes, the way you design the scenario is: define a general parent (base) class like `Task`, defining shared behavior for all the "alike" tasks. Then, you define child classes `xyz` and `ABC`, both of which inherit from `Task`, and each of which adds specialized behavior to handle their respective tasks.

**Importantly**, the class design pattern will encourage you that to get the most out of inheritance, you will want to employ method overriding (and polymorphism), where you override the definition of some general `Task` method in your `xyz` task, perhaps even making use of `super` to call to the base version of that method while adding more behavior to it. **You'll likely find quite a few places** where you can "abstract" out general behavior to the parent class and specialize (override) it in your child classes.

Here's some loose pseudo-code for that scenario:

```
class Task {
    id;

    // constructor `Task()`
    Task(ID) { id = ID; }
    outputTask() { output( id ); }
}

class XYZ inherits Task {
    label;

    // constructor `XYZ()`
    XYZ(ID,Label) { super( ID ); label = Label; }
    outputTask() { super(); output( label ); }
}

class ABC inherits Task {
    ...
}
```

Now, you can instantiate one or more **copies** of the `xyz` child class, and use those instance(s) to perform task "XYZ". These instances have **copies both** of the general `Task` defined behavior as well as the specific `xyz` defined behavior. Likewise, instances of the `ABC` class would have copies of the `Task` behavior and the specific `ABC` behavior. After construction, you will generally only interact with these instances (and not the classes), as the instances each have copies of all the behavior you need to do the intended task.

### Delegation Theory

But now let's try to think about the same problem domain, but using *behavior delegation* instead of *classes*.

You will first define an **object** (not a class, nor a `function` as most JS's would lead you to believe) called `Task`, and it will have concrete behavior on it that includes utility methods that various tasks can use (read: *delegate to!*). Then, for each task ("XYZ", "ABC"), you define an **object** to hold that task-specific data/behavior. You **link** your task-specific object(s) to the `Task` utility object, allowing them to delegate to it when they need to.

Basically, you think about performing task "XYZ" as needing behaviors from two sibling/peer objects (`xyz` and `Task`) to accomplish it. But rather than needing to compose them together, via class copies, we can keep them in their separate objects, and we can allow `xyz` object to **delegate to** `Task` when needed.

Here's some simple code to suggest how you accomplish that:

```
var Task = {
    setID: function(ID) { this.id = ID; },
    outputID: function() { console.log( this.id ); }
};

// make `XYZ` delegate to `Task`
var XYZ = Object.create( Task );

XYZ.prepareTask = function(ID, Label) {
    this.setID( ID );
    this.label = Label;
};

XYZ.outputTaskDetails = function() {
    this.outputID();
    console.log( this.label );
};

// ABC = Object.create( Task );
// ABC ... = ...
```

In this code, `Task` and `xyz` are not classes (or functions), they're **just objects**. `xyz` is set up via `Object.create(..)` to `[[Prototype]]` delegate to the `Task` object (see Chapter 5).

As compared to class-orientation (aka, OO -- object-oriented), I call this style of code "**OLOO**" (objects-linked-to-other-objects). All we *really* care about is that the `xyz` object delegates to the `Task` object (as does the `ABC` object).

In JavaScript, the `[[Prototype]]` mechanism links **objects** to other **objects**. There are no abstract mechanisms like "classes", no matter how much you try to convince yourself otherwise. It's like paddling a canoe upstream: you *can* do it, but you're *choosing* to go against the natural current, so it's obviously **going to be harder to get where you're going**.

Some other differences to note with **OLOO style code**:

- Both `id` and `label` data members from the previous class example are data properties directly on `xyz` (neither is on `Task`). In general, with `[[Prototype]]` delegation involved, **you want state to be on the delegators** (`XYZ`, `ABC`), not on the delegate (`Task`).
- With the class design pattern, we intentionally named `outputTask` the same on both parent (`Task`) and child (`xyz`), so that we could take advantage of overriding (polymorphism). In behavior delegation, we do the opposite: **we avoid if at all possible naming things the same** at different levels of the `[[Prototype]]` chain (called shadowing -- see Chapter 5), because having those name collisions creates awkward/brittle syntax to disambiguate references (see Chapter 4), and we want to avoid that if we can.

This design pattern calls for less of general method names which are prone to overriding and instead more of descriptive method names, *specific* to the type of behavior each object is doing. **This can actually create easier to understand/maintain code**, because the names of methods (not only at definition location but strewn throughout other code) are more obvious (self documenting).

3. `this.setID(ID);` inside of a method on the `xyz` object first looks on `xyz` for `setID(..)`, but since it doesn't find a method of that name on `xyz`, `[[Prototype]] delegation` means it can follow the link to `Task` to look for `setID(..)`, which it of course finds. Moreover, because of implicit call-site `this` binding rules (see Chapter 2), when `setID(..)` runs, even though the method was found on `Task`, the `this` binding for that function call is `xyz` exactly as we'd expect and want. We see the same thing with `this.outputID()` later in the code listing.

In other words, the general utility methods that exist on `Task` are available to us while interacting with `xyz`, because `xyz` can delegate to `Task`.

**Behavior Delegation** means: let some object (`xyz`) provide a delegation (to `Task`) for property or method references if not found on the object (`xyz`).

This is an *extremely powerful* design pattern, very distinct from the idea of parent and child classes, inheritance, polymorphism, etc. Rather than organizing the objects in your mind vertically, with Parents flowing down to Children, think of objects side-by-side, as peers, with any direction of delegation links between the objects as necessary.

**Note:** Delegation is more properly used as an internal implementation detail rather than exposed directly in the API design. In the above example, we don't necessarily *intend* with our API design for developers to call `xyz.setID()` (though we can, of course!). We sorta *hide* the delegation as an internal detail of our API, where `xyz.prepareTask(..)` delegates to `Task.setID(..)`. See the "Links As Fallbacks?" discussion in Chapter 5 for more detail.

## Mutual Delegation (Disallowed)

You cannot create a *cycle* where two or more objects are mutually delegated (bi-directionally) to each other. If you make `B` linked to `A`, and then try to link `A` to `B`, you will get an error.

It's a shame (not terribly surprising, but mildly annoying) that this is disallowed. If you made a reference to a property/method which didn't exist in either place, you'd have an infinite recursion on the `[[Prototype]]` loop. But if all references were strictly present, then `B` could delegate to `A`, and vice versa, and it *could* work. This would mean you could use either object to delegate to the other, for various tasks. There are a few niche use-cases where this might be helpful.

But it's disallowed because engine implementors have observed that it's more performant to check for (and reject!) the infinite circular reference once at set-time rather than needing to have the performance hit of that guard check every time you look-up a property on an object.

## Debugged

We'll briefly cover a subtle detail that can be confusing to developers. In general, the JS specification does not control how browser developer tools should represent specific values/structures to a developer, so each browser/engine is free to interpret such things as they see fit. As such, browsers/tools *don't always agree*. Specifically, the behavior we will now examine is currently observed only in Chrome's Developer Tools.

Consider this traditional "class constructor" style JS code, as it would appear in the *console* of Chrome Developer Tools:

```
function Foo() {}

var a1 = new Foo();

a1; // Foo {}
```

Let's look at the last line of that snippet: the output of evaluating the `a1` expression, which prints `Foo {}`. If you try this same code in Firefox, you will likely see `Object {}`. Why the difference? What do these outputs mean?

Chrome is essentially saying "`{}` is an empty object that was constructed by a function with name 'Foo'". Firefox is saying "`{}` is an empty object of general construction from Object". The subtle difference is that Chrome is actively tracking, as an *internal property*, the name of the actual function that did the construction, whereas other browsers don't track that additional information.

It would be tempting to attempt to explain this with JavaScript mechanisms:

```
function Foo() {}

var a1 = new Foo();

a1.constructor; // Foo(){}
a1.constructor.name; // "Foo"
```

So, is that how Chrome is outputting "Foo", by simply examining the object's `.constructor.name`? Confusingly, the answer is both "yes" and "no".

Consider this code:

```
function Foo() {}

var a1 = new Foo();

Foo.prototype.constructor = function Gotcha(){}

a1.constructor; // Gotcha(){}
a1.constructor.name; // "Gotcha"

a1; // Foo {}
```

Even though we change `a1.constructor.name` to legitimately be something else ("Gotcha"), Chrome's console still uses the "Foo" name.

So, it would appear the answer to previous question (does it use `.constructor.name`?) is **no**, it must track it somewhere else, internally.

But, Not so fast! Let's see how this kind of behavior works with OOO-style code:

```
var Foo = {};

var a1 = Object.create( Foo );

a1; // Object {}

Object.defineProperty( Foo, "constructor", {
    enumerable: false,
    value: function Gotcha(){}
});

a1; // Gotcha {}
```

Ah-ha! **Gotcha!** Here, Chrome's console **did** find and use the `.constructor.name`. Actually, while writing this book, this exact behavior was identified as a bug in Chrome, and by the time you're reading this, it may have already been fixed. So you may instead have seen the corrected `a1; // Object {}`.

Aside from that bug, the internal tracking (apparently only for debug output purposes) of the "constructor name" that Chrome does (shown in the earlier snippets) is an intentional Chrome-only extension of behavior beyond what the JS specification calls for.

If you don't use a "constructor" to make your objects, as we've discouraged with OOO-style code here in this chapter, then you'll get objects that Chrome does *not* track an internal "constructor name" for, and such objects will correctly only be outputted as "Object {}", meaning "object generated from Object() construction".

**Don't think** this represents a drawback of OOO-style coding. When you code with OOO and behavior delegation as your design pattern, *who* "constructed" (that is, *which function* was called with `new`?) some object is an irrelevant detail. Chrome's specific internal "constructor name" tracking is really only useful if you're fully embracing "class-style" coding, but is moot if

you're instead embracing OLOO delegation.

## Mental Models Compared

Now that you can see a difference between "class" and "delegation" design patterns, at least theoretically, let's see the implications these design patterns have on the mental models we use to reason about our code.

We'll examine some more theoretical ("Foo", "Bar") code, and compare both ways (OO vs. OLOO) of implementing the code. The first snippet uses the classical ("prototypal") OO style:

```
function Foo(who) {
  this.me = who;
}
Foo.prototype.identify = function() {
  return "I am " + this.me;
};

function Bar(who) {
  Foo.call( this, who );
}
Bar.prototype = Object.create( Foo.prototype );

Bar.prototype.speak = function() {
  alert( "Hello, " + this.identify() + "." );
};

var b1 = new Bar( "b1" );
var b2 = new Bar( "b2" );

b1.speak();
b2.speak();
```

Parent class `Foo`, inherited by child class `Bar`, which is then instantiated twice as `b1` and `b2`. What we have is `b1` delegating to `Bar.prototype` which delegates to `Foo.prototype`. This should look fairly familiar to you, at this point. Nothing too ground-breaking going on.

Now, let's implement **the exact same functionality** using *OLOO* style code:

```
var Foo = {
  init: function(who) {
    this.me = who;
  },
  identify: function() {
    return "I am " + this.me;
  }
};

var Bar = Object.create( Foo );

Bar.speak = function() {
  alert( "Hello, " + this.identify() + "." );
};

var b1 = Object.create( Bar );
b1.init( "b1" );
var b2 = Object.create( Bar );
b2.init( "b2" );

b1.speak();
b2.speak();
```

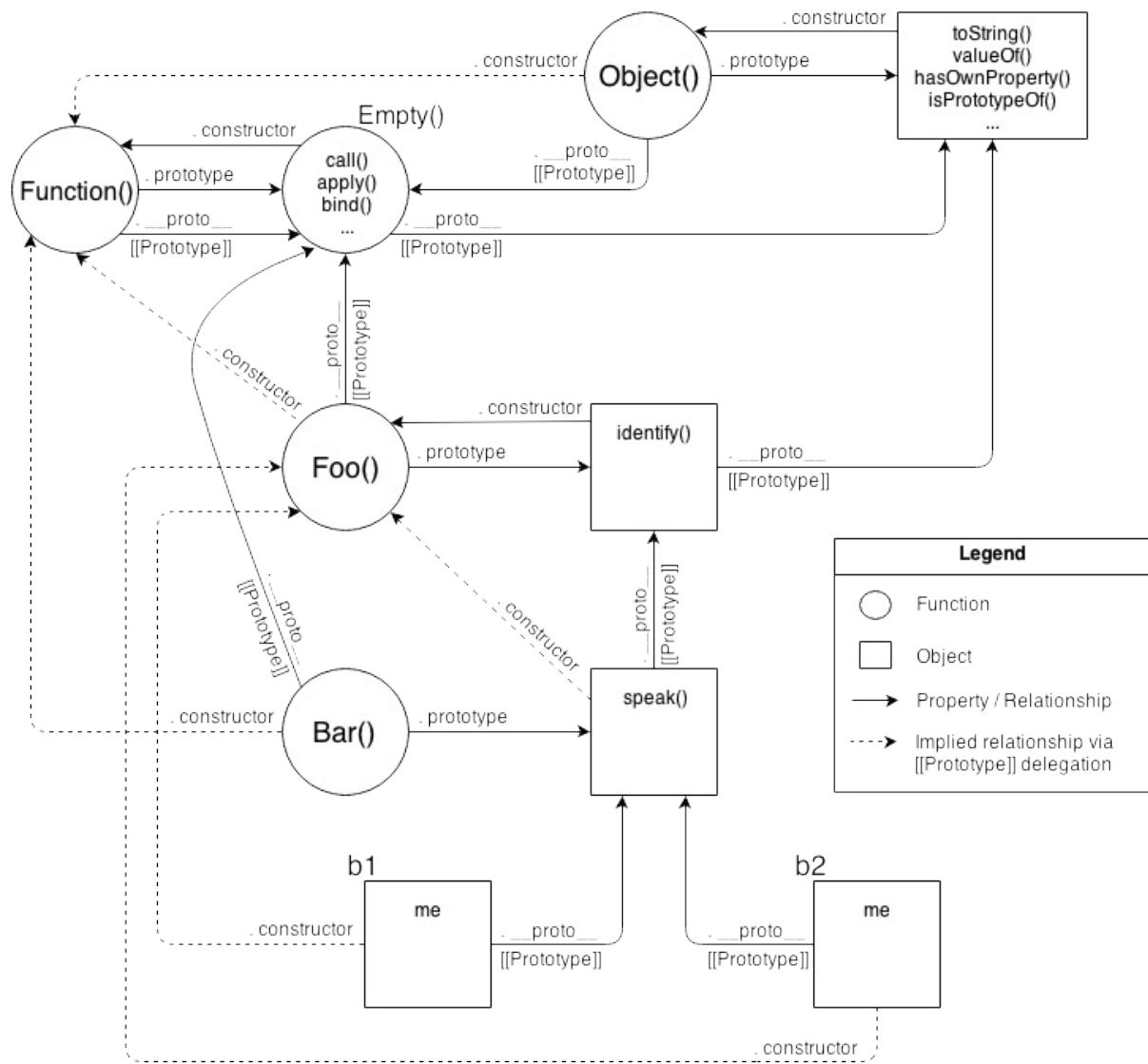
We take exactly the same advantage of `[[Prototype]]` delegation from `b1` to `Bar` to `Foo` as we did in the previous snippet between `b1`, `Bar.prototype`, and `Foo.prototype`. **We still have the same 3 objects linked together.**

But, importantly, we've greatly simplified *all the other stuff* going on, because now we just set up **objects** linked to each other, without needing all the cruft and confusion of things that look (but don't behave!) like classes, with constructors and prototypes and `new` calls.

Ask yourself: if I can get the same functionality with OOO style code as I do with "class" style code, but OOO is simpler and has less things to think about, **isn't OOO better?**

Let's examine the mental models involved between these two snippets.

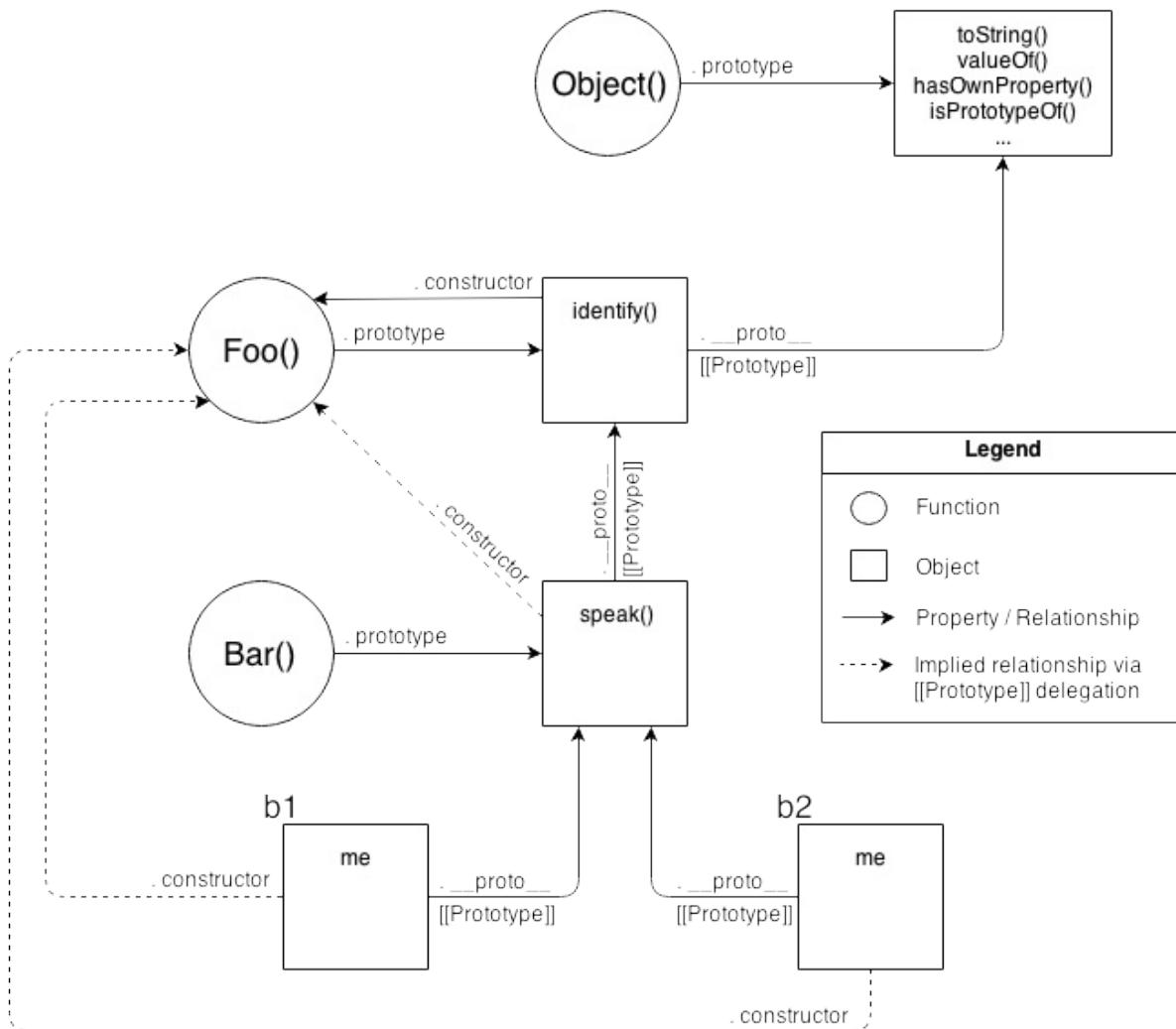
First, the class-style code snippet implies this mental model of entities and their relationships:



Actually, that's a little unfair/misleading, because it's showing a lot of extra detail that you don't *technically* need to know at all times (though you *do* need to understand it!). One take-away is that it's quite a complex series of relationships. But another take-away: if you spend the time to follow those relationship arrows around, **there's an amazing amount of internal consistency** in JS's mechanisms.

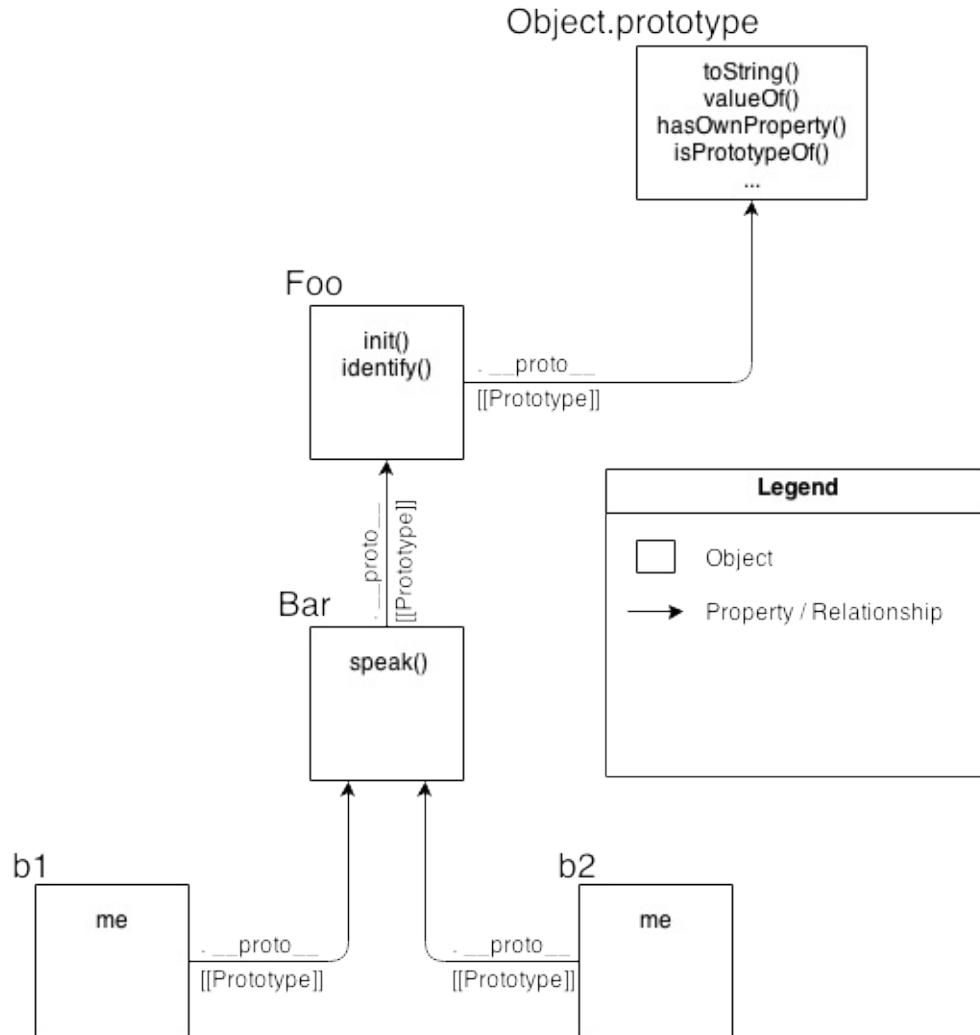
For instance, the ability of a JS function to access `call(..)`, `apply(..)`, and `bind(..)` (see Chapter 2) is because functions themselves are objects, and function-objects also have a `[[Prototype]]` linkage, to the `Function.prototype` object, which defines those default methods that any function-object can delegate to. JS can do those things, *and you can too!*

OK, let's now look at a *slightly* simplified version of that diagram which is a little more "fair" for comparison -- it shows only the *relevant* entities and relationships.



Still pretty complex, eh? The dotted lines are depicting the implied relationships when you setup the "inheritance" between `Foo.prototype` and `Bar.prototype` and haven't yet fixed the missing `.constructor` property reference (see "Constructor Redux" in Chapter 5). Even with those dotted lines removed, the mental model is still an awful lot to juggle every time you work with object linkages.

Now, let's look at the mental model for OOO-style code:



As you can see comparing them, it's quite obvious that OOO-style code has *vastly less stuff* to worry about, because OOO-style code embraces the **fact** that the only thing we ever really cared about was the **objects linked to other objects**.

All the other "class" cruft was a confusing and complex way of getting the same end result. Remove that stuff, and things get much simpler (without losing any capability).

## Classes vs. Objects

We've just seen various theoretical explorations and mental models of "classes" vs. "behavior delegation". But, let's now look at more concrete code scenarios to show how'd you actually use these ideas.

We'll first examine a typical scenario in front-end web dev: creating UI widgets (buttons, drop-downs, etc).

### Widget "Classes"

Because you're probably still so used to the OO design pattern, you'll likely immediately think of this problem domain in terms of a parent class (perhaps called `Widget`) with all the common base widget behavior, and then child derived classes for specific widget types (like `Button`).

**Note:** We're going to use jQuery here for DOM and CSS manipulation, only because it's a detail we don't really care about for the purposes of our current discussion. None of this code cares which JS framework (jQuery, Dojo, YUI, etc), if any, you might solve such mundane tasks with.

Let's examine how we'd implement the "class" design in classic-style pure JS without any "class" helper library or syntax:

```
// Parent class
function Widget(width,height) {
    this.width = width || 50;
    this.height = height || 50;
    this.$elem = null;
}

Widget.prototype.render = function($where){
    if (this.$elem) {
        this.$elem.css( {
            width: this.width + "px",
            height: this.height + "px"
        }).appendTo( $where );
    }
};

// Child class
function Button(width,height,label) {
    // "super" constructor call
    Widget.call( this, width, height );
    this.label = label || "Default";

    this.$elem = $( "<button>" ).text( this.label );
}

// make `Button` "inherit" from `Widget`
Button.prototype = Object.create( Widget.prototype );

// override base "inherited" `render(..)`
Button.prototype.render = function($where) {
    // "super" call
    Widget.prototype.render.call( this, $where );
    this.$elem.click( this.onClick.bind( this ) );
};

Button.prototype.onClick = function(evt) {
    console.log( "Button '" + this.label + "' clicked!" );
};

$( document ).ready( function(){
    var $body = $( document.body );
    var btn1 = new Button( 125, 30, "Hello" );
    var btn2 = new Button( 150, 40, "World" );

    btn1.render( $body );
    btn2.render( $body );
} );
}
```

OO design patterns tell us to declare a base `render(..)` in the parent class, then override it in our child class, but not to replace it per se, rather to augment the base functionality with button-specific behavior.

Notice the ugliness of *explicit pseudo-polymorphism* (see Chapter 4) with `Widget.call` and `Widget.prototype.render.call` references for faking "super" calls from the child "class" methods back up to the "parent" class base methods. Yuck.

## ES6 `class` sugar

We cover ES6 `class` syntax sugar in detail in Appendix A, but let's briefly demonstrate how we'd implement the same code using `class`:

```
class Widget {
    constructor(width,height) {
        this.width = width || 50;
```

```

        this.height = height || 50;
        this.$elem = null;
    }
    render($where){
        if (this.$elem) {
            this.$elem.css( {
                width: this.width + "px",
                height: this.height + "px"
            } ).appendTo( $where );
        }
    }
}

class Button extends Widget {
    constructor(width,height,label) {
        super( width, height );
        this.label = label || "Default";
        this.$elem = $( "<button>" ).text( this.label );
    }
    render($where) {
        super.render( $where );
        this.$elem.click( this.onClick.bind( this ) );
    }
    onClick(evt) {
        console.log( "Button '" + this.label + "' clicked!" );
    }
}

$( document ).ready( function(){
    var $body = $( document.body );
    var btn1 = new Button( 125, 30, "Hello" );
    var btn2 = new Button( 150, 40, "World" );

    btn1.render( $body );
    btn2.render( $body );
} );

```

Undoubtedly, a number of the syntax uglies of the previous classical approach have been smoothed over with ES6's `class`. The presence of a `super(..)` in particular seems quite nice (though when you dig into it, it's not all roses!).

Despite syntactic improvements, **these are not real classes**, as they still operate on top of the `[[Prototype]]` mechanism. They suffer from all the same mental-model mismatches we explored in Chapters 4, 5 and thus far in this chapter. Appendix A will expound on the ES6 `class` syntax and its implications in detail. We'll see why solving syntax hiccups doesn't substantially solve our class confusions in JS, though it makes a valiant effort masquerading as a solution!

Whether you use the classic prototypal syntax or the new ES6 sugar, you've still made a *choice* to model the problem domain (UI widgets) with "classes". And as the previous few chapters try to demonstrate, this *choice* in JavaScript is opting you into extra headaches and mental tax.

## Delegating Widget Objects

Here's our simpler `Widget` / `Button` example, using **OLOO style delegation**:

```

var Widget = {
    init: function(width,height){
        this.width = width || 50;
        this.height = height || 50;
        this.$elem = null;
    },
    insert: function($where){
        if (this.$elem) {
            this.$elem.css( {
                width: this.width + "px",
                height: this.height + "px"
            } ).appendTo( $where );
        }
    }
}

```

```

        } ).appendTo( $where );
    }
};

var Button = Object.create( Widget );

Button.setup = function(width,height,label){
    // delegated call
    this.init( width, height );
    this.label = label || "Default";

    this.$elem = $( "<button>" ).text( this.label );
};

Button.build = function($where) {
    // delegated call
    this.insert( $where );
    this.$elem.click( this.onClick.bind( this ) );
};

Button.onClick = function(evt) {
    console.log( "Button '" + this.label + "' clicked!" );
};

$( document ).ready( function(){
    var $body = $( document.body );

    var btn1 = Object.create( Button );
    btn1.setup( 125, 30, "Hello" );

    var btn2 = Object.create( Button );
    btn2.setup( 150, 40, "World" );

    btn1.build( $body );
    btn2.build( $body );
} );

```

With this OOO-style approach, we don't think of `Widget` as a parent and `Button` as a child. Rather, `Widget` is just an object and is sort of a utility collection that any specific type of widget might want to delegate to, and `Button` is also just a stand-alone object (with a delegation link to `Widget`, of course!).

From a design pattern perspective, we didn't share the same method name `render(..)` in both objects, the way classes suggest, but instead we chose different names (`insert(..)` and `build(..)`) that were more descriptive of what task each does specifically. The *initialization* methods are called `init(..)` and `setup(..)`, respectively, for the same reasons.

Not only does this delegation design pattern suggest different and more descriptive names (rather than shared and more generic names), but doing so with OOO happens to avoid the ugliness of the explicit pseudo-polymorphic calls (`Widget.call` and `Widget.prototype.render.call`), as you can see by the simple, relative, delegated calls to `this.init(..)` and `this.insert(..)`.

Syntactically, we also don't have any constructors, `.prototype` or `new` present, as they are, in fact, just unnecessary cruft.

Now, if you're paying close attention, you may notice that what was previously just one call (`var btn1 = new Button(..)`) is now two calls (`var btn1 = Object.create(Button)` and `btn1.setup(..)`). Initially this may seem like a drawback (more code).

However, even this is something that's a pro of OOO style code as compared to classical prototype style code. How?

With class constructors, you are "forced" (not really, but strongly suggested) to do both construction and initialization in the same step. However, there are many cases where being able to do these two steps separately (as you do with OOO!) is more flexible.

For example, let's say you create all your instances in a pool at the beginning of your program, but you wait to initialize them with specific setup until they are pulled from the pool and used. We showed the two calls happening right next to each other, but of course they can happen at very different times and in very different parts of our code, as needed.

**OLOO** supports *better* the principle of separation of concerns, where creation and initialization are not necessarily conflated into the same operation.

## Simpler Design

In addition to OLOO providing ostensibly simpler (and more flexible!) code, behavior delegation as a pattern can actually lead to simpler code architecture. Let's examine one last example that illustrates how OLOO simplifies your overall design.

The scenario we'll examine is two controller objects, one for handling the login form of a web page, and another for actually handling the authentication (communication) with the server.

We'll need a utility helper for making the Ajax communication to the server. We'll use jQuery (though any framework would do fine), since it handles not only the Ajax for us, but it returns a promise-like answer so that we can listen for the response in our calling code with `.then(...)`.

**Note:** We don't cover Promises here, but we will cover them in a future title of the "You Don't Know JS" series.

Following the typical class design pattern, we'll break up the task into base functionality in a class called `Controller`, and then we'll derive two child classes, `LoginController` and `AuthController`, which both inherit from `Controller` and specialize some of those base behaviors.

```
// Parent class
function Controller() {
    this.errors = [];
}
Controller.prototype.showDialog = function(title,msg) {
    // display title & message to user in dialog
};
Controller.prototype.success = function(msg) {
    this.showDialog( "Success", msg );
};
Controller.prototype.failure = function(err) {
    this.errors.push( err );
    this.showDialog( "Error", err );
};

// Child class
function LoginController() {
    Controller.call( this );
}
// Link child class to parent
LoginController.prototype = Object.create( Controller.prototype );
LoginController.prototype.getUser = function() {
    return document.getElementById( "login_username" ).value;
};
LoginController.prototype.getPassword = function() {
    return document.getElementById( "login_password" ).value;
};
LoginController.prototype.validateEntry = function(user,pw) {
    user = user || this.getUser();
    pw = pw || this.getPassword();

    if (!(user && pw)) {
        return this.failure( "Please enter a username & password!" );
    }
    else if (pw.length < 5) {
        return this.failure( "Password must be 5+ characters!" );
    }

    // got here? validated!
    return true;
};
```

```
// Override to extend base `failure()`
LoginController.prototype.failure = function(err) {
    // "super" call
    Controller.prototype.failure.call( this, "Login invalid: " + err );
};
```

```
// Child class
function AuthController(login) {
    Controller.call( this );
    // in addition to inheritance, we also need composition
    this.login = login;
}

// Link child class to parent
AuthController.prototype = Object.create( Controller.prototype );
AuthController.prototype.server = function(url,data) {
    return $.ajax( {
        url: url,
        data: data
    } );
};

AuthController.prototype.checkAuth = function() {
    var user = this.login.getUser();
    var pw = this.login.getPassword();

    if (this.login.validateEntry( user, pw )) {
        this.server( "/check-auth",{
            user: user,
            pw: pw
        } )
        .then( this.success.bind( this ) )
        .fail( this.failure.bind( this ) );
    }
};

// Override to extend base `success()`
AuthController.prototype.success = function() {
    // "super" call
    Controller.prototype.success.call( this, "Authenticated!" );
};

// Override to extend base `failure()`
AuthController.prototype.failure = function(err) {
    // "super" call
    Controller.prototype.failure.call( this, "Auth Failed: " + err );
};
```

```
var auth = new AuthController(
    // in addition to inheritance, we also need composition
    new LoginController()
);
auth.checkAuth();
```

We have base behaviors that all controllers share, which are `success(..)`, `failure(..)` and `showDialog(..)`. Our child classes `LoginController` and `AuthController` override `failure(..)` and `success(..)` to augment the default base class behavior. Also note that `AuthController` needs an instance of `LoginController` to interact with the login form, so that becomes a member data property.

The other thing to mention is that we chose some *composition* to sprinkle in on top of the inheritance. `AuthController` needs to know about `LoginController`, so we instantiate it (`new LoginController()`) and keep a class member property called `this.login` to reference it, so that `AuthController` can invoke behavior on `LoginController`.

**Note:** There *might* have been a slight temptation to make `AuthController` inherit from `LoginController`, or vice versa, such that we had *virtual composition* through the inheritance chain. But this is a strongly clear example of what's wrong with class inheritance as *the* model for the problem domain, because neither `AuthController` nor `LoginController` are specializing base

behavior of the other, so inheritance between them makes little sense except if classes are your only design pattern. Instead, we layered in some simple *composition* and now they can cooperate, while still both benefiting from the inheritance from the parent base `Controller`.

If you're familiar with class-oriented (OO) design, this should all look pretty familiar and natural.

## De-class-ified

But, **do we really need to model this problem** with a parent `controller` class, two child classes, **and some composition?** Is there a way to take advantage of OLOO-style behavior delegation and have a *much* simpler design? **Yes!**

```
var LoginController = {
  errors: [],
  getUser: function() {
    return document.getElementById( "login_username" ).value;
  },
  getPassword: function() {
    return document.getElementById( "login_password" ).value;
  },
  validateEntry: function(user,pw) {
    user = user || this.getUser();
    pw = pw || this.getPassword();

    if (!(user && pw)) {
      return this.failure( "Please enter a username & password!" );
    }
    else if (pw.length < 5) {
      return this.failure( "Password must be 5+ characters!" );
    }

    // got here? validated!
    return true;
  },
  showDialog: function(title,msg) {
    // display success message to user in dialog
  },
  failure: function(err) {
    this.errors.push( err );
    this.showDialog( "Error", "Login invalid: " + err );
  }
};
```

```
// Link `AuthController` to delegate to `LoginController`
var AuthController = Object.create( LoginController );

AuthController.errors = [];
AuthController.checkAuth = function() {
  var user = this.getUser();
  var pw = this.getPassword();

  if (this.validateEntry( user, pw )) {
    this.server( "/check-auth",{
      user: user,
      pw: pw
    } )
    .then( this.accepted.bind( this ) )
    .fail( this.rejected.bind( this ) );
  }
};

AuthController.server = function(url,data) {
  return $.ajax( {
    url: url,
    data: data
  } );
};
```

```

};

AuthController.accepted = function() {
  this.showDialog( "Success", "Authenticated!" )
};

AuthController.rejected = function(err) {
  this.failure( "Auth Failed: " + err );
};

```

Since `AuthController` is just an object (so is `LoginController`), we don't need to instantiate (like `new AuthController()`) to perform our task. All we need to do is:

```
AuthController.checkAuth();
```

Of course, with OOOO, if you do need to create one or more additional objects in the delegation chain, that's easy, and still doesn't require anything like class instantiation:

```

var controller1 = Object.create( AuthController );
var controller2 = Object.create( AuthController );

```

With behavior delegation, `AuthController` and `LoginController` are **just objects**, *horizontal* peers of each other, and are not arranged or related as parents and children in class-orientation. We somewhat arbitrarily chose to have `AuthController` delegate to `LoginController` -- it would have been just as valid for the delegation to go the reverse direction.

The main takeaway from this second code listing is that we only have two entities (`LoginController` and `AuthController`), **not three** as before.

We didn't need a base `controller` class to "share" behavior between the two, because delegation is a powerful enough mechanism to give us the functionality we need. We also, as noted before, don't need to instantiate our classes to work with them, because there are no classes, **just the objects themselves**. Furthermore, there's no need for *composition* as delegation gives the two objects the ability to cooperate *differentially* as needed.

Lastly, we avoided the polymorphism pitfalls of class-oriented design by not having the names `success(..)` and `failure(..)` be the same on both objects, which would have required ugly explicit pseudopolymorphism. Instead, we called them `accepted()` and `rejected()` on `AuthController` -- slightly more descriptive names for their specific tasks.

**Bottom line:** we end up with the same capability, but a (significantly) simpler design. That's the power of OOOO-style code and the power of the *behavior delegation* design pattern.

## Nicer Syntax

One of the nicer things that makes ES6's `class` so deceptively attractive (see Appendix A on why to avoid it!) is the short-hand syntax for declaring class methods:

```

class Foo {
  methodName() { /* ... */ }
}

```

We get to drop the word `function` from the declaration, which makes JS developers everywhere cheer!

And you may have noticed and been frustrated that the suggested OOOO syntax above has lots of `function` appearances, which seems like a bit of a detractor to the goal of OOOO simplification. **But it doesn't have to be that way!**

As of ES6, we can use *concise method declarations* in any object literal, so an object in OOOO style can be declared this way (same short-hand sugar as with `class` body syntax):

```
var LoginController = {
  errors: [],
  getUser() { // Look ma, no `function`!
    // ...
  },
  getPassword() {
    // ...
  }
  // ...
};
```

About the only difference is that object literals will still require `,` comma separators between elements whereas `class` syntax doesn't. Pretty minor concession in the whole scheme of things.

Moreover, as of ES6, the clunkier syntax you use (like for the `AuthController` definition), where you're assigning properties individually and not using an object literal, can be re-written using an object literal (so that you can use concise methods), and you can just modify that object's `[[Prototype]]` with `Object.setPrototypeOf(..)`, like this:

```
// use nicer object literal syntax w/ concise methods!
var AuthController = {
  errors: [],
  checkAuth() {
    // ...
  },
  server(url,data) {
    // ...
  }
  // ...
};

// NOW, link `AuthController` to delegate to `LoginController`
Object.setPrototypeOf( AuthController, LoginController );
```

OLOO-style as of ES6, with concise methods, **is a lot friendlier** than it was before (and even then, it was much simpler and nicer than classical prototype-style code). **You don't have to opt for class** (complexity) to get nice clean object syntax!

## Unlexical

There *is* one drawback to concise methods that's subtle but important to note. Consider this code:

```
var Foo = {
  bar() { /*...*/ },
  baz: function baz() { /*...*/ }
};
```

Here's the syntactic de-sugaring that expresses how that code will operate:

```
var Foo = {
  bar: function() { /*...*/ },
  baz: function baz() { /*...*/ }
};
```

See the difference? The `bar()` short-hand became an *anonymous function expression* (`function()...`) attached to the `bar` property, because the function object itself has no name identifier. Compare that to the manually specified *named function expression* (`function baz()...`) which has a lexical name identifier `baz` in addition to being attached to a `.baz` property.

So what? In the "Scope & Closures" title of this "You Don't Know JS" book series, we cover the three main downsides of *anonymous function expressions* in detail. We'll just briefly repeat them so we can compare to the concise method short-hand.

Lack of a `name` identifier on an anonymous function:

1. makes debugging stack traces harder
2. makes self-referencing (recursion, event (un)binding, etc) harder
3. makes code (a little bit) harder to understand

Items 1 and 3 don't apply to concise methods.

Even though the de-sugaring uses an *anonymous function expression* which normally would have no `name` in stack traces, concise methods are specified to set the internal `name` property of the function object accordingly, so stack traces should be able to use it (though that's implementation dependent so not guaranteed).

Item 2 is, unfortunately, **still a drawback to concise methods**. They will not have a lexical identifier to use as a self-reference.

Consider:

```
var Foo = {
  bar: function(x) {
    if (x < 10) {
      return Foo.bar( x * 2 );
    }
    return x;
  },
  baz: function baz(x) {
    if (x < 10) {
      return baz( x * 2 );
    }
    return x;
  }
};
```

The manual `Foo.bar(x*2)` reference above kind of suffices in this example, but there are many cases where a function wouldn't necessarily be able to do that, such as cases where the function is being shared in delegation across different objects, using `this` binding, etc. You would want to use a real self-reference, and the function object's `name` identifier is the best way to accomplish that.

Just be aware of this caveat for concise methods, and if you run into such issues with lack of self-reference, make sure to forgo the concise method syntax **just for that declaration** in favor of the manual *named function expression* declaration form: `baz: function baz(){...}`.

## Introspection

If you've spent much time with class oriented programming (either in JS or other languages), you're probably familiar with *type introspection*: inspecting an instance to find out what *kind* of object it is. The primary goal of *type introspection* with class instances is to reason about the structure/capabilities of the object based on *how it was created*.

Consider this code which uses `instanceof` (see Chapter 5) for introspecting on an object `a1` to infer its capability:

```
function Foo() {
  // ...
}
Foo.prototype.something = function(){
  // ...
}

var a1 = new Foo();

// later

if (a1 instanceof Foo) {
  a1.something();
```

```
}
```

Because `Foo.prototype` (not `Foo`!) is in the `[[Prototype]]` chain (see Chapter 5) of `a1`, the `instanceof` operator (confusingly) pretends to tell us that `a1` is an instance of the `Foo` "class". With this knowledge, we then assume that `a1` has the capabilities described by the `Foo` "class".

Of course, there is no `Foo` class, only a plain old normal function `Foo`, which happens to have a reference to an arbitrary object (`Foo.prototype`) that `a1` happens to be delegation-linked to. By its syntax, `instanceof` pretends to be inspecting the relationship between `a1` and `Foo`, but it's actually telling us whether `a1` and (the arbitrary object referenced by) `Foo.prototype` are related.

The semantic confusion (and indirection) of `instanceof` syntax means that to use `instanceof`-based introspection to ask if object `a1` is related to the capabilities object in question, you *have to* have a function that holds a reference to that object -- you can't just directly ask if the two objects are related.

Recall the abstract `Foo` / `Bar` / `b1` example from earlier in this chapter, which we'll abbreviate here:

```
function Foo() { /* .. */ }
Foo.prototype...

function Bar() { /* .. */ }
Bar.prototype = Object.create( Foo.prototype );

var b1 = new Bar( "b1" );
```

For *type introspection* purposes on the entities in that example, using `instanceof` and `.prototype` semantics, here are the various checks you might need to perform:

```
// relating `Foo` and `Bar` to each other
Bar.prototype instanceof Foo; // true
Object.getPrototypeOf( Bar.prototype ) === Foo.prototype; // true
Foo.prototype.isPrototypeOf( Bar.prototype ); // true

// relating `b1` to both `Foo` and `Bar`
b1 instanceof Foo; // true
b1 instanceof Bar; // true
Object.getPrototypeOf( b1 ) === Bar.prototype; // true
Foo.prototype.isPrototypeOf( b1 ); // true
Bar.prototype.isPrototypeOf( b1 ); // true
```

It's fair to say that some of that kinda sucks. For instance, intuitively (with classes) you might want to be able to say something like `Bar instanceof Foo` (because it's easy to mix up what "instance" means to think it includes "inheritance"), but that's not a sensible comparison in JS. You have to do `Bar.prototype instanceof Foo` instead.

Another common, but perhaps less robust, pattern for *type introspection*, which many devs seem to prefer over `instanceof`, is called "duck typing". This term comes from the adage, "if it looks like a duck, and it quacks like a duck, it must be a duck".

Example:

```
if (a1.something) {
  a1.something();
}
```

Rather than inspecting for a relationship between `a1` and an object that holds the delegatable `something()` function, we assume that the test for `a1.something` passing means `a1` has the capability to call `.something()` (regardless of if it found the method directly on `a1` or delegated to some other object). In and of itself, that assumption isn't so risky.

But "duck typing" is often extended to make **other assumptions about the object's capabilities** besides what's being tested, which of course introduces more risk (aka, brittle design) into the test.

One notable example of "duck typing" comes with ES6 Promises (which as an earlier note explained are not being covered in this book).

For various reasons, there's a need to determine if any arbitrary object reference *is a Promise*, but the way that test is done is to check if the object happens to have a `then()` function present on it. In other words, **if any object** happens to have a `then()` method, ES6 Promises will assume unconditionally that the object **is a "thenable"** and therefore will expect it to behave conformantly to all standard behaviors of Promises.

If you have any non-Promise object that happens for whatever reason to have a `then()` method on it, you are strongly advised to keep it far away from the ES6 Promise mechanism to avoid broken assumptions.

That example clearly illustrates the perils of "duck typing". You should only use such approaches sparingly and in controlled conditions.

Turning our attention once again back to OOO-style code as presented here in this chapter, *type introspection* turns out to be much cleaner. Let's recall (and abbreviate) the `Foo` / `Bar` / `b1` OOO example from earlier in the chapter:

```
var Foo = { /* ... */ };

var Bar = Object.create( Foo );
Bar...

var b1 = Object.create( Bar );
```

Using this OOO approach, where all we have are plain objects that are related via `[[Prototype]]` delegation, here's the quite simplified *type introspection* we might use:

```
// relating `Foo` and `Bar` to each other
Foo.isPrototypeOf( Bar ); // true
Object.getPrototypeOf( Bar ) === Foo; // true

// relating `b1` to both `Foo` and `Bar`
Foo.isPrototypeOf( b1 ); // true
Bar.isPrototypeOf( b1 ); // true
Object.getPrototypeOf( b1 ) === Bar; // true
```

We're not using `instanceof` anymore, because it's confusingly pretending to have something to do with classes. Now, we just ask the (informally stated) question, "are you *a prototype of me?*" There's no more indirection necessary with stuff like `Foo.prototype` or the painfully verbose `Foo.prototype.isPrototypeOf(..)`.

I think it's fair to say these checks are significantly less complicated/confusing than the previous set of introspection checks. **Yet again, we see that OOO is simpler than (but with all the same power of) class-style coding in JavaScript.**

## Review (TL;DR)

Classes and inheritance are a design pattern you can *choose*, or *not choose*, in your software architecture. Most developers take for granted that classes are the only (proper) way to organize code, but here we've seen there's another less-commonly talked about pattern that's actually quite powerful: **behavior delegation**.

Behavior delegation suggests objects as peers of each other, which delegate amongst themselves, rather than parent and child class relationships. JavaScript's `[[Prototype]]` mechanism is, by its very designed nature, a behavior delegation mechanism. That means we can either choose to struggle to implement class mechanics on top of JS (see Chapters 4 and 5), or we can just embrace the natural state of `[[Prototype]]` as a delegation mechanism.

When you design code with objects only, not only does it simplify the syntax you use, but it can actually lead to simpler code architecture design.

**OLOO** (objects-linked-to-other-objects) is a code style which creates and relates objects directly without the abstraction of classes. OLOO quite naturally implements `[[Prototype]]`-based behavior delegation.

# You Don't Know JS Yet: Objects & Classes - 2nd Edition

## Appendix A: ES6 `class`

**NOTE:**

Work in progress

**NOTE:**

Everything below here is previous text from 1st edition, and is only here for reference while 2nd edition work is underway.  
**Please ignore this stuff.**

If there's any take-away message from the second half of this book (Chapters 4-6), it's that classes are an optional design pattern for code (not a necessary given), and that furthermore they are often quite awkward to implement in a `[[Prototype]]` language like JavaScript.

This awkwardness is *not* just about syntax, although that's a big part of it. Chapters 4 and 5 examined quite a bit of syntactic ugliness, from verbosity of `.prototype` references cluttering the code, to *explicit pseudo-polymorphism* (see Chapter 4) when you give methods the same name at different levels of the chain and try to implement a polymorphic reference from a lower-level method to a higher-level method. `.constructor` being wrongly interpreted as "was constructed by" and yet being unreliable for that definition is yet another syntactic ugly.

But the problems with class design are much deeper. Chapter 4 points out that classes in traditional class-oriented languages actually produce a *copy* action from parent to child to instance, whereas in `[[Prototype]]`, the action is **not** a copy, but rather the opposite -- a delegation link.

When compared to the simplicity of OLOO-style code and behavior delegation (see Chapter 6), which embrace `[[Prototype]]` rather than hide from it, classes stand out as a sore thumb in JS.

### `class`

But we *don't* need to re-argue that case again. I re-mention those issues briefly only so that you keep them fresh in your mind now that we turn our attention to the ES6 `class` mechanism. We'll demonstrate here how it works, and look at whether or not `class` does anything substantial to address any of those "class" concerns.

Let's revisit the `Widget` / `Button` example from Chapter 6:

```
class Widget {
```

```

constructor(width,height) {
    this.width = width || 50;
    this.height = height || 50;
    this.$elem = null;
}
render($where){
    if (this.$elem) {
        this.$elem.css( {
            width: this.width + "px",
            height: this.height + "px"
        }).appendTo( $where );
    }
}

class Button extends Widget {
    constructor(width,height,label) {
        super( width, height );
        this.label = label || "Default";
        this.$elem = $( "<button>" ).text( this.label );
    }
    render($where) {
        super.render( $where );
        this.$elem.click( this.onClick.bind( this ) );
    }
    onClick(evt) {
        console.log( "Button '" + this.label + "' clicked!" );
    }
}

```

Beyond this syntax *looking* nicer, what problems does ES6 solve?

1. There's no more (well, sorta, see below!) references to `.prototype` cluttering the code.
2. `Button` is declared directly to "inherit from" (aka `extends`) `Widget`, instead of needing to use `Object.create(..)` to replace a `.prototype` object that's linked, or having to set with `.__proto__` or `Object.setPrototypeOf(..)`.
3. `super(..)` now gives us a very helpful **relative polymorphism** capability, so that any method at one level of the chain can refer relatively one level up the chain to a method of the same name. This includes a solution to the note from Chapter 4 about the weirdness of constructors not belonging to their class, and so being unrelated -- `super()` works inside constructors exactly as you'd expect.
4. `class` literal syntax has no affordance for specifying properties (only methods). This might seem limiting to some, but it's expected that the vast majority of cases where a property (state) exists elsewhere but the end-chain "instances", this is usually a mistake and surprising (as it's state that's implicitly "shared" among all "instances"). So, one *could* say the `class` syntax is protecting you from mistakes.
5. `extends` lets you extend even built-in object (sub)types, like `Array` or `RegExp`, in a very natural way. Doing so without `class .. extends` has long been an exceedingly complex and frustrating task, one that only the most adept of framework authors have ever been able to accurately tackle. Now, it will be rather trivial!

In all fairness, those are some substantial solutions to many of the most obvious (syntactic) issues and surprises people have with classical prototype-style code.

## class Gotchas

It's not all bubblegum and roses, though. There are still some deep and profoundly troubling issues with using "classes" as a design pattern in JS.

Firstly, the `class` syntax may convince you a new "class" mechanism exists in JS as of ES6. **Not so.** `class` is, mostly, just syntactic sugar on top of the existing `[[Prototype]]` (delegation!) mechanism.

That means `class` is not actually copying definitions statically at declaration time the way it does in traditional class-oriented languages. If you change/replace a method (on purpose or by accident) on the parent "class", the child "class" and/or instances will still be "affected", in that they didn't get copies at declaration time, they are all still using the live-delegation model based on

`[[Prototype]]` :

```
class C {
  constructor() {
    this.num = Math.random();
  }
  rand() {
    console.log( "Random: " + this.num );
  }
}

var c1 = new C();
c1.rand(); // "Random: 0.4324299..."

C.prototype.rand = function() {
  console.log( "Random: " + Math.round( this.num * 1000 ) );
};

var c2 = new C();
c2.rand(); // "Random: 867"

c1.rand(); // "Random: 432" -- oops!!!

```

This only seems like reasonable behavior *if you already know* about the delegation nature of things, rather than expecting *copies* from "real classes". So the question to ask yourself is, why are you choosing `class` syntax for something fundamentally different from classes?

Doesn't the ES6 `class` syntax **just make it harder** to see and understand the difference between traditional classes and delegated objects?

`class` syntax *does not* provide a way to declare class member properties (only methods). So if you need to do that to track shared state among instances, then you end up going back to the ugly `.prototype` syntax, like this:

```
class C {
  constructor() {
    // make sure to modify the shared state,
    // not set a shadowed property on the
    // instances!
    C.prototype.count++;

    // here, `this.count` works as expected
    // via delegation
    console.log( "Hello: " + this.count );
  }
}

// add a property for shared state directly to
// prototype object
C.prototype.count = 0;

var c1 = new C();
// Hello: 1

var c2 = new C();
// Hello: 2

c1.count === 2; // true
c1.count === c2.count; // true

```

The biggest problem here is that it betrays the `class` syntax by exposing (leakage!) `.prototype` as an implementation detail.

But, we also still have the surprise gotcha that `this.count++` would implicitly create a separate shadowed `.count` property on both `c1` and `c2` objects, rather than updating the shared state. `class` offers us no consolation from that issue, except (presumably) to imply by lack of syntactic support that you shouldn't be doing that *at all*.

Moreover, accidental shadowing is still a hazard:

```
class C {
  constructor(id) {
    // oops, gotcha, we're shadowing `id()` method
    // with a property value on the instance
    this.id = id;
  }
  id() {
    console.log("Id: " + this.id);
  }
}

var c1 = new C("c1");
c1.id(); // TypeError -- `c1.id` is now the string "c1"
```

There's also some very subtle nuanced issues with how `super` works. You might assume that `super` would be bound in an analogous way to how `this` gets bound (see Chapter 2), which is that `super` would always be bound to one level higher than whatever the current method's position in the `[[Prototype]]` chain is.

However, for performance reasons (`this` binding is already expensive), `super` is not bound dynamically. It's bound sort of "statically", as declaration time. No big deal, right?

Ehh... maybe, maybe not. If you, like most JS devs, start assigning functions around to different objects (which came from `class` definitions), in various different ways, you probably won't be very aware that in all those cases, the `super` mechanism under the covers is having to be re-bound each time.

And depending on what sorts of syntactic approaches you take to these assignments, there may very well be cases where the `super` can't be properly bound (at least, not where you suspect), so you may (at time of writing, TC39 discussion is ongoing on the topic) have to manually bind `super` with `toMethod(..)` (kinda like you have to do `bind(..)` for `this` -- see Chapter 2).

You're used to being able to assign around methods to different objects to *automatically* take advantage of the dynamism of `this` via the *implicit binding* rule (see Chapter 2). But the same will likely not be true with methods that use `super`.

Consider what `super` should do here (against `D` and `E`):

```
class P {
  foo() { console.log("P.foo"); }
}

class C extends P {
  foo() {
    super();
  }
}

var c1 = new C();
c1.foo(); // "P.foo"

var D = {
  foo: function() { console.log("D.foo"); }
};

var E = {
  foo: C.prototype.foo
};

// Link E to D for delegation
Object.setPrototypeOf(E, D);
```

```
E.foo(); // "P.foo"
```

If you were thinking (quite reasonably!) that `super` would be bound dynamically at call-time, you might expect that `super()` would automatically recognize that `E` delegates to `D`, so `E.foo()` using `super()` should call to `D.foo()`.

**Not so.** For performance pragmatism reasons, `super` is not *late bound* (aka, dynamically bound) like `this` is. Instead it's derived at call-time from `[[HomeObject]].[[Prototype]]`, where `[[HomeObject]]` is statically bound at creation time.

In this particular case, `super()` is still resolving to `P.foo()`, since the method's `[[HomeObject]]` is still `C` and `C.  
[[Prototype]]` is `P`.

There will *probably* be ways to manually address such gotchas. Using `toMethod(...)` to bind/rebind a method's `[[HomeObject]]` (along with setting the `[[Prototype]]` of that object!) appears to work in this scenario:

```
var D = {
  foo: function() { console.log("D.foo"); }
};

// Link E to D for delegation
var E = Object.create(D);

// manually bind `foo`'s `[[HomeObject]]` as
// `E`, and `E.[[Prototype]]` is `D`, so thus
// `super()` is `D.foo()`
E.foo = C.prototype.foo.toMethod(E, "foo");

E.foo(); // "D.foo"
```

**Note:** `toMethod(...)` clones the method, and takes `homeObject` as its first parameter (which is why we pass `E`), and the second parameter (optionally) sets a `name` for the new method (which keep at "foo").

It remains to be seen if there are other corner case gotchas that devs will run into beyond this scenario. Regardless, you will have to be diligent and stay aware of which places the engine automatically figures out `super` for you, and which places you have to manually take care of it. **Ugh!**

## Static > Dynamic?

But the biggest problem of all about ES6 `class` is that all these various gotchas mean `class` sorta opts you into a syntax which seems to imply (like traditional classes) that once you declare a `class`, it's a static definition of a (future instantiated) thing. You completely lose sight of the fact that `C` is an object, a concrete thing, which you can directly interact with.

In traditional class-oriented languages, you never adjust the definition of a class later, so the class design pattern doesn't suggest such capabilities. But **one of the most powerful parts** of JS is that it *is* dynamic, and the definition of any object is (unless you make it immutable) a fluid and mutable *thing*.

`class` seems to imply you shouldn't do such things, by forcing you into the uglier `.prototype` syntax to do so, or forcing you to think about `super` gotchas, etc. It also offers *very little* support for any of the pitfalls that this dynamism can bring.

In other words, it's as if `class` is telling you: "dynamic is too hard, so it's probably not a good idea. Here's a static-looking syntax, so code your stuff statically."

What a sad commentary on JavaScript: **dynamic is too hard, let's pretend to be (but not actually be!) static.**

These are the reasons why ES6 `class` is masquerading as a nice solution to syntactic headaches, but it's actually muddying the waters further and making things worse for JS and for clear and concise understanding.

**Note:** If you use the `.bind(..)` utility to make a hard-bound function (see Chapter 2), the function created is not subclassable with ES6 `extend` like normal functions are.

## Review (TL;DR)

`class` does a very good job of pretending to fix the problems with the class/inheritance design pattern in JS. But it actually does the opposite: **it hides many of the problems, and introduces other subtle but dangerous ones.**

`class` contributes to the ongoing confusion of "class" in JavaScript which has plagued the language for nearly two decades. In some respects, it asks more questions than it answers, and it feels in totality like a very unnatural fit on top of the elegant simplicity of the `[[Prototype]]` mechanism.

Bottom line: if ES6 `class` makes it harder to robustly leverage `[[Prototype]]`, and hides the most important nature of the JS object mechanism -- **the live delegation links between objects** -- shouldn't we see `class` as creating more troubles than it solves, and just relegate it to an anti-pattern?

I can't really answer that question for you. But I hope this book has fully explored the issue at a deeper level than you've ever gone before, and has given you the information you need to *answer it yourself*.

# You Don't Know JS Yet: Types & Grammar - 2nd Edition

<b>NOTE:</b>
Work in progress

## Table of Contents

- [Foreword \(by TBA\)](#)
- [Preface](#)
- [Chapter 1: Types](#)
- [Chapter 2: Values](#)
- [Chapter 3: Natives](#)
- [Chapter 4: Coercion](#)
- [Chapter 5: Grammar](#)
- [Appendix A: Mixed Environment JavaScript](#)

# You Don't Know JS Yet: Types & Grammar - 2nd Edition

## Chapter 1: Types

<b>NOTE:</b>
Work in progress

<b>NOTE:</b>
Everything below here is previous text from 1st edition, and is only here for reference while 2nd edition work is underway. <b>Please ignore this stuff.</b>

Everything below here is previous text from 1st edition, and is only here for reference while 2nd edition work is underway.  
**Please ignore this stuff.**

Most developers would say that a dynamic language (like JS) does not have *types*. Let's see what the ES5.1 specification (<http://www.ecma-international.org/ecma-262/5.1/>) has to say on the topic:

Algorithms within this specification manipulate values each of which has an associated type. The possible value types are exactly those defined in this clause. Types are further sub classified into ECMAScript language types and specification types.

An ECMAScript language type corresponds to values that are directly manipulated by an ECMAScript programmer using the ECMAScript language. The ECMAScript language types are Undefined, Null, Boolean, String, Number, and Object.

Now, if you're a fan of strongly typed (statically typed) languages, you may object to this usage of the word "type." In those languages, "type" means a whole lot *more* than it does here in JS.

Some people say JS shouldn't claim to have "types," and they should instead be called "tags" or perhaps "subtypes".

Bah! We're going to use this rough definition (the same one that seems to drive the wording of the spec): a *type* is an intrinsic, built-in set of characteristics that uniquely identifies the behavior of a particular value and distinguishes it from other values, both to the engine **and to the developer**.

In other words, if both the engine and the developer treat value `42` (the number) differently than they treat value `"42"` (the string), then those two values have different *types* -- `number` and `string`, respectively. When you use `42`, you are *intending* to do something numeric, like math. But when you use `"42"`, you are *intending* to do something string-ish, like outputting to the page, etc. **These two values have different types.**

That's by no means a perfect definition. But it's good enough for this discussion. And it's consistent with how JS describes itself.

## A Type By Any Other Name...

# A Type By Any Other Name...

Beyond academic definition disagreements, why does it matter if JavaScript has *types* or not?

Having a proper understanding of each *type* and its intrinsic behavior is absolutely essential to understanding how to properly and accurately convert values to different types (see Coercion, Chapter 4). Nearly every JS program ever written will need to handle value coercion in some shape or form, so it's important you do so responsibly and with confidence.

If you have the `number` value `42`, but you want to treat it like a `string`, such as pulling out the `"2"` as a character in position `1`, you obviously must first convert (coerce) the value from `number` to `string`.

That seems simple enough.

But there are many different ways that such coercion can happen. Some of these ways are explicit, easy to reason about, and reliable. But if you're not careful, coercion can happen in very strange and surprising ways.

Coercion confusion is perhaps one of the most profound frustrations for JavaScript developers. It has often been criticized as being so *dangerous* as to be considered a flaw in the design of the language, to be shunned and avoided.

Armed with a full understanding of JavaScript types, we're aiming to illustrate why coercion's *bad reputation* is largely overhyped and somewhat undeserved -- to flip your perspective, to seeing coercion's power and usefulness. But first, we have to get a much better grip on values and types.

## Built-in Types

JavaScript defines seven built-in types:

- `null`
- `undefined`
- `boolean`
- `number`
- `string`
- `object`
- `symbol` -- added in ES6!

**Note:** All of these types except `object` are called "primitives".

The `typeof` operator inspects the type of the given value, and always returns one of seven string values -- surprisingly, there's not an exact 1-to-1 match with the seven built-in types we just listed.

```
typeof undefined === "undefined"; // true
typeof true === "boolean"; // true
typeof 42 === "number"; // true
typeof "42" === "string"; // true
typeof { life: 42 } === "object"; // true

// added in ES6!
typeof Symbol() === "symbol"; // true
```

These six listed types have values of the corresponding type and return a string value of the same name, as shown. `Symbol` is a new data type as of ES6, and will be covered in Chapter 3.

As you may have noticed, I excluded `null` from the above listing. It's *special* -- special in the sense that it's buggy when combined with the `typeof` operator:

```
typeof null === "object"; // true
```

It would have been nice (and correct!) if it returned `"null"`, but this original bug in JS has persisted for nearly two decades, and will likely never be fixed because there's too much existing web content that relies on its buggy behavior that "fixing" the bug would *create* more "bugs" and break a lot of web software.

If you want to test for a `null` value using its type, you need a compound condition:

```
var a = null;
(!a && typeof a === "object"); // true
```

`null` is the only primitive value that is "falsy" (aka false-like; see Chapter 4) but that also returns `"object"` from the `typeof` check.

So what's the seventh string value that `typeof` can return?

```
typeof function a(){ /* ... */ } === "function"; // true
```

It's easy to think that `function` would be a top-level built-in type in JS, especially given this behavior of the `typeof` operator. However, if you read the spec, you'll see it's actually a "subtype" of object. Specifically, a function is referred to as a "callable object" -- an object that has an internal `[[Call]]` property that allows it to be invoked.

The fact that functions are actually objects is quite useful. Most importantly, they can have properties. For example:

```
function a(b,c) {
  /* ... */
}
```

The function object has a `length` property set to the number of formal parameters it is declared with.

```
a.length; // 2
```

Since you declared the function with two formal named parameters (`b` and `c`), the "length of the function" is `2`.

What about arrays? They're native to JS, so are they a special type?

```
typeof [1,2,3] === "object"; // true
```

Nope, just objects. It's most appropriate to think of them also as a "subtype" of object (see Chapter 3), in this case with the additional characteristics of being numerically indexed (as opposed to just being string-keyed like plain objects) and maintaining an automatically updated `.length` property.

## Values as Types

In JavaScript, variables don't have types -- **values have types**. Variables can hold any value, at any time.

Another way to think about JS types is that JS doesn't have "type enforcement," in that the engine doesn't insist that a *variable* always holds values of the *same initial type* that it starts out with. A variable can, in one assignment statement, hold a `string`, and in the next hold a `number`, and so on.

The `value 42` has an intrinsic type of `number`, and its `type` cannot be changed. Another value, like `"42"` with the `string` type, can be created *from* the `number` value `42` through a process called **coercion** (see Chapter 4).

If you use `typeof` against a variable, it's not asking "what's the type of the variable?" as it may seem, since JS variables have no types. Instead, it's asking "what's the type of the value *in* the variable?"

```
var a = 42;
typeof a; // "number"

a = true;
typeof a; // "boolean"
```

The `typeof` operator always returns a string. So:

```
typeof typeof 42; // "string"
```

The first `typeof 42` returns `"number"`, and `typeof "number"` is `"string"`.

## undefined vs "undeclared"

Variables that have no value *currently*, actually have the `undefined` value. Calling `typeof` against such variables will return `"undefined"`:

```
var a;

typeof a; // "undefined"

var b = 42;
var c;

// later
b = c;

typeof b; // "undefined"
typeof c; // "undefined"
```

It's tempting for most developers to think of the word "undefined" and think of it as a synonym for "undeclared." However, in JS, these two concepts are quite different.

An "undefined" variable is one that has been declared in the accessible scope, but *at the moment* has no other value in it. By contrast, an "undeclared" variable is one that has not been formally declared in the accessible scope.

Consider:

```
var a;

a; // undefined
b; // ReferenceError: b is not defined
```

An annoying confusion is the error message that browsers assign to this condition. As you can see, the message is "b is not defined," which is of course very easy and reasonable to confuse with "b is undefined." Yet again, "undefined" and "is not defined" are very different things. It'd be nice if the browsers said something like "b is not found" or "b is not declared," to reduce the confusion!

There's also a special behavior associated with `typeof` as it relates to undeclared variables that even further reinforces the confusion. Consider:

```
var a;

typeof a; // "undefined"

typeof b; // "undefined"
```

The `typeof` operator returns `"undefined"` even for "undeclared" (or "not defined") variables. Notice that there was no error thrown when we executed `typeof b`, even though `b` is an undeclared variable. This is a special safety guard in the behavior of `typeof`.

Similar to above, it would have been nice if `typeof` used with an undeclared variable returned "undeclared" instead of conflating the result value with the different "undefined" case.

## typeof Undeclared

Nevertheless, this safety guard is a useful feature when dealing with JavaScript in the browser, where multiple script files can load variables into the shared global namespace.

**Note:** Many developers believe there should never be any variables in the global namespace, and that everything should be contained in modules and private/separate namespaces. This is great in theory but nearly impossible in practicality; still it's a good goal to strive toward! Fortunately, ES6 added first-class support for modules, which will eventually make that much more practical.

As a simple example, imagine having a "debug mode" in your program that is controlled by a global variable (flag) called `DEBUG`. You'd want to check if that variable was declared before performing a debug task like logging a message to the console. A top-level global `var DEBUG = true` declaration would only be included in a "debug.js" file, which you only load into the browser when you're in development/testing, but not in production.

However, you have to take care in how you check for the global `DEBUG` variable in the rest of your application code, so that you don't throw a `ReferenceError`. The safety guard on `typeof` is our friend in this case.

```
// oops, this would throw an error!
if (DEBUG) {
  console.log( "Debugging is starting" );
}

// this is a safe existence check
if (typeof DEBUG !== "undefined") {
  console.log( "Debugging is starting" );
}
```

This sort of check is useful even if you're not dealing with user-defined variables (like `DEBUG`). If you are doing a feature check for a built-in API, you may also find it helpful to check without throwing an error:

```
if (typeof atob === "undefined") {
  atob = function() { /*...*/ };
}
```

**Note:** If you're defining a "polyfill" for a feature if it doesn't already exist, you probably want to avoid using `var` to make the `atob` declaration. If you declare `var atob` inside the `if` statement, this declaration is hoisted (see the *Scope & Closures* title of this series) to the top of the scope, even if the `if` condition doesn't pass (because the global `atob` already exists!). In some browsers and for some special types of global built-in variables (often called "host objects"), this duplicate declaration may throw an error. Omitting the `var` prevents this hoisted declaration.

Another way of doing these checks against global variables but without the safety guard feature of `typeof` is to observe that all global variables are also properties of the global object, which in the browser is basically the `window` object. So, the above checks could have been done (quite safely) as:

```
if (window.DEBUG) {
  // ...
}

if (!window.atob) {
```

```
// ...
}
```

Unlike referencing undeclared variables, there is no `ReferenceError` thrown if you try to access an object property (even on the global `window` object) that doesn't exist.

On the other hand, manually referencing the global variable with a `window` reference is something some developers prefer to avoid, especially if your code needs to run in multiple JS environments (not just browsers, but server-side node.js, for instance), where the global object may not always be called `window`.

Technically, this safety guard on `typeof` is useful even if you're not using global variables, though these circumstances are less common, and some developers may find this design approach less desirable. Imagine a utility function that you want others to copy-and-paste into their programs or modules, in which you want to check to see if the including program has defined a certain variable (so that you can use it) or not:

```
function doSomethingCool() {
  var helper =
    (typeof FeatureXYZ !== "undefined") ?
    FeatureXYZ :
    function() { /*.. default feature ..*/ };

  var val = helper();
  // ...
}
```

`doSomethingCool()` tests for a variable called `FeatureXYZ`, and if found, uses it, but if not, uses its own. Now, if someone includes this utility in their module/program, it safely checks if they've defined `FeatureXYZ` or not:

```
// an IIFE (see "Immediately Invoked Function Expressions"
// discussion in the *Scope & Closures* title of this series)
(function(){
  function FeatureXYZ() { /*.. my XYZ feature ..*/ }

  // include `doSomethingCool(..)`
  function doSomethingCool() {
    var helper =
      (typeof FeatureXYZ !== "undefined") ?
      FeatureXYZ :
      function() { /*.. default feature ..*/ };

    var val = helper();
    // ...
  }

  doSomethingCool();
})();
```

Here, `FeatureXYZ` is not at all a global variable, but we're still using the safety guard of `typeof` to make it safe to check for. And importantly, here there is *no* object we can use (like we did for global variables with `window.____`) to make the check, so `typeof` is quite helpful.

Other developers would prefer a design pattern called "dependency injection," where instead of `doSomethingCool()` inspecting implicitly for `FeatureXYZ` to be defined outside/around it, it would need to have the dependency explicitly passed in, like:

```
function doSomethingCool(FeatureXYZ) {
  var helper = FeatureXYZ ||
    function() { /*.. default feature ..*/ };

  var val = helper();
  // ...
}
```

There are lots of options when designing such functionality. No one pattern here is "correct" or "wrong" -- there are various tradeoffs to each approach. But overall, it's nice that the `typeof` undeclared safety guard gives us more options.

## Review

JavaScript has seven built-in *types*: `null`, `undefined`, `boolean`, `number`, `string`, `object`, `symbol`. They can be identified by the `typeof` operator.

Variables don't have types, but the values in them do. These types define intrinsic behavior of the values.

Many developers will assume "undefined" and "undeclared" are roughly the same thing, but in JavaScript, they're quite different. `undefined` is a value that a declared variable can hold. "Undeclared" means a variable has never been declared.

JavaScript unfortunately kind of conflates these two terms, not only in its error messages ("ReferenceError: a is not defined") but also in the return values of `typeof`, which is `"undefined"` for both cases.

However, the safety guard (preventing an error) on `typeof` when used against an undeclared variable can be helpful in certain cases.

# You Don't Know JS Yet: Types & Grammar - 2nd Edition

## Chapter 2: Values

**NOTE:**

Work in progress

**NOTE:**

Everything below here is previous text from 1st edition, and is only here for reference while 2nd edition work is underway.  
**Please ignore this stuff.**

`array` `s`, `string` `s`, and `number` `s` are the most basic building-blocks of any program, but JavaScript has some unique characteristics with these types that may either delight or confound you.

Let's look at several of the built-in value types in JS, and explore how we can more fully understand and correctly leverage their behaviors.

## Arrays

As compared to other type-enforced languages, JavaScript `array` `s` are just containers for any type of value, from `string` to `number` to `object` to even another `array` (which is how you get multidimensional `array` `s`).

```
var a = [ 1, "2", [3] ];

a.length;           // 3
a[0] === 1;         // true
a[2][0] === 3;     // true
```

You don't need to presize your `array` `s` (see "Arrays" in Chapter 3), you can just declare them and add values as you see fit:

```
var a = [ ];

a.length;           // 0

a[0] = 1;
a[1] = "2";
a[2] = [ 3 ];
```

```
a.length; // 3
```

**Warning:** Using `delete` on an `array` value will remove that slot from the `array`, but even if you remove the final element, it does **not** update the `length` property, so be careful! We'll cover the `delete` operator itself in more detail in Chapter 5.

Be careful about creating "sparse" `array`s (leaving or creating empty/missing slots):

```
var a = [ ];
a[0] = 1;
// no `a[1]` slot set here
a[2] = [ 3 ];
a[1]; // undefined
a.length; // 3
```

While that works, it can lead to some confusing behavior with the "empty slots" you leave in between. While the slot appears to have the `undefined` value in it, it will not behave the same as if the slot is explicitly set (`a[1] = undefined`). See "Arrays" in Chapter 3 for more information.

`array`s are numerically indexed (as you'd expect), but the tricky thing is that they also are objects that can have `string` keys/properties added to them (but which don't count toward the `length` of the `array`):

```
var a = [ ];
a[0] = 1;
a["foobar"] = 2;

a.length; // 1
a["foobar"]; // 2
a.foobar; // 2
```

However, a gotcha to be aware of is that if a `string` value intended as a key can be coerced to a standard base-10 `number`, then it is assumed that you wanted to use it as a `number` index rather than as a `string` key!

```
var a = [ ];
a["13"] = 42;

a.length; // 14
```

Generally, it's not a great idea to add `string` keys/properties to `array`s. Use `object`s for holding values in keys/properties, and save `array`s for strictly numerically indexed values.

## Array-Likes

There will be occasions where you need to convert an `array`-like value (a numerically indexed collection of values) into a true `array`, usually so you can call array utilities (like `indexOf(..)`, `concat(..)`, `forEach(..)`, etc.) against the collection of values.

For example, various DOM query operations return lists of DOM elements that are not true `array`s but are `array`-like enough for our conversion purposes. Another common example is when functions expose the `arguments` (`array`-like) object (as of ES6, deprecated) to access the arguments as a list.

One very common way to make such a conversion is to borrow the `slice(..)` utility against the value:

```
function foo() {
```

```

var arr = Array.prototype.slice.call( arguments );
arr.push( "bam" );
console.log( arr );
}

foo( "bar", "baz" ); // ["bar", "baz", "bam"]

```

If `slice()` is called without any other parameters, as it effectively is in the above snippet, the default values for its parameters have the effect of duplicating the `array` (or, in this case, `array`-like).

As of ES6, there's also a built-in utility called `Array.from(...)` that can do the same task:

```

...
var arr = Array.from( arguments );
...

```

**Note:** `Array.from(...)` has several powerful capabilities, and will be covered in detail in the *ES6 & Beyond* title of this series.

## Strings

It's a very common belief that `string`s are essentially just `array`s of characters. While the implementation under the covers may or may not use `array`s, it's important to realize that JavaScript `string`s are really not the same as `array`s of characters. The similarity is mostly just skin-deep.

For example, let's consider these two values:

```

var a = "foo";
var b = ["f", "o", "o"];

```

Strings do have a shallow resemblance to `array`s -- `array`-likes, as above -- for instance, both of them having a `length` property, an `indexOf(...)` method (`array` version only as of ES5), and a `concat(...)` method:

```

a.length;                      // 3
b.length;                      // 3

a.indexOf( "o" );              // 1
b.indexOf( "o" );              // 1

var c = a.concat( "bar" );      // "foobar"
var d = b.concat( ["b", "a", "r"] ); // ["f", "o", "o", "b", "a", "r"]

a === c;                       // false
b === d;                       // false

a;                            // "foo"
b;                            // ["f", "o", "o"]

```

So, they're both basically just "arrays of characters", right? **Not exactly**:

```

a[1] = "0";
b[1] = "0";

a; // "foo"
b; // ["f", "0", "o"]

```

JavaScript `string`s are immutable, while `array`s are quite mutable. Moreover, the `a[1]` character position access form was not always widely valid JavaScript. Older versions of IE did not allow that syntax (but now they do). Instead, the *correct* approach has been `a.charAt(1)`.

A further consequence of immutable `string`s is that none of the `string` methods that alter its contents can modify in-place, but rather must create and return new `string`s. By contrast, many of the methods that change `array` contents actually *do* modify in-place.

```
c = a.toUpperCase();
a === c;      // false
a;            // "foo"
c;            // "FOO"

b.push( "!" );
b;            // ["f", "o", "o", "!"]
```

Also, many of the `array` methods that could be helpful when dealing with `string`s are not actually available for them, but we can "borrow" non-mutation `array` methods against our `string`:

```
a.join;        // undefined
a.map;         // undefined

var c = Array.prototype.join.call( a, "-" );
var d = Array.prototype.map.call( a, function(v){
    return v.toUpperCase() + ".";
} ).join( "" );

c;            // "f-o-o"
d;            // "F.O.O."
```

Let's take another example: reversing a `string` (incidentally, a common JavaScript interview trivia question!). `array`s have a `reverse()` in-place mutator method, but `string`s do not:

```
a.reverse;     // undefined

b.reverse();   // ["!", "o", "O", "f"]
b;            // ["!", "o", "O", "f"]
```

Unfortunately, this "borrowing" doesn't work with `array` mutators, because `string`s are immutable and thus can't be modified in place:

```
Array.prototype.reverse.call( a );
// still returns a String object wrapper (see Chapter 3)
// for "foo" :(
```

Another workaround (aka hack) is to convert the `string` into an `array`, perform the desired operation, then convert it back to a `string`.

```
var c = a
    // split `a` into an array of characters
    .split( "" )
    // reverse the array of characters
    .reverse()
    // join the array of characters back to a string
    .join( "" );

c; // "oof"
```

If that feels ugly, it is. Nevertheless, *it works* for simple `string`s, so if you need something quick-n-dirty, often such an approach gets the job done.

**Warning:** Be careful! This approach **doesn't work** for `string`s with complex (unicode) characters in them (astral symbols, multibyte characters, etc.). You need more sophisticated library utilities that are unicode-aware for such operations to be handled accurately. Consult Mathias Bynens' work on the subject: *Esrever* (<https://github.com/mathiasbynens/esrever>).

The other way to look at this is: if you are more commonly doing tasks on your "strings" that treat them as basically *arrays of characters*, perhaps it's better to just actually store them as `array`s rather than as `string`s. You'll probably save yourself a lot of hassle of converting from `string` to `array` each time. You can always call `join("")` on the `array of characters` whenever you actually need the `string` representation.

## Numbers

JavaScript has just one numeric type: `number`. This type includes both "integer" values and fractional decimal numbers. I say "integer" in quotes because it's long been a criticism of JS that there are not true integers, as there are in other languages. That may change at some point in the future, but for now, we just have `number`s for everything.

So, in JS, an "integer" is just a value that has no fractional decimal value. That is, `42.0` is as much an "integer" as `42`.

Like most modern languages, including practically all scripting languages, the implementation of JavaScript's `number`s is based on the "IEEE 754" standard, often called "floating-point." JavaScript specifically uses the "double precision" format (aka "64-bit binary") of the standard.

There are many great write-ups on the Web about the nitty-gritty details of how binary floating-point numbers are stored in memory, and the implications of those choices. Because understanding bit patterns in memory is not strictly necessary to understand how to correctly use `number`s in JS, we'll leave it as an exercise for the interested reader if you'd like to dig further into IEEE 754 details.

## Numeric Syntax

Number literals are expressed in JavaScript generally as base-10 decimal literals. For example:

```
var a = 42;
var b = 42.3;
```

The leading portion of a decimal value, if `0`, is optional:

```
var a = 0.42;
var b = .42;
```

Similarly, the trailing portion (the fractional) of a decimal value after the `.`, if `0`, is optional:

```
var a = 42.0;
var b = 42.;
```

**Warning:** `42.` is pretty uncommon, and probably not a great idea if you're trying to avoid confusion when other people read your code. But it is, nevertheless, valid.

By default, most `number`s will be outputted as base-10 decimals, with trailing fractional `0`s removed. So:

```
var a = 42.300;
var b = 42.0;

a; // 42.3
```

```
b; // 42
```

Very large or very small `number` s will by default be outputted in exponent form, the same as the output of the `toExponential()` method, like:

```
var a = 5E10;
a;                      // 50000000000
a.toExponential();      // "5e+10"

var b = a * a;
b;                      // 2.5e+21

var c = 1 / a;
c;                      // 2e-11
```

Because `number` values can be boxed with the `Number` object wrapper (see Chapter 3), `number` values can access methods that are built into the `Number.prototype` (see Chapter 3). For example, the `toFixed(..)` method allows you to specify how many fractional decimal places you'd like the value to be represented with:

```
var a = 42.59;

a.toFixed( 0 ); // "43"
a.toFixed( 1 ); // "42.6"
a.toFixed( 2 ); // "42.59"
a.toFixed( 3 ); // "42.590"
a.toFixed( 4 ); // "42.5900"
```

Notice that the output is actually a `string` representation of the `number`, and that the value is `0`-padded on the right-hand side if you ask for more decimals than the value holds.

`toPrecision(..)` is similar, but specifies how many *significant digits* should be used to represent the value:

```
var a = 42.59;

a.toPrecision( 1 ); // "4e+1"
a.toPrecision( 2 ); // "43"
a.toPrecision( 3 ); // "42.6"
a.toPrecision( 4 ); // "42.59"
a.toPrecision( 5 ); // "42.590"
a.toPrecision( 6 ); // "42.5900"
```

You don't have to use a variable with the value in it to access these methods; you can access these methods directly on `number` literals. But you have to be careful with the `.` operator. Since `.` is a valid numeric character, it will first be interpreted as part of the `number` literal, if possible, instead of being interpreted as a property accessor.

```
// invalid syntax:
42.toFixed( 3 ); // SyntaxError

// these are all valid:
(42).toFixed( 3 ); // "42.000"
0.42.toFixed( 3 ); // "0.420"
42..toFixed( 3 ); // "42.000"
```

`42.toFixed(3)` is invalid syntax, because the `.` is swallowed up as part of the `42.` literal (which is valid -- see above!), and so then there's no `.` property operator present to make the `.toFixed` access.

`42..toFixed(3)` works because the first `.` is part of the `number` and the second `.` is the property operator. But it probably looks strange, and indeed it's very rare to see something like that in actual JavaScript code. In fact, it's pretty uncommon to access methods directly on any of the primitive values. Uncommon doesn't mean *bad* or *wrong*.

**Note:** There are libraries that extend the built-in `Number.prototype` (see Chapter 3) to provide extra operations on/with `number`s, and so in those cases, it's perfectly valid to use something like `10..makeItRain()` to set off a 10-second money raining animation, or something else silly like that.

This is also technically valid (notice the space):

```
42 .toFixed(3); // "42.000"
```

However, with the `number` literal specifically, **this is particularly confusing coding style** and will serve no other purpose but to confuse other developers (and your future self). Avoid it.

`number`s can also be specified in exponent form, which is common when representing larger `number`s, such as:

```
var onethousand = 1E3; // means 1 * 10^3
var onemilliononehundredthousand = 1.1E6; // means 1.1 * 10^6
```

`number` literals can also be expressed in other bases, like binary, octal, and hexadecimal.

These formats work in current versions of JavaScript:

```
0xf3; // hexadecimal for: 243
0XF3; // ditto

0363; // octal for: 243
```

**Note:** Starting with ES6 + `strict` mode, the `0363` form of octal literals is no longer allowed (see below for the new form). The `0363` form is still allowed in non-`strict` mode, but you should stop using it anyway, to be future-friendly (and because you should be using `strict` mode by now!).

As of ES6, the following new forms are also valid:

```
0o363; // octal for: 243
0o363; // ditto

0b11110011; // binary for: 243
0B11110011; // ditto
```

Please do your fellow developers a favor: never use the `00363` form. `0` next to capital `o` is just asking for confusion. Always use the lowercase predicates `0x`, `0b`, and `0o`.

## Small Decimal Values

The most (in)famous side effect of using binary floating-point numbers (which, remember, is true of **all** languages that use IEEE 754 -- not *just* JavaScript as many assume/pretend) is:

```
0.1 + 0.2 === 0.3; // false
```

Mathematically, we know that statement should be `true`. Why is it `false`?

Simply put, the representations for `0.1` and `0.2` in binary floating-point are not exact, so when they are added, the result is not exactly `0.3`. It's **really** close: `0.3000000000000004`, but if your comparison fails, "close" is irrelevant.

**Note:** Should JavaScript switch to a different `number` implementation that has exact representations for all values? Some think so. There have been many alternatives presented over the years. None of them have been accepted yet, and perhaps never will. As easy as it may seem to just wave a hand and say, "fix that bug already!", it's not nearly that easy. If it were, it most definitely would have been changed a long time ago.

Now, the question is, if some `number`'s can't be *trusted* to be exact, does that mean we can't use `number`'s at all? **Of course not.**

There are some applications where you need to be more careful, especially when dealing with fractional decimal values. There are also plenty of (maybe most?) applications that only deal with whole numbers ("integers"), and moreover, only deal with numbers in the millions or trillions at maximum. These applications have been, and always will be, **perfectly safe** to use numeric operations in JS.

What if we *did* need to compare two `number`'s, like `0.1 + 0.2` to `0.3`, knowing that the simple equality test fails?

The most commonly accepted practice is to use a tiny "rounding error" value as the *tolerance* for comparison. This tiny value is often called "machine epsilon," which is commonly `2^-52` (`2.220446049250313e-16`) for the kind of `number`'s in JavaScript.

As of ES6, `Number.EPSILON` is predefined with this tolerance value, so you'd want to use it, but you can safely polyfill the definition for pre-ES6:

```
if (!Number.EPSILON) {
  Number.EPSILON = Math.pow(2, -52);
}
```

We can use this `Number.EPSILON` to compare two `number`'s for "equality" (within the rounding error tolerance):

```
function numbersCloseEnoughToEqual(n1,n2) {
  return Math.abs( n1 - n2 ) < Number.EPSILON;
}

var a = 0.1 + 0.2;
var b = 0.3;

numbersCloseEnoughToEqual( a, b );           // true
numbersCloseEnoughToEqual( 0.0000001, 0.0000002 ); // false
```

The maximum floating-point value that can be represented is roughly `1.798e+308` (which is really, really, really huge!), predefined for you as `Number.MAX_VALUE`. On the small end, `Number.MIN_VALUE` is roughly `5e-324`, which isn't negative but is really close to zero!

## Safe Integer Ranges

Because of how `number`'s are represented, there is a range of "safe" values for the whole `number` "integers", and it's significantly less than `Number.MAX_VALUE`.

The maximum integer that can "safely" be represented (that is, there's a guarantee that the requested value is actually representable unambiguously) is `2^53 - 1`, which is `9007199254740991`. If you insert your commas, you'll see that this is just over 9 quadrillion. So that's pretty darn big for `number`'s to range up to.

This value is actually automatically predefined in ES6, as `Number.MAX_SAFE_INTEGER`. Unsurprisingly, there's a minimum value, `-9007199254740991`, and it's defined in ES6 as `Number.MIN_SAFE_INTEGER`.

The main way that JS programs are confronted with dealing with such large numbers is when dealing with 64-bit IDs from databases, etc. 64-bit numbers cannot be represented accurately with the `number` type, so must be stored in (and transmitted to/from) JavaScript using `string` representation.

Numeric operations on such large ID `number` values (besides comparison, which will be fine with `string` `s`) aren't all that common, thankfully. But if you *do* need to perform math on these very large values, for now you'll need to use a *big number* utility. Big numbers may get official support in a future version of JavaScript.

## Testing for Integers

To test if a value is an integer, you can use the ES6-specified `Number.isInteger(..)` :

```
Number.isInteger( 42 );           // true
Number.isInteger( 42.000 );       // true
Number.isInteger( 42.3 );         // false
```

To polyfill `Number.isInteger(..)` for pre-ES6:

```
if (!Number.isInteger) {
  Number.isInteger = function(num) {
    return typeof num == "number" && num % 1 == 0;
  };
}
```

To test if a value is a *safe integer*, use the ES6-specified `Number.isSafeInteger(..)` :

```
Number.isSafeInteger( Number.MAX_SAFE_INTEGER );      // true
Number.isSafeInteger( Math.pow( 2, 53 ) );            // false
Number.isSafeInteger( Math.pow( 2, 53 ) - 1 );        // true
```

To polyfill `Number.isSafeInteger(..)` in pre-ES6 browsers:

```
if (!Number.isSafeInteger) {
  Number.isSafeInteger = function(num) {
    return Number.isInteger( num ) &&
      Math.abs( num ) <= Number.MAX_SAFE_INTEGER;
  };
}
```

## 32-bit (Signed) Integers

While integers can range up to roughly 9 quadrillion safely (53 bits), there are some numeric operations (like the bitwise operators) that are only defined for 32-bit `number` s, so the "safe range" for `number` s used in that way must be much smaller.

The range then is `Math.pow(-2, 31)` ( `-2147483648` , about -2.1 billion) up to `Math.pow(2, 31)-1` ( `2147483647` , about +2.1 billion).

To force a `number` value in `a` to a 32-bit signed integer value, use `a | 0` . This works because the `|` bitwise operator only works for 32-bit integer values (meaning it can only pay attention to 32 bits and any other bits will be lost). Then, "or'ing" with zero is essentially a no-op bitwise speaking.

**Note:** Certain special values (which we will cover in the next section) such as `NaN` and `Infinity` are not "32-bit safe," in that those values when passed to a bitwise operator will pass through the abstract operation `ToInt32` (see Chapter 4) and become simply the `+0` value for the purpose of that bitwise operation.

## Special Values

There are several special values spread across the various types that the *alert* JS developer needs to be aware of, and use properly.

## The Non-value Values

For the `undefined` type, there is one and only one value: `undefined`. For the `null` type, there is one and only one value: `null`. So for both of them, the label is both its type and its value.

Both `undefined` and `null` are often taken to be interchangeable as either "empty" values or "non" values. Other developers prefer to distinguish between them with nuance. For example:

- `null` is an empty value
- `undefined` is a missing value

Or:

- `undefined` hasn't had a value yet
- `null` had a value and doesn't anymore

Regardless of how you choose to "define" and use these two values, `null` is a special keyword, not an identifier, and thus you cannot treat it as a variable to assign to (why would you!?). However, `undefined` is (unfortunately) an identifier. Uh oh.

## Undefined

In non-`strict` mode, it's actually possible (though incredibly ill-advised!) to assign a value to the globally provided `undefined` identifier:

```
function foo() {
  undefined = 2; // really bad idea!
}

foo();
```

```
function foo() {
  "use strict";
  undefined = 2; // TypeError!
}

foo();
```

In both non-`strict` mode and `strict` mode, however, you can create a local variable of the name `undefined`. But again, this is a terrible idea!

```
function foo() {
  "use strict";
  var undefined = 2;
  console.log( undefined ); // 2
}

foo();
```

**Friends don't let friends override `undefined`.** Ever.

## void Operator

While `undefined` is a built-in identifier that holds (unless modified -- see above!) the built-in `undefined` value, another way to get this value is the `void` operator.

The expression `void __` "voids" out any value, so that the result of the expression is always the `undefined` value. It doesn't modify the existing value; it just ensures that no value comes back from the operator expression.

```
var a = 42;

console.log( void a, a ); // undefined 42
```

By convention (mostly from C-language programming), to represent the `undefined` value stand-alone by using `void`, you'd use `void 0` (though clearly even `void true` or any other `void` expression does the same thing). There's no practical difference between `void 0`, `void 1`, and `undefined`.

But the `void` operator can be useful in a few other circumstances, if you need to ensure that an expression has no result value (even if it has side effects).

For example:

```
function doSomething() {
    // note: `APP.ready` is provided by our application
    if (!APP.ready) {
        // try again later
        return void setTimeout( doSomething, 100 );
    }

    var result;

    // do some other stuff
    return result;
}

// were we able to do it right away?
if (doSomething()) {
    // handle next tasks right away
}
```

Here, the `setTimeout(..)` function returns a numeric value (the unique identifier of the timer interval, if you wanted to cancel it), but we want to `void` that out so that the return value of our function doesn't give a false-positive with the `if` statement.

Many devs prefer to just do these actions separately, which works the same but doesn't use the `void` operator:

```
if (!APP.ready) {
    // try again later
    setTimeout( doSomething, 100 );
    return;
}
```

In general, if there's ever a place where a value exists (from some expression) and you'd find it useful for the value to be `undefined` instead, use the `void` operator. That probably won't be terribly common in your programs, but in the rare cases you do need it, it can be quite helpful.

## Special Numbers

The `number` type includes several special values. We'll take a look at each in detail.

### The Not Number, Number

Any mathematic operation you perform without both operands being `number`s (or values that can be interpreted as regular `number`s in base 10 or base 16) will result in the operation failing to produce a valid `number`, in which case you will get the `Nan` value.

`NaN` literally stands for "not a `number`", though this label/description is very poor and misleading, as we'll see shortly. It would be much more accurate to think of `NaN` as being "invalid number," "failed number," or even "bad number," than to think of it as "not a number."

For example:

```
var a = 2 / "foo";           // NaN
typeof a === "number";      // true
```

In other words: "the type of not-a-number is 'number'!" Hooray for confusing names and semantics.

`NaN` is a kind of "sentinel value" (an otherwise normal value that's assigned a special meaning) that represents a special kind of error condition within the `number` set. The error condition is, in essence: "I tried to perform a mathematic operation but failed, so here's the failed `number` result instead."

So, if you have a value in some variable and want to test to see if it's this special failed-number `NaN`, you might think you could directly compare to `NaN` itself, as you can with any other value, like `null` or `undefined`. Nope.

```
var a = 2 / "foo";
a == NaN;    // false
a === NaN;   // false
```

`NaN` is a very special value in that it's never equal to another `NaN` value (i.e., it's never equal to itself). It's the only value, in fact, that is not reflexive (without the Identity characteristic `x === x`). So, `NaN !== NaN`. A bit strange, huh?

So how do we test for it, if we can't compare to `NaN` (since that comparison would always fail)?

```
var a = 2 / "foo";
isNaN(a); // true
```

Easy enough, right? We use the built-in global utility called `isNaN(...)` and it tells us if the value is `NaN` or not. Problem solved!

Not so fast.

The `isNaN(...)` utility has a fatal flaw. It appears it tried to take the meaning of `NaN` ("Not a Number") too literally -- that its job is basically: "test if the thing passed in is either not a `number` or is a `number`." But that's not quite accurate.

```
var a = 2 / "foo";
var b = "foo";

a; // NaN
b; // "foo"

window.isNaN(a); // true
window.isNaN(b); // true -- ouch!
```

Clearly, `"foo"` is literally *not a `number`*, but it's definitely not the `NaN` value either! This bug has been in JS since the very beginning (over 19 years of ouch).

As of ES6, finally a replacement utility has been provided: `Number.isNaN(...)`. A simple polyfill for it so that you can safely check `NaN` values *now* even in pre-ES6 browsers is:

```
if (!Number.isNaN) {
  Number.isNaN = function(n) {
```

```

        return (
            typeof n === "number" &&
            window.isNaN( n )
        );
    };
}

var a = 2 / "foo";
var b = "foo";

Number.isNaN( a ); // true
Number.isNaN( b ); // false -- phew!

```

Actually, we can implement a `Number.isNaN(..)` polyfill even easier, by taking advantage of that peculiar fact that `Nan` isn't equal to itself. `Nan` is the *only* value in the whole language where that's true; every other value is always **equal to itself**.

So:

```

if (!Number.isNaN) {
    Number.isNaN = function(n) {
        return n !== n;
    };
}

```

Weird, huh? But it works!

`Nan`s are probably a reality in a lot of real-world JS programs, either on purpose or by accident. It's a really good idea to use a reliable test, like `Number.isNaN(..)` as provided (or polyfilled), to recognize them properly.

If you're currently using just `isNaN(..)` in a program, the sad reality is your program *has a bug*, even if you haven't been bitten by it yet!

## Infinities

Developers from traditional compiled languages like C are probably used to seeing either a compiler error or runtime exception, like "Divide by zero," for an operation like:

```
var a = 1 / 0;
```

However, in JS, this operation is well-defined and results in the value `Infinity` (aka `Number.POSITIVE_INFINITY`).

Unsurprisingly:

```

var a = 1 / 0;      // Infinity
var b = -1 / 0;     // -Infinity

```

As you can see, `-Infinity` (aka `Number.NEGATIVE_INFINITY`) results from a divide-by-zero where either (but not both!) of the divide operands is negative.

JS uses finite numeric representations (IEEE 754 floating-point, which we covered earlier), so contrary to pure mathematics, it seems it *is* possible to overflow even with an operation like addition or subtraction, in which case you'd get `Infinity` or `-Infinity`.

For example:

```

var a = Number.MAX_VALUE;      // 1.7976931348623157e+308
a + a;                      // Infinity
a + Math.pow( 2, 970 );       // Infinity
a + Math.pow( 2, 969 );       // 1.7976931348623157e+308

```

According to the specification, if an operation like addition results in a value that's too big to represent, the IEEE 754 "round-to-nearest" mode specifies what the result should be. So, in a crude sense, `Number.MAX_VALUE + Math.pow( 2, 969 )` is closer to `Number.MAX_VALUE` than to `Infinity`, so it "rounds down," whereas `Number.MAX_VALUE + Math.pow( 2, 970 )` is closer to `Infinity` so it "rounds up".

If you think too much about that, it's going to make your head hurt. So don't. Seriously, stop!

Once you overflow to either one of the *infinities*, however, there's no going back. In other words, in an almost poetic sense, you can go from finite to infinite but not from infinite back to finite.

It's almost philosophical to ask: "What is infinity divided by infinity". Our naive brains would likely say "1" or maybe "infinity." Turns out neither is true. Both mathematically and in JavaScript, `Infinity / Infinity` is not a defined operation. In JS, this results in `Nan`.

But what about any positive finite `number` divided by `Infinity`? That's easy! `0`. And what about a negative finite `number` divided by `Infinity`? Keep reading!

## Zeros

While it may confuse the mathematics-minded reader, JavaScript has both a normal zero `0` (otherwise known as a positive zero `+0`) and a negative zero `-0`. Before we explain why the `-0` exists, we should examine how JS handles it, because it can be quite confusing.

Besides being specified literally as `-0`, negative zero also results from certain mathematic operations. For example:

```
var a = 0 / -3; // -0
var b = 0 * -3; // -0
```

Addition and subtraction cannot result in a negative zero.

A negative zero when examined in the developer console will usually reveal `-0`, though that was not the common case until fairly recently, so some older browsers you encounter may still report it as `0`.

However, if you try to stringify a negative zero value, it will always be reported as `"0"`, according to the spec.

```
var a = 0 / -3;

// (some browser) consoles at least get it right
a; // -0

// but the spec insists on lying to you!
a.toString(); // "0"
a + ""; // "0"
String( a ); // "0"

// strangely, even JSON gets in on the deception
JSON.stringify( a ); // "0"
```

Interestingly, the reverse operations (going from `string` to `number`) don't lie:

```
+"-0"; // -0
Number( "-0" ); // -0
JSON.parse( "-0" ); // -0
```

**Warning:** The `JSON.stringify( -0 )` behavior of `"0"` is particularly strange when you observe that it's inconsistent with the reverse: `JSON.parse( "-0" )` reports `-0` as you'd correctly expect.

In addition to stringification of negative zero being deceptive to hide its true value, the comparison operators are also (intentionally) configured to *lie*.

```
var a = 0;
var b = 0 / -3;

a == b;          // true
-0 == 0;        // true

a === b;         // true
-0 === 0;        // true

0 > -0;          // false
a > b;            // false
```

Clearly, if you want to distinguish a `-0` from a `0` in your code, you can't just rely on what the developer console outputs, so you're going to have to be a bit more clever:

```
function isNegZero(n) {
  n = Number(n);
  return (n === 0) && (1 / n === -Infinity);
}

isNegZero(-0);      // true
isNegZero(0 / -3);  // true
isNegZero(0);        // false
```

Now, why do we need a negative zero, besides academic trivia?

There are certain applications where developers use the magnitude of a value to represent one piece of information (like speed of movement per animation frame) and the sign of that `number` to represent another piece of information (like the direction of that movement).

In those applications, as one example, if a variable arrives at zero and it loses its sign, then you would lose the information of what direction it was moving in before it arrived at zero. Preserving the sign of the zero prevents potentially unwanted information loss.

## Special Equality

As we saw above, the `Nan` value and the `-0` value have special behavior when it comes to equality comparison. `Nan` is never equal to itself, so you have to use ES6's `Number.isNaN(..)` (or a polyfill). Similarly, `-0` lies and pretends that it's equal (even `==` strict equal -- see Chapter 4) to regular positive `0`, so you have to use the somewhat hackish `isNegZero(..)` utility we suggested above.

As of ES6, there's a new utility that can be used to test two values for absolute equality, without any of these exceptions. It's called `Object.is(..)`:

```
var a = 2 / "foo";
var b = -3 * 0;

Object.is(a, NaN);    // true
Object.is(b, -0);     // true

Object.is(b, 0);       // false
```

There's a pretty simple polyfill for `Object.is(..)` for pre-ES6 environments:

```
if (!Object.is) {
  Object.is = function(v1, v2) {
    // test for `~-0`
```

```

        if (v1 === 0 && v2 === 0) {
            return 1 / v1 === 1 / v2;
        }
        // test for `NaN`
        if (v1 !== v1) {
            return v2 !== v2;
        }
        // everything else
        return v1 === v2;
    );
}

```

`Object.is(..)` probably shouldn't be used in cases where `==` or `===` are known to be *safe* (see Chapter 4 "Coercion"), as the operators are likely much more efficient and certainly are more idiomatic/common. `Object.is(..)` is mostly for these special cases of equality.

## Value vs. Reference

In many other languages, values can either be assigned/passed by value-copy or by reference-copy depending on the syntax you use.

For example, in C++ if you want to pass a `number` variable into a function and have that variable's value updated, you can declare the function parameter like `int& myNum`, and when you pass in a variable like `x`, `myNum` will be a **reference to `x`**; references are like a special form of pointers, where you obtain a pointer to another variable (like an *alias*). If you don't declare a reference parameter, the value passed in will *always* be copied, even if it's a complex object.

In JavaScript, there are no pointers, and references work a bit differently. You cannot have a reference from one JS variable to another variable. That's just not possible.

A reference in JS points at a (shared) **value**, so if you have 10 different references, they are all always distinct references to a single shared value; **none of them are references/pointers to each other**.

Moreover, in JavaScript, there are no syntactic hints that control value vs. reference assignment/passing. Instead, the *type* of the value *solely* controls whether that value will be assigned by value-copy or by reference-copy.

Let's illustrate:

```

var a = 2;
var b = a; // `b` is always a copy of the value in `a`
b++;
a; // 2
b; // 3

var c = [1,2,3];
var d = c; // `d` is a reference to the shared `[1,2,3]` value
d.push( 4 );
c; // [1,2,3,4]
d; // [1,2,3,4]

```

Simple values (aka scalar primitives) are *always* assigned/passed by value-copy: `null`, `undefined`, `string`, `number`, `boolean`, and ES6's `symbol`.

Compound values -- `object`'s (including `array`'s, and all boxed object wrappers -- see Chapter 3) and `function`'s -- *always* create a copy of the reference on assignment or passing.

In the above snippet, because `2` is a scalar primitive, `a` holds one initial copy of that value, and `b` is assigned another *copy* of the value. When changing `b`, you are in no way changing the value in `a`.

But **both** `c` **and** `d` are separate references to the same shared value `[1, 2, 3]`, which is a compound value. It's important to note that neither `c` nor `d` more "owns" the `[1, 2, 3]` value -- both are just equal peer references to the value. So, when using either reference to modify (`.push(4)`) the actual shared `array` value itself, it's affecting just the one shared value, and both references will reference the newly modified value `[1, 2, 3, 4]`.

Since references point to the values themselves and not to the variables, you cannot use one reference to change where another reference is pointed:

```
var a = [1, 2, 3];
var b = a;
a; // [1, 2, 3]
b; // [1, 2, 3]

// later
b = [4, 5, 6];
a; // [1, 2, 3]
b; // [4, 5, 6]
```

When we make the assignment `b = [4, 5, 6]`, we are doing absolutely nothing to affect *where* `a` is still referencing (`[1, 2, 3]`). To do that, `b` would have to be a pointer to `a` rather than a reference to the `array` -- but no such capability exists in JS!

The most common way such confusion happens is with function parameters:

```
function foo(x) {
  x.push( 4 );
  x; // [1, 2, 3, 4]

  // later
  x = [4, 5, 6];
  x.push( 7 );
  x; // [4, 5, 6, 7]
}

var a = [1, 2, 3];

foo( a );

a; // [1, 2, 3, 4] not [4, 5, 6, 7]
```

When we pass in the argument `a`, it assigns a copy of the `a` reference to `x`. `x` and `a` are separate references pointing at the same `[1, 2, 3]` value. Now, inside the function, we can use that reference to mutate the value itself (`push(4)`). But when we make the assignment `x = [4, 5, 6]`, this is in no way affecting where the initial reference `a` is pointing -- still points at the (now modified) `[1, 2, 3, 4]` value.

There is no way to use the `x` reference to change where `a` is pointing. We could only modify the contents of the shared value that both `a` and `x` are pointing to.

To accomplish changing `a` to have the `[4, 5, 6, 7]` value contents, you can't create a new `array` and assign -- you must modify the existing `array` value:

```
function foo(x) {
  x.push( 4 );
  x; // [1, 2, 3, 4]

  // later
  x.length = 0; // empty existing array in-place
  x.push( 4, 5, 6, 7 );
  x; // [4, 5, 6, 7]
}
```

```
var a = [1,2,3];
foo( a );
a; // [4,5,6,7] not [1,2,3,4]
```

As you can see, `x.length = 0` and `x.push(4,5,6,7)` were not creating a new `array`, but modifying the existing shared `array`. So of course, `a` references the new `[4,5,6,7]` contents.

Remember: you cannot directly control/override value-copy vs. reference -- those semantics are controlled entirely by the type of the underlying value.

To effectively pass a compound value (like an `array`) by value-copy, you need to manually make a copy of it, so that the reference passed doesn't still point to the original. For example:

```
foo( a.slice() );
```

`slice(..)` with no parameters by default makes an entirely new (shallow) copy of the `array`. So, we pass in a reference only to the copied `array`, and thus `foo(..)` cannot affect the contents of `a`.

To do the reverse -- pass a scalar primitive value in a way where its value updates can be seen, kinda like a reference -- you have to wrap the value in another compound value (`object`, `array`, etc) that *can* be passed by reference-copy:

```
function foo(wrapper) {
  wrapper.a = 42;
}

var obj = {
  a: 2
};

foo( obj );

obj.a; // 42
```

Here, `obj` acts as a wrapper for the scalar primitive property `a`. When passed to `foo(..)`, a copy of the `obj` reference is passed in and set to the `wrapper` parameter. We now can use the `wrapper` reference to access the shared object, and update its property. After the function finishes, `obj.a` will see the updated value `42`.

It may occur to you that if you wanted to pass in a reference to a scalar primitive value like `2`, you could just box the value in its `Number` object wrapper (see Chapter 3).

It is true a copy of the reference to this `Number` object *will* be passed to the function, but unfortunately, having a reference to the shared object is not going to give you the ability to modify the shared primitive value, like you may expect:

```
function foo(x) {
  x = x + 1;
  x; // 3
}

var a = 2;
var b = new Number( a ); // or equivalently `Object(a)`

foo( b );
console.log( b ); // 2, not 3
```

The problem is that the underlying scalar primitive value is *not mutable* (same goes for `String` and `Boolean`). If a `Number` object holds the scalar primitive value `2`, that exact `Number` object can never be changed to hold another value; you can only create a whole new `Number` object with a different value.

When `x` is used in the expression `x + 1`, the underlying scalar primitive value `2` is unboxed (extracted) from the `Number` object automatically, so the line `x = x + 1` very subtly changes `x` from being a shared reference to the `Number` object, to just holding the scalar primitive value `3` as a result of the addition operation `2 + 1`. Therefore, `b` on the outside still references the original unmodified/immutable `Number` object holding the value `2`.

You *can* add properties on top of the `Number` object (just not change its inner primitive value), so you could exchange information indirectly via those additional properties.

This is not all that common, however; it probably would not be considered a good practice by most developers.

Instead of using the wrapper object `Number` in this way, it's probably much better to use the manual object wrapper (`obj`) approach in the earlier snippet. That's not to say that there's no clever uses for the boxed object wrappers like `Number` -- just that you should probably prefer the scalar primitive value form in most cases.

References are quite powerful, but sometimes they get in your way, and sometimes you need them where they don't exist. The only control you have over reference vs. value-copy behavior is the type of the value itself, so you must indirectly influence the assignment/passing behavior by which value types you choose to use.

## Review

In JavaScript, `array`s are simply numerically indexed collections of any value-type. `string`s are somewhat "array -like", but they have distinct behaviors and care must be taken if you want to treat them as `array`s. Numbers in JavaScript include both "integers" and floating-point values.

Several special values are defined within the primitive types.

The `null` type has just one value: `null`, and likewise the `undefined` type has just the `undefined` value. `undefined` is basically the default value in any variable or property if no other value is present. The `void` operator lets you create the `undefined` value from any other value.

`number`s include several special values, like `NaN` (supposedly "Not a Number", but really more appropriately "invalid number"); `+Infinity` and `-Infinity`; and `-0`.

Simple scalar primitives (`string`s, `number`s, etc.) are assigned/passed by value-copy, but compound values (`object`s, etc.) are assigned/passed by reference-copy. References are not like references/pointers in other languages -- they're never pointed at other variables/references, only at the underlying values.

# You Don't Know JS Yet: Types & Grammar - 2nd Edition

## Chapter 3: Natives

<b>NOTE:</b>
Work in progress

<b>NOTE:</b>
Everything below here is previous text from 1st edition, and is only here for reference while 2nd edition work is underway. <b>Please ignore this stuff.</b>

Everything below here is previous text from 1st edition, and is only here for reference while 2nd edition work is underway.  
**Please ignore this stuff.**

Several times in Chapters 1 and 2, we alluded to various built-ins, usually called "natives," like `String` and `Number`. Let's examine those in detail now.

Here's a list of the most commonly used natives:

- `String()`
- `Number()`
- `Boolean()`
- `Array()`
- `Object()`
- `Function()`
- `RegExp()`
- `Date()`
- `Error()`
- `Symbol()` -- added in ES6!

As you can see, these natives are actually built-in functions.

If you're coming to JS from a language like Java, JavaScript's `String()` will look like the `String(...)` constructor you're used to for creating string values. So, you'll quickly observe that you can do things like:

```
var s = new String( "Hello World!" );
console.log( s.toString() ); // "Hello World!"
```

It is true that each of these natives can be used as a native constructor. But what's being constructed may be different than you think.

```
var a = new String( "abc" );
typeof a; // "object" ... not "String"
a instanceof String; // true
Object.prototype.toString.call( a ); // "[object String]"
```

The result of the constructor form of value creation (`new String("abc")`) is an object wrapper around the primitive (`"abc"`) value.

Importantly, `typeof` shows that these objects are not their own special *types*, but more appropriately they are subtypes of the `object` type.

This object wrapper can further be observed with:

```
console.log( a );
```

The output of that statement varies depending on your browser, as developer consoles are free to choose however they feel it's appropriate to serialize the object for developer inspection.

**Note:** At the time of writing, the latest Chrome prints something like this: `String {0: "a", 1: "b", 2: "c", length: 3, [[PrimitiveValue]]: "abc"}`. But older versions of Chrome used to just print this: `String {0: "a", 1: "b", 2: "c"}`. The latest Firefox currently prints `String ["a", "b", "c"]`, but used to print `"abc"` in italics, which was clickable to open the object inspector. Of course, these results are subject to rapid change and your experience may vary.

The point is, `new String("abc")` creates a string wrapper object around `"abc"`, not just the primitive `"abc"` value itself.

## Internal `[[Class]]`

Values that are `typeof "object"` (such as an array) are additionally tagged with an internal `[[Class]]` property (think of this more as an internal *classification* rather than related to classes from traditional class-oriented coding). This property cannot be accessed directly, but can generally be revealed indirectly by borrowing the default `Object.prototype.toString()` method called against the value. For example:

```
Object.prototype.toString.call( [1,2,3] );           // "[object Array]"
Object.prototype.toString.call( /regex-literal/i );    // "[object RegExp]"
```

So, for the array in this example, the internal `[[Class]]` value is `"Array"`, and for the regular expression, it's `"RegExp"`. In most cases, this internal `[[Class]]` value corresponds to the built-in native constructor (see below) that's related to the value, but that's not always the case.

What about primitive values? First, `null` and `undefined`:

```
Object.prototype.toString.call( null );           // "[object Null]"
Object.prototype.toString.call( undefined );      // "[object Undefined]"
```

You'll note that there are no `Null()` or `Undefined()` native constructors, but nevertheless the `"Null"` and `"Undefined"` are the internal `[[Class]]` values exposed.

But for the other simple primitives like `string`, `number`, and `boolean`, another behavior actually kicks in, which is usually called "boxing" (see "Boxing Wrappers" section next):

```
Object.prototype.toString.call( "abc" ); // "[object String]"
Object.prototype.toString.call( 42 ); // "[object Number]"
Object.prototype.toString.call( true ); // "[object Boolean]"
```

In this snippet, each of the simple primitives are automatically boxed by their respective object wrappers, which is why `"String"`, `"Number"`, and `"Boolean"` are revealed as the respective internal `[[Class]]` values.

**Note:** The behavior of `toString()` and `[[Class]]` as illustrated here has changed a bit from ES5 to ES6, but we cover those details in the *ES6 & Beyond* title of this series.

## Boxing Wrappers

These object wrappers serve a very important purpose. Primitive values don't have properties or methods, so to access `.length` or `.toString()` you need an object wrapper around the value. Thankfully, JS will automatically *box* (aka wrap) the primitive value to fulfill such accesses.

```
var a = "abc";
a.length; // 3
a.toUpperCase(); // "ABC"
```

So, if you're going to be accessing these properties/methods on your string values regularly, like a `i < a.length` condition in a `for` loop for instance, it might seem to make sense to just have the object form of the value from the start, so the JS engine doesn't need to implicitly create it for you.

But it turns out that's a bad idea. Browsers long ago performance-optimized the common cases like `.length`, which means your program will *actually go slower* if you try to "preoptimize" by directly using the object form (which isn't on the optimized path).

In general, there's basically no reason to use the object form directly. It's better to just let the boxing happen implicitly where necessary. In other words, never do things like `new String("abc")`, `new Number(42)`, etc -- always prefer using the literal primitive values `"abc"` and `42`.

## Object Wrapper Gotchas

There are some gotchas with using the object wrappers directly that you should be aware of if you *do* choose to ever use them.

For example, consider `Boolean` wrapped values:

```
var a = new Boolean( false );
if (!a) {
  console.log( "Oops" ); // never runs
}
```

The problem is that you've created an object wrapper around the `false` value, but objects themselves are "truthy" (see Chapter 4), so using the object behaves oppositely to using the underlying `false` value itself, which is quite contrary to normal expectation.

If you want to manually box a primitive value, you can use the `Object(..)` function (no `new` keyword):

```
var a = "abc";
var b = new String( a );
```

```
var c = Object( a );

typeof a; // "string"
typeof b; // "object"
typeof c; // "object"

b instanceof String; // true
c instanceof String; // true

Object.prototype.toString.call( b ); // "[object String]"
Object.prototype.toString.call( c ); // "[object String]"
```

Again, using the boxed object wrapper directly (like `b` and `c` above) is usually discouraged, but there may be some rare occasions you'll run into where they may be useful.

## Unboxing

If you have an object wrapper and you want to get the underlying primitive value out, you can use the `valueOf()` method:

```
var a = new String( "abc" );
var b = new Number( 42 );
var c = new Boolean( true );

a.valueOf(); // "abc"
b.valueOf(); // 42
c.valueOf(); // true
```

Unboxing can also happen implicitly, when using an object wrapper value in a way that requires the primitive value. This process (coercion) will be covered in more detail in Chapter 4, but briefly:

```
var a = new String( "abc" );
var b = a + ""; // `b` has the unboxed primitive value "abc"

typeof a; // "object"
typeof b; // "string"
```

## Natives as Constructors

For `array`, `object`, `function`, and regular-expression values, it's almost universally preferred that you use the literal form for creating the values, but the literal form creates the same sort of object as the constructor form does (that is, there is no nonwrapped value).

Just as we've seen above with the other natives, these constructor forms should generally be avoided, unless you really know you need them, mostly because they introduce exceptions and gotchas that you probably don't really *want* to deal with.

### Array(..)

```
var a = new Array( 1, 2, 3 );
a; // [1, 2, 3]

var b = [1, 2, 3];
b; // [1, 2, 3]
```

**Note:** The `Array(..)` constructor does not require the `new` keyword in front of it. If you omit it, it will behave as if you have used it anyway. So `Array(1,2,3)` is the same outcome as `new Array(1,2,3)`.

The `Array` constructor has a special form where if only one `number` argument is passed, instead of providing that value as `contents` of the array, it's taken as a length to "presize the array" (well, sorta).

This is a terrible idea. Firstly, you can trip over that form accidentally, as it's easy to forget.

But more importantly, there's no such thing as actually presizing the array. Instead, what you're creating is an otherwise empty array, but setting the `length` property of the array to the numeric value specified.

An array that has no explicit values in its slots, but has a `length` property that *implies* the slots exist, is a weird exotic type of data structure in JS with some very strange and confusing behavior. The capability to create such a value comes purely from old, deprecated, historical functionalities ("array-like objects" like the `arguments` object).

**Note:** An array with at least one "empty slot" in it is often called a "sparse array."

It doesn't help matters that this is yet another example where browser developer consoles vary on how they represent such an object, which breeds more confusion.

For example:

```
var a = new Array( 3 );
a.length; // 3
a;
```

The serialization of `a` in Chrome is (at the time of writing): `[ undefined × 3 ]`. **This is really unfortunate.** It implies that there are three `undefined` values in the slots of this array, when in fact the slots do not exist (so-called "empty slots" -- also a bad name!).

To visualize the difference, try this:

```
var a = new Array( 3 );
var b = [ undefined, undefined, undefined ];
var c = [];
c.length = 3;

a;
b;
c;
```

**Note:** As you can see with `c` in this example, empty slots in an array can happen after creation of the array. Changing the `length` of an array to go beyond its number of actually-defined slot values, you implicitly introduce empty slots. In fact, you could even call `delete b[1]` in the above snippet, and it would introduce an empty slot into the middle of `b`.

For `b` (in Chrome, currently), you'll find `[ undefined, undefined, undefined ]` as the serialization, as opposed to `[ undefined × 3 ]` for `a` and `c`. Confused? Yeah, so is everyone else.

Worse than that, at the time of writing, Firefox reports `[ , , , ]` for `a` and `c`. Did you catch why that's so confusing? Look closely. Three commas implies four slots, not three slots like we'd expect.

**What!?** Firefox puts an extra `,` on the end of their serialization here because as of ES5, trailing commas in lists (array values, property lists, etc.) are allowed (and thus dropped and ignored). So if you were to type in a `[ , , , ]` value into your program or the console, you'd actually get the underlying value that's like `[ , , ]` (that is, an array with three empty slots). This choice, while confusing if reading the developer console, is defended as instead making copy-n-paste behavior accurate.

If you're shaking your head or rolling your eyes about now, you're not alone! Shrugs.

Unfortunately, it gets worse. More than just confusing console output, `a` and `b` from the above code snippet actually behave the same in some cases **but differently in others**:

```
a.join( "-" ); // "--"
b.join( "-" ); // "--"

a.map(function(v,i){ return i; }); // [ undefined x 3 ]
b.map(function(v,i){ return i; }); // [ 0, 1, 2 ]
```

**Ugh.**

The `a.map(..)` call *fails* because the slots don't actually exist, so `map(..)` has nothing to iterate over. `join(..)` works differently. Basically, we can think of it implemented sort of like this:

```
function fakeJoin(arr, connector) {
  var str = "";
  for (var i = 0; i < arr.length; i++) {
    if (i > 0) {
      str += connector;
    }
    if (arr[i] !== undefined) {
      str += arr[i];
    }
  }
  return str;
}

var a = new Array( 3 );
fakeJoin( a, "-" ); // "--"
```

As you can see, `join(..)` works by just *assuming* the slots exist and looping up to the `length` value. Whatever `map(..)` does internally, it (apparently) doesn't make such an assumption, so the result from the strange "empty slots" array is unexpected and likely to cause failure.

So, if you wanted to *actually* create an array of actual `undefined` values (not just "empty slots"), how could you do it (besides manually)?

```
var a = Array.apply( null, { length: 3 } );
a; // [ undefined, undefined, undefined ]
```

Confused? Yeah. Here's roughly how it works.

`apply(..)` is a utility available to all functions, which calls the function it's used with but in a special way.

The first argument is a `this` object binding (covered in the *this & Object Prototypes* title of this series), which we don't care about here, so we set it to `null`. The second argument is supposed to be an array (or something *like* an array -- aka an "array-like object"). The contents of this "array" are "spread" out as arguments to the function in question.

So, `Array.apply(..)` is calling the `Array(..)` function and spreading out the values (of the `{ length: 3 }` object value) as its arguments.

Inside of `apply(..)`, we can envision there's another `for` loop (kinda like `join(..)` from above) that goes from `0` up to, but not including, `length` (`3` in our case).

For each index, it retrieves that key from the object. So if the array-object parameter was named `arr` internally inside of the `apply(..)` function, the property access would effectively be `arr[0]`, `arr[1]`, and `arr[2]`. Of course, none of those properties exist on the `{ length: 3 }` object value, so all three of those property accesses would return the value `undefined`.

In other words, it ends up calling `Array(..)` basically like this: `Array(undefined,undefined,undefined)`, which is how we end up with an array filled with `undefined` values, and not just those (crazy) empty slots.

While `Array.apply( null, { length: 3 } )` is a strange and verbose way to create an array filled with `undefined` values, it's **vastly** better and more reliable than what you get with the footgun'ish `Array(3)` empty slots.

Bottom line: **never ever, under any circumstances**, should you intentionally create and use these exotic empty-slot arrays. Just don't do it. They're nuts.

## Object(..) , Function(..) , and RegExp(..)

The `Object(..) / Function(..) / RegExp(..)` constructors are also generally optional (and thus should usually be avoided unless specifically called for):

```
var c = new Object();
c.foo = "bar";
c; // { foo: "bar" }

var d = { foo: "bar" };
d; // { foo: "bar" }

var e = new Function( "a", "return a * 2;" );
var f = function(a) { return a * 2; };
function g(a) { return a * 2; }

var h = new RegExp( "^a*b+", "g" );
var i = /^a*b+/g;
```

There's practically no reason to ever use the `new Object()` constructor form, especially since it forces you to add properties one-by-one instead of many at once in the object literal form.

The `Function` constructor is helpful only in the rarest of cases, where you need to dynamically define a function's parameters and/or its function body. **Do not just treat `Function(..)` as an alternate form of `eval(..)`**. You will almost never need to dynamically define a function in this way.

Regular expressions defined in the literal form (`/^a*b+/g`) are strongly preferred, not just for ease of syntax but for performance reasons -- the JS engine precompiles and caches them before code execution. Unlike the other constructor forms we've seen so far, `RegExp(..)` has some reasonable utility: to dynamically define the pattern for a regular expression.

```
var name = "Kyle";
var namePattern = new RegExp( "\b(?: " + name + ")+\b", "ig" );

var matches = someText.match( namePattern );
```

This kind of scenario legitimately occurs in JS programs from time to time, so you'd need to use the `new RegExp("pattern", "flags")` form.

## Date(..) and Error(..)

The `Date(..)` and `Error(..)` native constructors are much more useful than the other natives, because there is no literal form for either.

To create a date object value, you must use `new Date()`. The `Date(..)` constructor accepts optional arguments to specify the date/time to use, but if omitted, the current date/time is assumed.

By far the most common reason you construct a date object is to get the current timestamp value (a signed integer number of milliseconds since Jan 1, 1970). You can do this by calling `getTime()` on a date object instance.

But an even easier way is to just call the static helper function defined as of ES5: `date.now()`. And to polyfill that for pre-ES5 is pretty easy:

```

if (!Date.now) {
  Date.now = function(){
    return (new Date()).getTime();
  };
}

```

**Note:** If you call `Date()` without `new`, you'll get back a string representation of the date/time at that moment. The exact form of this representation is not specified in the language spec, though browsers tend to agree on something close to: `"Fri Jul 18 2014 00:31:02 GMT-0500 (CDT)"`.

The `Error(..)` constructor (much like `Array()` above) behaves the same with the `new` keyword present or omitted.

The main reason you'd want to create an error object is that it captures the current execution stack context into the object (in most JS engines, revealed as a read-only `.stack` property once constructed). This stack context includes the function call-stack and the line-number where the error object was created, which makes debugging that error much easier.

You would typically use such an error object with the `throw` operator:

```

function foo(x) {
  if (!x) {
    throw new Error( "x wasn't provided" );
  }
  // ...
}

```

Error object instances generally have at least a `message` property, and sometimes other properties (which you should treat as read-only), like `type`. However, other than inspecting the above-mentioned `stack` property, it's usually best to just call `toString()` on the error object (either explicitly, or implicitly through coercion -- see Chapter 4) to get a friendly-formatted error message.

**Tip:** Technically, in addition to the general `Error(..)` native, there are several other specific-error-type natives:

`Evaluator()`, `RangeError()`, `ReferenceError()`, `SyntaxError()`, `TypeError()`, and `URIError()`. But it's very rare to manually use these specific error natives. They are automatically used if your program actually suffers from a real exception (such as referencing an undeclared variable and getting a `ReferenceError` error).

## Symbol(..)

New as of ES6, an additional primitive value type has been added, called "Symbol". Symbols are special "unique" (not strictly guaranteed!) values that can be used as properties on objects with little fear of any collision. They're primarily designed for special built-in behaviors of ES6 constructs, but you can also define your own symbols.

Symbols can be used as property names, but you cannot see or access the actual value of a symbol from your program, nor from the developer console. If you evaluate a symbol in the developer console, what's shown looks like `Symbol(Symbol.create)`, for example.

There are several predefined symbols in ES6, accessed as static properties of the `Symbol` function object, like `Symbol.create`, `Symbol.iterator`, etc. To use them, do something like:

```
obj[Symbol.iterator] = function(){ /*...*/ };
```

To define your own custom symbols, use the `Symbol(..)` native. The `Symbol(..)` native "constructor" is unique in that you're not allowed to use `new` with it, as doing so will throw an error.

```

var mysym = Symbol( "my own symbol" );
mysym;           // Symbol(my own symbol)
mysym.toString(); // "Symbol(my own symbol)"
typeof mysym;    // "symbol"

```

```
var a = { };
a[mysym] = "foobar";

Object.getOwnPropertySymbols( a );
// [ Symbol(my own symbol) ]
```

While symbols are not actually private (`Object.getOwnPropertySymbols(..)` reflects on the object and reveals the symbols quite publicly), using them for private or special properties is likely their primary use-case. For most developers, they may take the place of property names with `_` underscore prefixes, which are almost always by convention signals to say, "hey, this is a private/special/internal property, so leave it alone!"

**Note:** `Symbol`s are *not* `object`s, they are simple scalar primitives.

## Native Prototypes

Each of the built-in native constructors has its own `.prototype` object -- `Array.prototype`, `String.prototype`, etc.

These objects contain behavior unique to their particular object subtype.

For example, all string objects, and by extension (via boxing) `string` primitives, have access to default behavior as methods defined on the `String.prototype` object.

**Note:** By documentation convention, `String.prototype.XYZ` is shortened to `String#XYZ`, and likewise for all the other `.prototype`s.

- `String#indexOf(..)` : find the position in the string of another substring
- `String#charAt(..)` : access the character at a position in the string
- `String#substr(..)`, `String#substring(..)`, and `String#slice(..)` : extract a portion of the string as a new string
- `String#toUpperCase()` and `String#toLowerCase()` : create a new string that's converted to either uppercase or lowercase
- `String#trim()` : create a new string that's stripped of any trailing or leading whitespace

None of the methods modify the string *in place*. Modifications (like case conversion or trimming) create a new value from the existing value.

By virtue of prototype delegation (see the *this & Object Prototypes* title in this series), any string value can access these methods:

```
var a = " abc ";

a.indexOf( "c" ); // 3
a.toUpperCase(); // " ABC "
a.trim(); // "abc"
```

The other constructor prototypes contain behaviors appropriate to their types, such as `Number#toFixed(..)` (stringifying a number with a fixed number of decimal digits) and `Array#concat(..)` (merging arrays). All functions have access to `apply(..)`, `call(..)`, and `bind(..)` because `Function.prototype` defines them.

But, some of the native prototypes aren't *just* plain objects:

```
typeof Function.prototype;           // "function"
Function.prototype();               // it's an empty function!

RegExp.prototype.toString();        // "/(?:)/" -- empty regex
"abc".match( RegExp.prototype );   // [""]
```

A particularly bad idea, you can even modify these native prototypes (not just adding properties as you're probably familiar with):

```
Array.isArray( Array.prototype );    // true
```

```
Array.prototype.push( 1, 2, 3 );      // 3
Array.prototype;                      // [1,2,3]

// don't leave it that way, though, or expect weirdness!
// reset the `Array.prototype` to empty
Array.prototype.length = 0;
```

As you can see, `Function.prototype` is a function, `RegExp.prototype` is a regular expression, and `Array.prototype` is an array. Interesting and cool, huh?

## Prototypes As Defaults

`Function.prototype` being an empty function, `RegExp.prototype` being an "empty" (e.g., non-matching) regex, and `Array.prototype` being an empty array, make them all nice "default" values to assign to variables if those variables wouldn't already have had a value of the proper type.

For example:

```
function isThisCool(vals,fn,rx) {
  vals = vals || Array.prototype;
  fn = fn || Function.prototype;
  rx = rx || RegExp.prototype;

  return rx.test(
    vals.map( fn ).join( "" )
  );
}

isThisCool();           // true

isThisCool(
  ["a","b","c"],
  function(v){ return v.toUpperCase(); },
  /D/
);                     // false
```

**Note:** As of ES6, we don't need to use the `vals = vals || ..` default value syntax trick (see Chapter 4) anymore, because default values can be set for parameters via native syntax in the function declaration (see Chapter 5).

One minor side-benefit of this approach is that the `.prototype`s are already created and built-in, thus created *only once*. By contrast, using `[]`, `function(){}, and /(?:)/ values themselves for those defaults would (likely, depending on engine implementations) be recreating those values (and probably garbage-collecting them later) for each call of isThisCool(..). That could be memory/CPU wasteful.`

Also, be very careful not to use `Array.prototype` as a default value **that will subsequently be modified**. In this example, `vals` is used read-only, but if you were to instead make in-place changes to `vals`, you would actually be modifying `Array.prototype` itself, which would lead to the gotchas mentioned earlier!

**Note:** While we're pointing out these native prototypes and some usefulness, be cautious of relying on them and even more wary of modifying them in any way. See Appendix A "Native Prototypes" for more discussion.

## Review

JavaScript provides object wrappers around primitive values, known as natives (`String`, `Number`, `Boolean`, etc). These object wrappers give the values access to behaviors appropriate for each object subtype (`String#trim()` and `Array#concat(..)`).

If you have a simple scalar primitive value like `"abc"` and you access its `length` property or some `String.prototype` method, JS automatically "boxes" the value (wraps it in its respective object wrapper) so that the property/method accesses can be fulfilled.

# You Don't Know JS Yet: Types & Grammar - 2nd Edition

## Chapter 4: Coercion

<b>NOTE:</b>
Work in progress

<b>NOTE:</b>
Everything below here is previous text from 1st edition, and is only here for reference while 2nd edition work is underway. <b>Please ignore this stuff.</b>

Everything below here is previous text from 1st edition, and is only here for reference while 2nd edition work is underway.  
**Please ignore this stuff.**

Now that we much more fully understand JavaScript's types and values, we turn our attention to a very controversial topic: coercion.

As we mentioned in Chapter 1, the debates over whether coercion is a useful feature or a flaw in the design of the language (or somewhere in between!) have raged since day one. If you've read other popular books on JS, you know that the overwhelmingly prevalent *message* out there is that coercion is magical, evil, confusing, and just downright a bad idea.

In the same overall spirit of this book series, rather than running away from coercion because everyone else does, or because you get bitten by some quirk, I think you should run toward that which you don't understand and seek to *get it* more fully.

Our goal is to fully explore the pros and cons (yes, there *are* pros!) of coercion, so that you can make an informed decision on its appropriateness in your program.

## Converting Values

Converting a value from one type to another is often called "type casting," when done explicitly, and "coercion" when done implicitly (forced by the rules of how a value is used).

**Note:** It may not be obvious, but JavaScript coercions always result in one of the scalar primitive (see Chapter 2) values, like `string`, `number`, or `boolean`. There is no coercion that results in a complex value like `object` or `function`. Chapter 3 covers "boxing," which wraps scalar primitive values in their `object` counterparts, but this is not really coercion in an accurate sense.

Another way these terms are often distinguished is as follows: "type casting" (or "type conversion") occur in statically typed languages at compile time, while "type coercion" is a runtime conversion for dynamically typed languages.

However, in JavaScript, most people refer to all these types of conversions as *coercion*, so the way I prefer to distinguish is to say "implicit coercion" vs. "explicit coercion."

The difference should be obvious: "explicit coercion" is when it is obvious from looking at the code that a type conversion is intentionally occurring, whereas "implicit coercion" is when the type conversion will occur as a less obvious side effect of some other intentional operation.

For example, consider these two approaches to coercion:

```
var a = 42;

var b = a + "";           // implicit coercion

var c = String( a );     // explicit coercion
```

For `b`, the coercion that occurs happens implicitly, because the `+` operator combined with one of the operands being a `string` value (`" "`) will insist on the operation being a `string` concatenation (adding two strings together), which *as a (hidden) side effect* will force the `42` value in `a` to be coerced to its `string` equivalent: `"42"`.

By contrast, the `String(..)` function makes it pretty obvious that it's explicitly taking the value in `a` and coercing it to a `string` representation.

Both approaches accomplish the same effect: `"42"` comes from `42`. But it's the *how* that is at the heart of the heated debates over JavaScript coercion.

**Note:** Technically, there's some nuanced behavioral difference here beyond the stylistic difference. We cover that in more detail later in the chapter, in the "Implicitly: Strings <--> Numbers" section.

The terms "explicit" and "implicit," or "obvious" and "hidden side effect," are *relative*.

If you know exactly what `a + ""` is doing and you're intentionally doing that to coerce to a `string`, you might feel the operation is sufficiently "explicit." Conversely, if you've never seen the `String(..)` function used for `string` coercion, its behavior might seem hidden enough as to feel "implicit" to you.

But we're having this discussion of "explicit" vs. "implicit" based on the likely opinions of an *average, reasonably informed, but not expert or JS specification devotee* developer. To whatever extent you do or do not find yourself fitting neatly in that bucket, you will need to adjust your perspective on our observations here accordingly.

Just remember: it's often rare that we write our code and are the only ones who ever read it. Even if you're an expert on all the ins and outs of JS, consider how a less experienced teammate of yours will feel when they read your code. Will it be "explicit" or "implicit" to them in the same way it is for you?

## Abstract Value Operations

Before we can explore *explicit* vs *implicit* coercion, we need to learn the basic rules that govern how values *become* either a `string`, `number`, or `boolean`. The ES5 spec in section 9 defines several "abstract operations" (fancy spec-speak for "internal-only operation") with the rules of value conversion. We will specifically pay attention to: `ToString`, `ToNumber`, and `ToBoolean`, and to a lesser extent, `ToPrimitive`.

### To**String**

When any non-`string` value is coerced to a `string` representation, the conversion is handled by the `ToString` abstract operation in section 9.8 of the specification.

Built-in primitive values have natural stringification: `null` becomes `"null"`, `undefined` becomes `"undefined"` and `true` becomes `"true"`. `number`s are generally expressed in the natural way you'd expect, but as we discussed in Chapter 2, very small or very large `numbers` are represented in exponent form:

```
// multiplying `1.07` by `1000`, seven times over
var a = 1.07 * 1000 * 1000 * 1000 * 1000 * 1000 * 1000;

// seven times three digits => 21 digits
a.toString(); // "1.07e21"
```

For regular objects, unless you specify your own, the default `toString()` (located in `Object.prototype.toString()`) will return the *internal* `[[Class]]` (see Chapter 3), like for instance `"[object Object]"`.

But as shown earlier, if an object has its own `toString()` method on it, and you use that object in a `string`-like way, its `toString()` will automatically be called, and the `string` result of that call will be used instead.

**Note:** The way an object is coerced to a `string` technically goes through the `ToPrimitive` abstract operation (ES5 spec, section 9.1), but those nuanced details are covered in more detail in the `ToNumber` section later in this chapter, so we will skip over them here.

Arrays have an overridden default `toString()` that stringifies as the (string) concatenation of all its values (each stringified themselves), with `", "` in between each value:

```
var a = [1,2,3];

a.toString(); // "1,2,3"
```

Again, `toString()` can either be called explicitly, or it will automatically be called if a non-`string` is used in a `string` context.

## JSON Stringification

Another task that seems awfully related to `ToString` is when you use the `JSON.stringify(..)` utility to serialize a value to a JSON-compatible `string` value.

It's important to note that this stringification is not exactly the same thing as coercion. But since it's related to the `ToString` rules above, we'll take a slight diversion to cover JSON stringification behaviors here.

For most simple values, JSON stringification behaves basically the same as `toString()` conversions, except that the serialization result is *always a string*:

```
JSON.stringify( 42 );      // "42"
JSON.stringify( "42" );    // ""42"" (a string with a quoted string value in it)
JSON.stringify( null );   // "null"
JSON.stringify( true );   // "true"
```

Any *JSON-safe* value can be stringified by `JSON.stringify(..)`. But what is *JSON-safe*? Any value that can be represented validly in a JSON representation.

It may be easier to consider values that are **not** JSON-safe. Some examples: `undefined`s, `function`s, (ES6+) `symbol`s, and `object`s with circular references (where property references in an object structure create a never-ending cycle through each other). These are all illegal values for a standard JSON structure, mostly because they aren't portable to other languages that consume JSON values.

The `JSON.stringify(..)` utility will automatically omit `undefined`, `function`, and `symbol` values when it comes across them. If such a value is found in an `array`, that value is replaced by `null` (so that the array position information isn't altered). If found as a property of an `object`, that property will simply be excluded.

Consider:

```
JSON.stringify( undefined );           // undefined
JSON.stringify( function(){ } );       // undefined

JSON.stringify( [1,undefined,function(){},4] );   // "[1,null,null,4]"
JSON.stringify( { a:2, b:function(){ } } );        // {"a":2}
```

But if you try to `JSON.stringify(..)` an `object` with circular reference(s) in it, an error will be thrown.

JSON stringification has the special behavior that if an `object` value has a `toJSON()` method defined, this method will be called first to get a value to use for serialization.

If you intend to JSON stringify an object that may contain illegal JSON value(s), or if you just have values in the `object` that aren't appropriate for the serialization, you should define a `toJSON()` method for it that returns a *JSON-safe* version of the `object`.

For example:

```
var o = { };

var a = {
  b: 42,
  c: o,
  d: function(){}
};

// create a circular reference inside `a`
o.e = a;

// would throw an error on the circular reference
// JSON.stringify( a );

// define a custom JSON value serialization
a.toJSON = function() {
  // only include the `b` property for serialization
  return { b: this.b };
};

JSON.stringify( a ); // {"b":42}
```

It's a very common misconception that `toJSON()` should return a JSON stringification representation. That's probably incorrect, unless you're wanting to actually stringify the `string` itself (usually not!). `toJSON()` should return the actual regular value (of whatever type) that's appropriate, and `JSON.stringify(..)` itself will handle the stringification.

In other words, `toJSON()` should be interpreted as "to a JSON-safe value suitable for stringification," not "to a JSON string" as many developers mistakenly assume.

Consider:

```
var a = {
  val: [1,2,3],

  // probably correct!
  toJSON: function(){
    return this.val.slice( 1 );
  }
};
```

```

var b = {
  val: [1,2,3],

  // probably incorrect!
  toJSON: function(){
    return "[" +
      this.val.slice( 1 ).join() +
    "]";
  }
};

JSON.stringify( a ); // "[2,3]"

JSON.stringify( b ); // ""[2,3]"""

```

In the second call, we stringified the returned `string` rather than the `array` itself, which was probably not what we wanted to do.

While we're talking about `JSON.stringify(..)`, let's discuss some lesser-known functionalities that can still be very useful.

An optional second argument can be passed to `JSON.stringify(..)` that is called *replacer*. This argument can either be an `array` or a `function`. It's used to customize the recursive serialization of an `object` by providing a filtering mechanism for which properties should and should not be included, in a similar way to how `toJSON()` can prepare a value for serialization.

If *replacer* is an `array`, it should be an `array` of `string`s, each of which will specify a property name that is allowed to be included in the serialization of the `object`. If a property exists that isn't in this list, it will be skipped.

If *replacer* is a `function`, it will be called once for the `object` itself, and then once for each property in the `object`, and each time is passed two arguments, *key* and *value*. To skip a *key* in the serialization, return `undefined`. Otherwise, return the *value* provided.

```

var a = {
  b: 42,
  c: "42",
  d: [1,2,3]
};

JSON.stringify( a, ["b","c"] ); // '{"b":42,"c":"42"}'

JSON.stringify( a, function(k,v){
  if (k !== "c") return v;
} );
// {"b":42,"d":[1,2,3]}

```

**Note:** In the `function replacer` case, the key argument `k` is `undefined` for the first call (where the `a` object itself is being passed in). The `if` statement **filters out** the property named `"c"`. Stringification is recursive, so the `[1,2,3]` array has each of its values (`1`, `2`, and `3`) passed as `v` to *replacer*, with indexes (`0`, `1`, and `2`) as `k`.

A third optional argument can also be passed to `JSON.stringify(..)`, called *space*, which is used as indentation for prettier human-friendly output. *space* can be a positive integer to indicate how many space characters should be used at each indentation level. Or, *space* can be a `string`, in which case up to the first ten characters of its value will be used for each indentation level.

```

var a = {
  b: 42,
  c: "42",
  d: [1,2,3]
};

JSON.stringify( a, null, 3 );
// "{\n  \"b\": 42,

```

```
//   "c": "42",
//   "d": [
//     1,
//     2,
//     3
//   ]
// }"

JSON.stringify( a, null, "----" );
//{
// ----"b": 42,
// ----"c": "42",
// ----"d": [
// -----1,
// -----2,
// -----3
// ----]
// }"
```

Remember, `JSON.stringify(..)` is not directly a form of coercion. We covered it here, however, for two reasons that relate its behavior to `ToString` coercion:

1. `string`, `number`, `boolean`, and `null` values all `stringify` for JSON basically the same as how they coerce to `string` values via the rules of the `ToString` abstract operation.
2. If you pass an `object` value to `JSON.stringify(..)`, and that `object` has a `toJSON()` method on it, `toJSON()` is automatically called to (sort of) "coerce" the value to be *JSON-safe* before stringification.

## ToNumber

If any non-`number` value is used in a way that requires it to be a `number`, such as a mathematical operation, the ES5 spec defines the `ToNumber` abstract operation in section 9.3.

For example, `true` becomes `1` and `false` becomes `0`. `undefined` becomes `Nan`, but (curiously) `null` becomes `0`.

`ToNumber` for a `string` value essentially works for the most part like the rules/syntax for numeric literals (see Chapter 3). If it fails, the result is `Nan` (instead of a syntax error as with `number` literals). One example difference is that `0`-prefixed octal numbers are not handled as octals (just as normal base-10 decimals) in this operation, though such octals are valid as `number` literals (see Chapter 2).

**Note:** The differences between `number` literal grammar and `ToNumber` on a `string` value are subtle and highly nuanced, and thus will not be covered further here. Consult section 9.3.1 of the ES5 spec for more information.

Objects (and arrays) will first be converted to their primitive value equivalent, and the resulting value (if a primitive but not already a `number`) is coerced to a `number` according to the `ToNumber` rules just mentioned.

To convert to this primitive value equivalent, the `ToPrimitive` abstract operation (ES5 spec, section 9.1) will consult the value (using the internal `DefaultValue` operation -- ES5 spec, section 8.12.8) in question to see if it has a `valueOf()` method. If `valueOf()` is available and it returns a primitive value, *that* value is used for the coercion. If not, but `toString()` is available, it will provide the value for the coercion.

If neither operation can provide a primitive value, a `TypeError` is thrown.

As of ES5, you can create such a noncoercible object -- one without `valueOf()` and `toString()` -- if it has a `null` value for its `[[Prototype]]`, typically created with `Object.create(null)`. See the *this & Object Prototypes* title of this series for more information on `[[Prototype]]`s.

**Note:** We cover how to coerce to `number`s later in this chapter in detail, but for this next code snippet, just assume the `Number(..)` function does so.

Consider:

```

var a = {
  valueOf: function(){
    return "42";
  }
};

var b = {
  toString: function(){
    return "42";
  }
};

var c = [4,2];
c.toString = function(){
  return this.join( "" ); // "42"
};

Number( a ); // 42
Number( b ); // 42
Number( c ); // 42
Number( "" ); // 0
Number( [] ); // 0
Number( [ "abc" ] ); // NaN

```

## ToBoolean

Next, let's have a little chat about how `boolean`s behave in JS. There's **lots of confusion and misconception** floating out there around this topic, so pay close attention!

First and foremost, JS has actual keywords `true` and `false`, and they behave exactly as you'd expect of `boolean` values. It's a common misconception that the values `1` and `0` are identical to `true` / `false`. While that may be true in other languages, in JS the `number`s are `number`s and the `boolean`s are `boolean`s. You can coerce `1` to `true` (and vice versa) or `0` to `false` (and vice versa). But they're not the same.

## Falsy Values

But that's not the end of the story. We need to discuss how values other than the two `boolean`s behave whenever you coerce to their `boolean` equivalent.

All of JavaScript's values can be divided into two categories:

1. values that will become `false` if coerced to `boolean`
2. everything else (which will obviously become `true`)

I'm not just being facetious. The JS spec defines a specific, narrow list of values that will coerce to `false` when coerced to a `boolean` value.

How do we know what the list of values is? In the ES5 spec, section 9.2 defines a `toBoolean` abstract operation, which says exactly what happens for all the possible values when you try to coerce them "to boolean."

From that table, we get the following as the so-called "falsy" values list:

- `undefined`
- `null`
- `false`
- `+0`, `-0`, and `NaN`
- `""`

That's it. If a value is on that list, it's a "falsy" value, and it will coerce to `false` if you force a `boolean` coercion on it.

By logical conclusion, if a value is *not* on that list, it must be on *another list*, which we call the "truthy" values list. But JS doesn't really define a "truthy" list per se. It gives some examples, such as saying explicitly that all objects are truthy, but mostly the spec just implies: **anything not explicitly on the falsy list is therefore truthy.**

## Falsy Objects

Wait a minute, that section title even sounds contradictory. I literally *just said* the spec calls all objects truthy, right? There should be no such thing as a "falsy object."

What could that possibly even mean?

You might be tempted to think it means an object wrapper (see Chapter 3) around a falsy value (such as `""`, `0` or `false`). But don't fall into that *trap*.

**Note:** That's a subtle specification joke some of you may get.

Consider:

```
var a = new Boolean( false );
var b = new Number( 0 );
var c = new String( "" );
```

We know all three values here are objects (see Chapter 3) wrapped around obviously falsy values. But do these objects behave as `true` or as `false`? That's easy to answer:

```
var d = Boolean( a && b && c );

d; // true
```

So, all three behave as `true`, as that's the only way `d` could end up as `true`.

**Tip:** Notice the `Boolean( ... )` wrapped around the `a && b && c` expression -- you might wonder why that's there. We'll come back to that later in this chapter, so make a mental note of it. For a sneak-peek (trivia-wise), try for yourself what `d` will be if you just do `d = a && b && c` without the `Boolean( ... )` call!

So, if "falsy objects" are **not just objects wrapped around falsy values**, what the heck are they?

The tricky part is that they can show up in your JS program, but they're not actually part of JavaScript itself.

### What!?

There are certain cases where browsers have created their own sort of *exotic* values behavior, namely this idea of "falsy objects," on top of regular JS semantics.

A "falsy object" is a value that looks and acts like a normal object (properties, etc.), but when you coerce it to a `boolean`, it coerces to a `false` value.

### Why!?

The most well-known case is `document.all`: an array-like (object) provided to your JS program by the DOM (not the JS engine itself), which exposes elements in your page to your JS program. It *used* to behave like a normal object--it would act truthy. But not anymore.

`document.all` itself was never really "standard" and has long since been deprecated/abandoned.

"Can't they just remove it, then?" Sorry, nice try. Wish they could. But there's far too many legacy JS code bases out there that rely on using it.

So, why make it act falsy? Because coercions of `document.all` to `boolean` (like in `if` statements) were almost always used as a means of detecting old, nonstandard IE.

IE has long since come up to standards compliance, and in many cases is pushing the web forward as much or more than any other browser. But all that old `if (document.all) { /* it's IE */ }` code is still out there, and much of it is probably never going away. All this legacy code is still assuming it's running in decade-old IE, which just leads to bad browsing experience for IE users.

So, we can't remove `document.all` completely, but IE doesn't want `if (document.all) { ... }` code to work anymore, so that users in modern IE get new, standards-compliant code logic.

"What should we do?" \*\*"I've got it! Let's bastardize the JS type system and pretend that `document.all` is falsy!"

Ugh. That sucks. It's a crazy gotcha that most JS developers don't understand. But the alternative (doing nothing about the above no-win problems) sucks *just a little bit more*.

So... that's what we've got: crazy, nonstandard "falsy objects" added to JavaScript by the browsers. Yay!

## Truthy Values

Back to the truthy list. What exactly are the truthy values? Remember: **a value is truthy if it's not on the falsy list**.

Consider:

```
var a = "false";
var b = "0";
var c = "!";
var d = Boolean( a && b && c );
d;
```

What value do you expect `d` to have here? It's gotta be either `true` or `false`.

It's `true`. Why? Because despite the contents of those `string` values looking like falsy values, the `string` values themselves are all truthy, because `""` is the only `string` value on the falsy list.

What about these?

```
var a = [];           // empty array -- truthy or falsy?
var b = {};          // empty object -- truthy or falsy?
var c = function(){}; // empty function -- truthy or falsy?

var d = Boolean( a && b && c );
d;
```

Yep, you guessed it, `d` is still `true` here. Why? Same reason as before. Despite what it may seem like, `[]`, `{}`, and `function(){}()` are *not* on the falsy list, and thus are truthy values.

In other words, the truthy list is infinitely long. It's impossible to make such a list. You can only make a finite falsy list and consult *it*.

Take five minutes, write the falsy list on a post-it note for your computer monitor, or memorize it if you prefer. Either way, you'll easily be able to construct a virtual truthy list whenever you need it by simply asking if it's on the falsy list or not.

The importance of truthy and falsy is in understanding how a value will behave if you coerce it (either explicitly or implicitly) to a `boolean` value. Now that you have those two lists in mind, we can dive into coercion examples themselves.

## Explicit Coercion

*Explicit* coercion refers to type conversions that are obvious and explicit. There's a wide range of type conversion usage that clearly falls under the *explicit* coercion category for most developers.

The goal here is to identify patterns in our code where we can make it clear and obvious that we're converting a value from one type to another, so as to not leave potholes for future developers to trip into. The more explicit we are, the more likely someone later will be able to read our code and understand without undue effort what our intent was.

It would be hard to find any salient disagreements with *explicit* coercion, as it most closely aligns with how the commonly accepted practice of type conversion works in statically typed languages. As such, we'll take for granted (for now) that *explicit* coercion can be agreed upon to not be evil or controversial. We'll revisit this later, though.

### Explicitly: Strings <--> Numbers

We'll start with the simplest and perhaps most common coercion operation: coercing values between `string` and `number` representation.

To coerce between `string`s and `number`s, we use the built-in `String(..)` and `Number(..)` functions (which we referred to as "native constructors" in Chapter 3), but **very importantly**, we do not use the `new` keyword in front of them. As such, we're not creating object wrappers.

Instead, we're actually *explicitly coercing* between the two types:

```
var a = 42;
var b = String(a);

var c = "3.14";
var d = Number(c);

b; // "42"
d; // 3.14
```

`String(..)` coerces from any other value to a primitive `string` value, using the rules of the `ToString` operation discussed earlier. `Number(..)` coerces from any other value to a primitive `number` value, using the rules of the `ToNumber` operation discussed earlier.

I call this *explicit* coercion because in general, it's pretty obvious to most developers that the end result of these operations is the applicable type conversion.

In fact, this usage actually looks a lot like it does in some other statically typed languages.

For example, in C/C++, you can say either `(int)x` or `int(x)`, and both will convert the value in `x` to an integer. Both forms are valid, but many prefer the latter, which kinda looks like a function call. In JavaScript, when you say `Number(x)`, it looks awfully similar. Does it matter that it's *actually* a function call in JS? Not really.

Besides `String(..)` and `Number(..)`, there are other ways to "explicitly" convert these values between `string` and `number`:

```
var a = 42;
var b = a.toString();

var c = "3.14";
var d = +c;

b; // "42"
d; // 3.14
```

Calling `a.toString()` is ostensibly explicit (pretty clear that "toString" means "to a string"), but there's some hidden implicitness here. `toString()` cannot be called on a *primitive* value like `42`. So JS automatically "boxes" (see Chapter 3) `42` in an object wrapper, so that `toString()` can be called against the object. In other words, you might call it "explicitly implicit."

`+c` here is showing the *unary operator* form (operator with only one operand) of the `+` operator. Instead of performing mathematic addition (or string concatenation -- see below), the unary `+` explicitly coerces its operand (`c`) to a `number` value.

Is `+c` *explicit* coercion? Depends on your experience and perspective. If you know (which you do, now!) that unary `+` is explicitly intended for `number` coercion, then it's pretty explicit and obvious. However, if you've never seen it before, it can seem awfully confusing, implicit, with hidden side effects, etc.

**Note:** The generally accepted perspective in the open-source JS community is that unary `+` is an accepted form of *explicit* coercion.

Even if you really like the `+c` form, there are definitely places where it can look awfully confusing. Consider:

```
var c = "3.14";
var d = 5 + +c;

d; // 8.14
```

The unary `-` operator also coerces like `+` does, but it also flips the sign of the number. However, you cannot put two `--` next to each other to unflip the sign, as that's parsed as the decrement operator. Instead, you would need to do: `--"3.14"` with a space in between, and that would result in coercion to `3.14`.

You can probably dream up all sorts of hideous combinations of binary operators (like `+` for addition) next to the unary form of an operator. Here's another crazy example:

```
1 + - + + + - + 1; // 2
```

You should strongly consider avoiding unary `+` (or `-`) coercion when it's immediately adjacent to other operators. While the above works, it would almost universally be considered a bad idea. Even `d = +c` (or `d += c` for that matter!) can far too easily be confused for `d += c`, which is entirely different!

**Note:** Another extremely confusing place for unary `+` to be used adjacent to another operator would be the `++` increment operator and `--` decrement operator. For example: `a +++b`, `a + ++b`, and `a + + +b`. See "Expression Side-Effects" in Chapter 5 for more about `++`.

Remember, we're trying to be explicit and **reduce** confusion, not make it much worse!

## Date To number

Another common usage of the unary `+` operator is to coerce a `Date` object into a `number`, because the result is the unix timestamp (milliseconds elapsed since 1 January 1970 00:00:00 UTC) representation of the date/time value:

```
var d = new Date("Mon, 18 Aug 2014 08:53:06 CDT");
+d; // 1408369986000
```

The most common usage of this idiom is to get the current *now* moment as a timestamp, such as:

```
var timestamp = +new Date();
```

**Note:** Some developers are aware of a peculiar syntactic "trick" in JavaScript, which is that the `()` set on a constructor call (a function called with `new`) is *optional* if there are no arguments to pass. So you may run across the `var timestamp = +new Date();` form. However, not all developers agree that omitting the `()` improves readability, as it's an uncommon syntax exception that only applies to the `new fn()` call form and not the regular `fn()` call form.

But coercion is not the only way to get the timestamp out of a `Date` object. A noncoercion approach is perhaps even preferable, as it's even more explicit:

```
var timestamp = new Date().getTime();
// var timestamp = (new Date()).getTime();
// var timestamp = (new Date).getTime();
```

But an *even more* preferable noncoercion option is to use the ES5 added `Date.now()` static function:

```
var timestamp = Date.now();
```

And if you want to polyfill `Date.now()` into older browsers, it's pretty simple:

```
if (!Date.now) {
  Date.now = function() {
    return +new Date();
  };
}
```

I'd recommend skipping the coercion forms related to dates. Use `Date.now()` for current *now* timestamps, and `new Date(..).getTime()` for getting a timestamp of a specific *non-now* date/time that you need to specify.

## The Curious Case of the `~`

One coercive JS operator that is often overlooked and usually very confused is the tilde `~` operator (aka "bitwise NOT"). Many of those who even understand what it does will often times still want to avoid it. But sticking to the spirit of our approach in this book and series, let's dig into it to find out if `~` has anything useful to give us.

In the "32-bit (Signed) Integers" section of Chapter 2, we covered how bitwise operators in JS are defined only for 32-bit operations, which means they force their operands to conform to 32-bit value representations. The rules for how this happens are controlled by the `ToInt32` abstract operation (ES5 spec, section 9.5).

`ToInt32` first does a `ToNumber` coercion, which means if the value is `"123"`, it's going to first become `123` before the `ToInt32` rules are applied.

While not *technically* coercion itself (since the type doesn't change!), using bitwise operators (like `|` or `~`) with certain special `number` values produces a coercive effect that results in a different `number` value.

For example, let's first consider the `|` "bitwise OR" operator used in the otherwise no-op idiom `0 | x`, which (as Chapter 2 showed) essentially only does the `ToInt32` conversion:

```
0 | -0;           // 0
0 | NaN;          // 0
0 | Infinity;    // 0
0 | -Infinity;   // 0
```

These special numbers aren't 32-bit representable (since they come from the 64-bit IEEE 754 standard -- see Chapter 2), so `ToInt32` just specifies `0` as the result from these values.

It's debatable if `0 | __` is an *explicit* form of this coercive `ToInt32` operation or if it's more *implicit*. From the spec perspective, it's unquestionably *explicit*, but if you don't understand bitwise operations at this level, it can seem a bit more *implicitly* magical. Nevertheless, consistent with other assertions in this chapter, we will call it *explicit*.

So, let's turn our attention back to `~`. The `~` operator first "coerces" to a 32-bit `number` value, and then performs a bitwise negation (flipping each bit's parity).

**Note:** This is very similar to how `!` not only coerces its value to `boolean` but also flips its parity (see discussion of the "unary `!`" later).

But... what!? Why do we care about bits being flipped? That's some pretty specialized, nuanced stuff. It's pretty rare for JS developers to need to reason about individual bits.

Another way of thinking about the definition of `~` comes from old-school computer science/discrete Mathematics: `~` performs two's-complement. Great, thanks, that's totally clearer!

Let's try again: `-x` is roughly the same as `-(x+1)`. That's weird, but slightly easier to reason about. So:

```
~42; // -(42+1) ==> -43
```

You're probably still wondering what the heck all this `~` stuff is about, or why it really matters for a coercion discussion. Let's quickly get to the point.

Consider `-(x+1)`. What's the only value that you can perform that operation on that will produce a `0` (or `-0` technically!) result? `-1`. In other words, `~` used with a range of `number` values will produce a falsy (easily coercible to `false`) `0` value for the `-1` input value, and any other truthy `number` otherwise.

Why is that relevant?

`-1` is commonly called a "sentinel value," which basically means a value that's given an arbitrary semantic meaning within the greater set of values of its same type (`number`s). The C-language uses `-1` sentinel values for many functions that return `>= 0` values for "success" and `-1` for "failure."

JavaScript adopted this precedent when defining the `string` operation `indexOf(..)`, which searches for a substring and if found returns its zero-based index position, or `-1` if not found.

It's pretty common to try to use `indexOf(..)` not just as an operation to get the position, but as a `boolean` check of presence/absence of a substring in another `string`. Here's how developers usually perform such checks:

```
var a = "Hello World";

if (a.indexOf( "lo" ) >= 0) {    // true
    // found it!
}
if (a.indexOf( "lo" ) != -1) {    // true
    // found it
}

if (a.indexOf( "ol" ) < 0) {    // true
    // not found!
}
if (a.indexOf( "ol" ) == -1) {    // true
    // not found!
}
```

I find it kind of gross to look at `>= 0` or `== -1`. It's basically a "leaky abstraction," in that it's leaking underlying implementation behavior -- the usage of sentinel `-1` for "failure" -- into my code. I would prefer to hide such a detail.

And now, finally, we see why `~` could help us! Using `~` with `indexOf()` "coerces" (actually just transforms) the value **to be appropriately boolean-coercible**:

```
var a = "Hello World";

~a.indexOf( "lo" );           // -4    <-- truthy!

if (~a.indexOf( "lo" )) {     // true
  // found it!
}

~a.indexOf( "ol" );           // 0    <-- falsy!
!~a.indexOf( "ol" );         // true

if (!~a.indexOf( "ol" )) {    // true
  // not found!
}
```

~ takes the return value of `indexOf(..)` and transforms it: for the "failure" `-1` we get the falsy `0`, and every other value is truthy.

**Note:** The `-(x+1)` pseudo-algorithm for ~ would imply that `~-1` is `-0`, but actually it produces `0` because the underlying operation is actually bitwise, not mathematic.

Technically, `if (~a.indexOf(..))` is still relying on *implicit* coercion of its resultant `0` to `false` or nonzero to `true`. But overall, ~ still feels to me more like an *explicit* coercion mechanism, as long as you know what it's intended to do in this idiom.

I find this to be cleaner code than the previous `>= 0 / == -1` clutter.

### Truncating Bits

There's one more place ~ may show up in code you run across: some developers use the double tilde ~~ to truncate the decimal part of a number (i.e., "coerce" it to a whole number "integer"). It's commonly (though mistakenly) said this is the same result as calling `Math.floor(..)`.

How ~~ works is that the first ~ applies the `ToInt32` "coercion" and does the bitwise flip, and then the second ~ does another bitwise flip, flipping all the bits back to the original state. The end result is just the `ToInt32` "coercion" (aka truncation).

**Note:** The bitwise double-flip of ~~ is very similar to the parity double-negate !!! behavior, explained in the "Explicitly: \* --> Boolean" section later.

However, ~~ needs some caution/clarification. First, it only works reliably on 32-bit values. But more importantly, it doesn't work the same on negative numbers as `Math.floor(..)` does!

```
Math.floor( -49.6 );      // -50
~~-49.6;                  // -49
```

Setting the `Math.floor(..)` difference aside, ~~ can truncate to a (32-bit) integer. But so does `x | 0`, and seemingly with (slightly) less effort.

So, why might you choose ~~ over `x | 0`, then? Operator precedence (see Chapter 5):

```
~~1E20 / 10;             // 166199296
1E20 | 0 / 10;          // 1661992960
(1E20 | 0) / 10;        // 166199296
```

Just as with all other advice here, use ~ and ~~ as explicit mechanisms for "coercion" and value transformation only if everyone who reads/writes such code is properly aware of how these operators work!

## Explicitly: Parsing Numeric Strings

A similar outcome to coercing a `string` to a `number` can be achieved by parsing a `number` out of a `string`'s character contents. There are, however, distinct differences between this parsing and the type conversion we examined above.

Consider:

```
var a = "42";
var b = "42px";

Number( a );      // 42
parseInt( a );    // 42

Number( b );      // NaN
parseInt( b );    // 42
```

Parsing a numeric value out of a string is *tolerant* of non-numeric characters -- it just stops parsing left-to-right when encountered -- whereas coercion is *not tolerant* and fails resulting in the `Nan` value.

Parsing should not be seen as a substitute for coercion. These two tasks, while similar, have different purposes. Parse a `string` as a `number` when you don't know/care what other non-numeric characters there may be on the right-hand side. Coerce a `string` (to a `number`) when the only acceptable values are numeric and something like `"42px"` should be rejected as a `number`.

**Tip:** `parseInt(..)` has a twin, `parseFloat(..)`, which (as it sounds) pulls out a floating-point number from a string.

Don't forget that `parseInt(..)` operates on `string` values. It makes absolutely no sense to pass a `number` value to `parseInt(..)`. Nor would it make sense to pass any other type of value, like `true`, `function(){..}` or `[1,2,3]`.

If you pass a non-`string`, the value you pass will automatically be coerced to a `string` first (see "`ToString`" earlier), which would clearly be a kind of hidden *implicit* coercion. It's a really bad idea to rely upon such a behavior in your program, so never use `parseInt(..)` with a non-`string` value.

Prior to ES5, another gotcha existed with `parseInt(..)`, which was the source of many JS programs' bugs. If you didn't pass a second argument to indicate which numeric base (aka radix) to use for interpreting the numeric `string` contents, `parseInt(..)` would look at the beginning character(s) to make a guess.

If the first two characters were `"0x"` or `"0X"`, the guess (by convention) was that you wanted to interpret the `string` as a hexadecimal (base-16) `number`. Otherwise, if the first character was `"0"`, the guess (again, by convention) was that you wanted to interpret the `string` as an octal (base-8) `number`.

Hexadecimal `string`s (with the leading `0x` or `0X`) aren't terribly easy to get mixed up. But the octal number guessing proved devilishly common. For example:

```
var hour = parseInt( selectedHour.value );
var minute = parseInt( selectedMinute.value );

console.log( "The time you selected was: " + hour + ":" + minute);
```

Seems harmless, right? Try selecting `08` for the hour and `09` for the minute. You'll get `0:0`. Why? because neither `8` nor `9` are valid characters in octal base-8.

The pre-ES5 fix was simple, but so easy to forget: **always pass `10` as the second argument**. This was totally safe:

```
var hour = parseInt( selectedHour.value, 10 );
var minute = parseInt( selectedMinute.value, 10 );
```

As of ES5, `parseInt(..)` no longer guesses octal. Unless you say otherwise, it assumes base-10 (or base-16 for `"0x"` prefixes). That's much nicer. Just be careful if your code has to run in pre-ES5 environments, in which case you still need to pass `10` for the radix.

## Parsing Non-Strings

One somewhat infamous example of `parseInt(...)`'s behavior is highlighted in a sarcastic joke post a few years ago, poking fun at this JS behavior:

```
parseInt( 1/0, 19 ); // 18
```

The assumptive (but totally invalid) assertion was, "If I pass in `Infinity`, and parse an integer out of that, I should get `Infinity` back, not 18." Surely, JS must be crazy for this outcome, right?

Though this example is obviously contrived and unreal, let's indulge the madness for a moment and examine whether JS really is that crazy.

First off, the most obvious sin committed here is to pass a non-`string` to `parseInt(...)`. That's a no-no. Do it and you're asking for trouble. But even if you do, JS politely coerces what you pass in into a `string` that it can try to parse.

Some would argue that this is unreasonable behavior, and that `parseInt(...)` should refuse to operate on a non-`string` value. Should it perhaps throw an error? That would be very Java-like, frankly. I shudder at thinking JS should start throwing errors all over the place so that `try..catch` is needed around almost every line.

Should it return `Nan`? Maybe. But... what about:

```
parseInt( new String( "42" ) );
```

Should that fail, too? It's a non-`string` value. If you want that `String` object wrapper to be unboxed to `"42"`, then is it really so unusual for `42` to first become `"42"` so that `42` can be parsed back out?

I would argue that this half-explicit, half-implicit coercion that can occur can often be a very helpful thing. For example:

```
var a = {
  num: 21,
  toString: function() { return String( this.num * 2 ); }
};

parseInt( a ); // 42
```

The fact that `parseInt(...)` forcibly coerces its value to a `string` to perform the parse on is quite sensible. If you pass in garbage, and you get garbage back out, don't blame the trash can -- it just did its job faithfully.

So, if you pass in a value like `Infinity` (the result of `1 / 0` obviously), what sort of `string` representation would make the most sense for its coercion? Only two reasonable choices come to mind: `"Infinity"` and `"∞"`. JS chose `"Infinity"`. I'm glad it did.

I think it's a good thing that **all values** in JS have some sort of default `string` representation, so that they aren't mysterious black boxes that we can't debug and reason about.

Now, what about base-19? Obviously, completely bogus and contrived. No real JS programs use base-19. It's absurd. But again, let's indulge the ridiculousness. In base-19, the valid numeric characters are `0 - 9` and `a - i` (case insensitive).

So, back to our `parseInt( 1/0, 19 )` example. It's essentially `parseInt( "Infinity", 19 )`. How does it parse? The first character is `"I"`, which is value `18` in the silly base-19. The second character `"n"` is not in the valid set of numeric characters, and as such the parsing simply politely stops, just like when it ran across `"p"` in `"42px"`.

The result? `18`. Exactly like it sensibly should be. The behaviors involved to get us there, and not to an error or to `Infinity` itself, are **very important** to JS, and should not be so easily discarded.

Other examples of this behavior with `parseInt(...)` that may be surprising but are quite sensible include:

```

parseInt( 0.000008 );           // 0   ("0" from "0.000008")
parseInt( 0.000008 );           // 8   ("8" from "8e-7")
parseInt( false, 16 );          // 250 ("fa" from "false")
parseInt( parseInt, 16 );        // 15  ("f" from "function..")

parseInt( "0x10" );            // 16
parseInt( "103", 2 );          // 2

```

`parseInt(..)` is actually pretty predictable and consistent in its behavior. If you use it correctly, you'll get sensible results. If you use it incorrectly, the crazy results you get are not the fault of JavaScript.

## Explicitly: \* --> Boolean

Now, let's examine coercing from any non-`boolean` value to a `boolean`.

Just like with `String(..)` and `Number(..)` above, `Boolean(..)` (without the `new`, of course!) is an explicit way of forcing the `ToBoolean` coercion:

```

var a = "0";
var b = [];
var c = {};

var d = "";
var e = 0;
var f = null;
var g;

Boolean( a ); // true
Boolean( b ); // true
Boolean( c ); // true

Boolean( d ); // false
Boolean( e ); // false
Boolean( f ); // false
Boolean( g ); // false

```

While `Boolean(..)` is clearly explicit, it's not at all common or idiomatic.

Just like the unary `+` operator coerces a value to a `number` (see above), the unary `!` negate operator explicitly coerces a value to a `boolean`. The *problem* is that it also flips the value from truthy to falsy or vice versa. So, the most common way JS developers explicitly coerce to `boolean` is to use the `!!` double-negate operator, because the second `!` will flip the parity back to the original:

```

var a = "0";
var b = [];
var c = {};

var d = "";
var e = 0;
var f = null;
var g;

!!a;    // true
!!b;    // true
!!c;    // true

!!d;    // false
!!e;    // false
!!f;    // false
!!g;    // false

```

Any of these `ToBoolean` coercions would happen *implicitly* without the `Boolean(..)` or `!!`, if used in a `boolean` context such as an `if (...) ..` statement. But the goal here is to explicitly force the value to a `boolean` to make it clearer that the `ToBoolean` coercion is intended.

Another example use-case for explicit `ToBoolean` coercion is if you want to force a `true` / `false` value coercion in the JSON serialization of a data structure:

```
var a = [
  1,
  function(){ /*...*/ },
  2,
  function(){ /*...*/ }
];

JSON.stringify( a ); // "[1,null,2,null]"

JSON.stringify( a, function(key,val){
  if (typeof val == "function") {
    // force `ToBoolean` coercion of the function
    return !!val;
  }
  else {
    return val;
  }
} );
// "[1,true,2,true]"
```

If you come to JavaScript from Java, you may recognize this idiom:

```
var a = 42;

var b = a ? true : false;
```

The `? :` ternary operator will test `a` for truthiness, and based on that test will either assign `true` or `false` to `b`, accordingly.

On its surface, this idiom looks like a form of *explicit* `ToBoolean`-type coercion, since it's obvious that only either `true` or `false` come out of the operation.

However, there's a hidden *implicit* coercion, in that the `a` expression has to first be coerced to `boolean` to perform the truthiness test. I'd call this idiom "explicitly implicit." Furthermore, I'd suggest **you should avoid this idiom completely** in JavaScript. It offers no real benefit, and worse, masquerades as something it's not.

`Boolean(a)` and `!!a` are far better as *explicit* coercion options.

## Implicit Coercion

*Implicit* coercion refers to type conversions that are hidden, with non-obvious side-effects that implicitly occur from other actions. In other words, *implicit coercions* are any type conversions that aren't obvious (to you).

While it's clear what the goal of *explicit* coercion is (making code explicit and more understandable), it might be *too* obvious that *implicit* coercion has the opposite goal: making code harder to understand.

Taken at face value, I believe that's where much of the ire towards coercion comes from. The majority of complaints about "JavaScript coercion" are actually aimed (whether they realize it or not) at *implicit* coercion.

**Note:** Douglas Crockford, author of "*JavaScript: The Good Parts*", has claimed in many conference talks and writings that JavaScript coercion should be avoided. But what he seems to mean is that *implicit* coercion is bad (in his opinion). However, if you read his own code, you'll find plenty of examples of coercion, both *implicit* and *explicit!* In truth, his angst seems to primarily

be directed at the `==` operation, but as you'll see in this chapter, that's only part of the coercion mechanism.

So, is **implicit coercion** evil? Is it dangerous? Is it a flaw in JavaScript's design? Should we avoid it at all costs?

I bet most of you readers are inclined to enthusiastically cheer, "Yes!"

**Not so fast.** Hear me out.

Let's take a different perspective on what *implicit coercion* is, and can be, than just that it's "the opposite of the good explicit kind of coercion." That's far too narrow and misses an important nuance.

Let's define the goal of *implicit coercion* as: to reduce verbosity, boilerplate, and/or unnecessary implementation detail that clutters up our code with noise that distracts from the more important intent.

## Simplifying Implicitly

Before we even get to JavaScript, let me suggest something pseudo-code-ish from some theoretical strongly typed language to illustrate:

```
SomeType x = SomeType( AnotherType( y ) )
```

In this example, I have some arbitrary type of value in `y` that I want to convert to the `SomeType` type. The problem is, this language can't go directly from whatever `y` currently is to `SomeType`. It needs an intermediate step, where it first converts to `AnotherType`, and then from `AnotherType` to `SomeType`.

Now, what if that language (or definition you could create yourself with the language) *did* just let you say:

```
SomeType x = SomeType( y )
```

Wouldn't you generally agree that we simplified the type conversion here to reduce the unnecessary "noise" of the intermediate conversion step? I mean, is it *really* all that important, right here at this point in the code, to see and deal with the fact that `y` goes to `AnotherType` first before then going to `SomeType`?

Some would argue, at least in some circumstances, yes. But I think an equal argument can be made of many other circumstances that here, the simplification **actually aids in the readability of the code** by abstracting or hiding away such details, either in the language itself or in our own abstractions.

Undoubtedly, behind the scenes, somewhere, the intermediate conversion step is still happening. But if that detail is hidden from view here, we can just reason about getting `y` to type `SomeType` as a generic operation and hide the messy details.

While not a perfect analogy, what I'm going to argue throughout the rest of this chapter is that JS *implicit coercion* can be thought of as providing a similar aid to your code.

But, **and this is very important**, that is not an unbounded, absolute statement. There are definitely plenty of *evils* lurking around *implicit coercion*, that will harm your code much more than any potential readability improvements. Clearly, we have to learn how to avoid such constructs so we don't poison our code with all manner of bugs.

Many developers believe that if a mechanism can do some useful thing **A** but can also be abused or misused to do some awful thing **Z**, then we should throw out that mechanism altogether, just to be safe.

My encouragement to you is: don't settle for that. Don't "throw the baby out with the bathwater." Don't assume *implicit coercion* is all bad because all you think you've ever seen is its "bad parts." I think there are "good parts" here, and I want to help and inspire more of you to find and embrace them!

## Implicitly: Strings <--> Numbers

Earlier in this chapter, we explored *explicitly* coercing between `string` and `number` values. Now, let's explore the same task but with *implicit* coercion approaches. But before we do, we have to examine some nuances of operations that will *implicitly* force coercion.

The `+` operator is overloaded to serve the purposes of both `number` addition and `string` concatenation. So how does JS know which type of operation you want to use? Consider:

```
var a = "42";
var b = "0";

var c = 42;
var d = 0;

a + b; // "420"
c + d; // 42
```

What's different that causes `"420"` vs `42`? It's a common misconception that the difference is whether one or both of the operands is a `string`, as that means `+` will assume `string` concatenation. While that's partially true, it's more complicated than that.

Consider:

```
var a = [1,2];
var b = [3,4];

a + b; // "1,23,4"
```

Neither of these operands is a `string`, but clearly they were both coerced to `string`s and then the `string` concatenation kicked in. So what's really going on?

(**Warning:** deeply nitty gritty spec-speak coming, so skip the next two paragraphs if that intimidates you!)

---

According to ES5 spec section 11.6.1, the `+` algorithm (when an `object` value is an operand) will concatenate if either operand is either already a `string`, or if the following steps produce a `string` representation. So, when `+` receives an `object` (including `array`) for either operand, it first calls the `ToPrimitive` abstract operation (section 9.1) on the value, which then calls the `[[DefaultValue]]` algorithm (section 8.12.8) with a context hint of `number`.

If you're paying close attention, you'll notice that this operation is now identical to how the `ToNumber` abstract operation handles `object`s (see the "`ToNumber`" section earlier). The `valueOf()` operation on the `array` will fail to produce a simple primitive, so it then falls to a `toString()` representation. The two `array`s thus become `"1,2"` and `"3,4"`, respectively. Now, `+` concatenates the two `string`s as you'd normally expect: `"1,23,4"`.

---

Let's set aside those messy details and go back to an earlier, simplified explanation: if either operand to `+` is a `string` (or becomes one with the above steps!), the operation will be `string` concatenation. Otherwise, it's always numeric addition.

**Note:** A commonly cited coercion gotcha is `[] + {}` vs. `{ } + []`, as those two expressions result, respectively, in `"[object Object]"` and `0`. There's more to it, though, and we cover those details in "Blocks" in Chapter 5.

What's that mean for *implicit* coercion?

You can coerce a `number` to a `string` simply by "adding" the `number` and the `""` empty `string`:

```
var a = 42;
var b = a + "";
```

```
b; // "42"
```

**Tip:** Numeric addition with the `+` operator is commutative, which means `2 + 3` is the same as `3 + 2`. String concatenation with `+` is obviously not generally commutative, **but** with the specific case of `""`, it's effectively commutative, as `a + ""` and `"" + a` will produce the same result.

It's extremely common/idiomatic to (*implicitly*) coerce `number` to `string` with a `+ ""` operation. In fact, interestingly, even some of the most vocal critics of *implicit* coercion still use that approach in their own code, instead of one of its *explicit* alternatives.

I think this is a great example of a useful form in *implicit* coercion, despite how frequently the mechanism gets criticized!

Comparing this *implicit* coercion of `a + ""` to our earlier example of `String(a)` *explicit* coercion, there's one additional quirk to be aware of. Because of how the `ToPrimitive` abstract operation works, `a + ""` invokes `valueOf()` on the `a` value, whose return value is then finally converted to a `string` via the internal `ToString` abstract operation. But `String(a)` just invokes `toString()` directly.

Both approaches ultimately result in a `string`, but if you're using an `object` instead of a regular primitive `number` value, you may not necessarily get the *same* `string` value!

Consider:

```
var a = {
  valueOf: function() { return 42; },
  toString: function() { return 4; }
};

a + "";           // "42"

String(a);      // "4"
```

Generally, this sort of gotcha won't bite you unless you're really trying to create confusing data structures and operations, but you should be careful if you're defining both your own `valueOf()` and `toString()` methods for some `object`, as how you coerce the value could affect the outcome.

What about the other direction? How can we *implicitly coerce* from `string` to `number`?

```
var a = "3.14";
var b = a - 0;

b; // 3.14
```

The `-` operator is defined only for numeric subtraction, so `a - 0` forces `a`'s value to be coerced to a `number`. While far less common, `a * 1` or `a / 1` would accomplish the same result, as those operators are also only defined for numeric operations.

What about `object` values with the `-` operator? Similar story as for `+` above:

```
var a = [3];
var b = [1];

a - b; // 2
```

Both `array` values have to become `number`s, but they end up first being coerced to `strings` (using the expected `toString()` serialization), and then are coerced to `number`s, for the `-` subtraction to perform on.

So, is *implicit* coercion of `string` and `number` values the ugly evil you've always heard horror stories about? I don't personally think so.

Compare `b = String(a)` (*explicit*) to `b = a + ""` (*implicit*). I think cases can be made for both approaches being useful in your code. Certainly `b = a + ""` is quite a bit more common in JS programs, proving its own utility regardless of *feelings* about the merits or hazards of *implicit* coercion in general.

## Implicitly: Booleans --> Numbers

I think a case where *implicit* coercion can really shine is in simplifying certain types of complicated `boolean` logic into simple numeric addition. Of course, this is not a general-purpose technique, but a specific solution for specific cases.

Consider:

```
function onlyOne(a,b,c) {
  return (!((a && !b && !c) ||
            (!a && b && !c) || (!a && !b && c)));
}

var a = true;
var b = false;

onlyOne( a, b, b );      // true
onlyOne( b, a, b );      // true

onlyOne( a, b, a );      // false
```

This `onlyOne(..)` utility should only return `true` if exactly one of the arguments is `true` / *truthy*. It's using *implicit* coercion on the *truthy* checks and *explicit* coercion on the others, including the final return value.

But what if we needed that utility to be able to handle four, five, or twenty flags in the same way? It's pretty difficult to imagine implementing code that would handle all those permutations of comparisons.

But here's where coercing the `boolean` values to `number`'s (`0` or `1`, obviously) can greatly help:

```
function onlyOne() {
  var sum = 0;
  for (var i=0; i < arguments.length; i++) {
    // skip falsy values. same as treating
    // them as 0's, but avoids NaN's.
    if (arguments[i]) {
      sum += arguments[i];
    }
  }
  return sum == 1;
}

var a = true;
var b = false;

onlyOne( b, a );          // true
onlyOne( b, a, b, b, b ); // true

onlyOne( b, b );          // false
onlyOne( b, a, b, b, b, a ); // false
```

**Note:** Of course, instead of the `for` loop in `onlyOne(..)`, you could more tersely use the ES5 `reduce(..)` utility, but I didn't want to obscure the concepts.

What we're doing here is relying on the `1` for `true` /*truthy* coercions, and numerically adding them all up. `sum += arguments[i]` uses *implicit* coercion to make that happen. If one and only one value in the `arguments` list is `true`, then the numeric sum will be `1`, otherwise the sum will not be `1` and thus the desired condition is not met.

We could of course do this with *explicit* coercion instead:

```

function onlyOne() {
  var sum = 0;
  for (var i=0; i < arguments.length; i++) {
    sum += Number( !arguments[i] );
  }
  return sum === 1;
}

```

We first use `!arguments[i]` to force the coercion of the value to `true` or `false`. That's so you could pass non-`boolean` values in, like `onlyOne( "42", 0 )`, and it would still work as expected (otherwise you'd end up with `string` concatenation and the logic would be incorrect).

Once we're sure it's a `boolean`, we do another *explicit* coercion with `Number(..)` to make sure the value is `0` or `1`.

Is the *explicit* coercion form of this utility "better"? It does avoid the `NAN` trap as explained in the code comments. But, ultimately, it depends on your needs. I personally think the former version, relying on *implicit* coercion is more elegant (if you won't be passing `undefined` or `NAN`), and the *explicit* version is needlessly more verbose.

But as with almost everything we're discussing here, it's a judgment call.

**Note:** Regardless of *implicit* or *explicit* approaches, you could easily make `onlyTwo(..)` or `onlyFive(..)` variations by simply changing the final comparison from `1`, to `2` or `5`, respectively. That's drastically easier than adding a bunch of `&&` and `||` expressions. So, generally, coercion is very helpful in this case.

## Implicitly: \* --> Boolean

Now, let's turn our attention to *implicit* coercion to `boolean` values, as it's by far the most common and also by far the most potentially troublesome.

Remember, *implicit* coercion is what kicks in when you use a value in such a way that it forces the value to be converted. For numeric and `string` operations, it's fairly easy to see how the coercions can occur.

But, what sort of expression operations require/force (*implicitly*) a `boolean` coercion?

1. The test expression in an `if(..)` statement.
2. The test expression (second clause) in a `for(..; ..; ..)` header.
3. The test expression in `while(..)` and `do..while(..)` loops.
4. The test expression (first clause) in `? :` ternary expressions.
5. The left-hand operand (which serves as a test expression -- see below!) to the `||` ("logical or") and `&&` ("logical and") operators.

Any value used in these contexts that is not already a `boolean` will be *implicitly* coerced to a `boolean` using the rules of the `ToBoolean` abstract operation covered earlier in this chapter.

Let's look at some examples:

```

var a = 42;
var b = "abc";
var c;
var d = null;

if (a) {
  console.log( "yep" );           // yep
}

while (c) {
  console.log( "nope, never runs" );
}

c = d ? a : b;

```

```
c;           // "abc"

if ((a && d) || c) {
  console.log( "yep" );
}           // yep
```

In all these contexts, the non-`boolean` values are *implicitly coerced* to their `boolean` equivalents to make the test decisions.

## Operators `||` and `&&`

It's quite likely that you have seen the `||` ("logical or") and `&&` ("logical and") operators in most or all other languages you've used. So it'd be natural to assume that they work basically the same in JavaScript as in other similar languages.

There's some very little known, but very important, nuance here.

In fact, I would argue these operators shouldn't even be called "logical \_ operators", as that name is incomplete in describing what they do. If I were to give them a more accurate (if more clumsy) name, I'd call them "selector operators," or more completely, "operand selector operators."

Why? Because they don't actually result in a *logic* value (aka `boolean`) in JavaScript, as they do in some other languages.

So what *do* they result in? They result in the value of one (and only one) of their two operands. In other words, **they select one of the two operand's values**.

Quoting the ES5 spec from section 11.11:

The value produced by a `&&` or `||` operator is not necessarily of type Boolean. The value produced will always be the value of one of the two operand expressions.

Let's illustrate:

```
var a = 42;
var b = "abc";
var c = null;

a || b;      // 42
a && b;      // "abc"

c || b;      // "abc"
c && b;      // null
```

**Wait, what!?** Think about that. In languages like C and PHP, those expressions result in `true` or `false`, but in JS (and Python and Ruby, for that matter!), the result comes from the values themselves.

Both `||` and `&&` operators perform a `boolean` test on the **first operand** (`a` or `c`). If the operand is not already `boolean` (as it's not, here), a normal `ToBoolean` coercion occurs, so that the test can be performed.

For the `||` operator, if the test is `true`, the `||` expression results in the value of the *first operand* (`a` or `c`). If the test is `false`, the `||` expression results in the value of the *second operand* (`b`).

Inversely, for the `&&` operator, if the test is `true`, the `&&` expression results in the value of the *second operand* (`b`). If the test is `false`, the `&&` expression results in the value of the *first operand* (`a` or `c`).

The result of a `||` or `&&` expression is always the underlying value of one of the operands, **not** the (possibly coerced) result of the test. In `c && b`, `c` is `null`, and thus falsy. But the `&&` expression itself results in `null` (the value in `c`), not in the coerced `false` used in the test.

Do you see how these operators act as "operand selectors", now?

Another way of thinking about these operators:

```
a || b;
// roughly equivalent to:
a ? a : b;

a && b;
// roughly equivalent to:
a ? b : a;
```

**Note:** I call `a || b` "roughly equivalent" to `a ? a : b` because the outcome is identical, but there's a nuanced difference. In `a ? a : b`, if `a` was a more complex expression (like for instance one that might have side effects like calling a `function`, etc.), then the `a` expression would possibly be evaluated twice (if the first evaluation was truthy). By contrast, for `a || b`, the `a` expression is evaluated only once, and that value is used both for the coercive test as well as the result value (if appropriate). The same nuance applies to the `a && b` and `a ? b : a` expressions.

An extremely common and helpful usage of this behavior, which there's a good chance you may have used before and not fully understood, is:

```
function foo(a,b) {
  a = a || "hello";
  b = b || "world";

  console.log( a + " " + b );
}

foo();           // "hello world"
foo( "yeah", "yeah!" ); // "yeah yeah!"
```

The `a = a || "hello"` idiom (sometimes said to be JavaScript's version of the C# "null coalescing operator") acts to test `a` and if it has no value (or only an undesired falsy value), provides a backup default value ("hello").

**Be careful**, though!

```
foo( "That's it!", "" ); // "That's it! world" <-- Oops!
```

See the problem? `""` as the second argument is a falsy value (see `ToBoolean` earlier in this chapter), so the `b = b || "world"` test fails, and the `"world"` default value is substituted, even though the intent probably was to have the explicitly passed `""` be the value assigned to `b`.

This `||` idiom is extremely common, and quite helpful, but you have to use it only in cases where *all falsy values* should be skipped. Otherwise, you'll need to be more explicit in your test, and probably use `a ? : ternary` instead.

This *default value assignment* idiom is so common (and useful!) that even those who publicly and vehemently decry JavaScript coercion often use it in their own code!

What about `&&`?

There's another idiom that is quite a bit less commonly authored manually, but which is used by JS minifiers frequently. The `&&` operator "selects" the second operand if and only if the first operand tests as truthy, and this usage is sometimes called the "guard operator" (also see "Short Circuited" in Chapter 5) -- the first expression test "guards" the second expression:

```
function foo() {
  console.log( a );
}

var a = 42;

a && foo(); // 42
```

`foo()` gets called only because `a` tests as truthy. If that test failed, this `a && foo()` expression statement would just silently stop -- this is known as "short circuiting" -- and never call `foo()`.

Again, it's not nearly as common for people to author such things. Usually, they'd do `if (a) { foo(); }` instead. But JS minifiers choose `a && foo()` because it's much shorter. So, now, if you ever have to decipher such code, you'll know what it's doing and why.

OK, so `||` and `&&` have some neat tricks up their sleeve, as long as you're willing to allow the *implicit* coercion into the mix.

**Note:** Both the `a = b || "something"` and `a && b()` idioms rely on short circuiting behavior, which we cover in more detail in Chapter 5.

The fact that these operators don't actually result in `true` and `false` is possibly messing with your head a little bit by now. You're probably wondering how all your `if` statements and `for` loops have been working, if they've included compound logical expressions like `a && (b || c)`.

Don't worry! The sky is not falling. Your code is (probably) just fine. It's just that you probably never realized before that there was an *implicit* coercion to `boolean` going on **after** the compound expression was evaluated.

Consider:

```
var a = 42;
var b = null;
var c = "foo";

if (a && (b || c)) {
  console.log( "yep" );
}
```

This code still works the way you always thought it did, except for one subtle extra detail. The `a && (b || c)` expression *actually* results in `"foo"`, not `true`. So, the `if` statement *then* forces the `"foo"` value to coerce to a `boolean`, which of course will be `true`.

See? No reason to panic. Your code is probably still safe. But now you know more about how it does what it does.

And now you also realize that such code is using *implicit* coercion. If you're in the "avoid (*implicit*) coercion camp" still, you're going to need to go back and make all of those tests *explicit*:

```
if (!!a && (!!b || !!c)) {
  console.log( "yep" );
}
```

Good luck with that! ... Sorry, just teasing.

## Symbol Coercion

Up to this point, there's been almost no observable outcome difference between *explicit* and *implicit* coercion -- only the readability of code has been at stake.

But ES6 Symbols introduce a gotcha into the coercion system that we need to discuss briefly. For reasons that go well beyond the scope of what we'll discuss in this book, *explicit* coercion of a `symbol` to a `string` is allowed, but *implicit* coercion of the same is disallowed and throws an error.

Consider:

```
var s1 = Symbol( "cool" );
String( s1 ); // "Symbol(cool)"

var s2 = Symbol( "not cool" );
```

```
s2 + ""; // TypeError
```

`symbol` values cannot coerce to `number` at all (throws an error either way), but strangely they can both *explicitly* and *implicitly* coerce to `boolean` (always `true`).

Consistency is always easier to learn, and exceptions are never fun to deal with, but we just need to be careful around the new ES6 `symbol` values and how we coerce them.

The good news: it's probably going to be exceedingly rare for you to need to coerce a `symbol` value. The way they're typically used (see Chapter 3) will probably not call for coercion on a normal basis.

## Loose Equals vs. Strict Equals

Loose equals is the `==` operator, and strict equals is the `===` operator. Both operators are used for comparing two values for "equality," but the "loose" vs. "strict" indicates a **very important** difference in behavior between the two, specifically in how they decide "equality."

A very common misconception about these two operators is: "`==` checks values for equality and `===` checks both values and types for equality." While that sounds nice and reasonable, it's inaccurate. Countless well-respected JavaScript books and blogs have said exactly that, but unfortunately they're all *wrong*.

The correct description is: "`==` allows coercion in the equality comparison and `===` disallows coercion."

## Equality Performance

Stop and think about the difference between the first (inaccurate) explanation and this second (accurate) one.

In the first explanation, it seems obvious that `===` is *doing more work* than `==`, because it has to *also* check the type. In the second explanation, `==` is the one *doing more work* because it has to follow through the steps of coercion if the types are different.

Don't fall into the trap, as many have, of thinking this has anything to do with performance, though, as if `==` is going to be slower than `===` in any relevant way. While it's measurable that coercion does take *a little bit* of processing time, it's mere microseconds (yes, that's millionths of a second!).

If you're comparing two values of the same types, `==` and `===` use the identical algorithm, and so other than minor differences in engine implementation, they should do the same work.

If you're comparing two values of different types, the performance isn't the important factor. What you should be asking yourself is: when comparing these two values, do I want coercion or not?

If you want coercion, use `==` loose equality, but if you don't want coercion, use `===` strict equality.

**Note:** The implication here then is that both `==` and `===` check the types of their operands. The difference is in how they respond if the types don't match.

## Abstract Equality

The `==` operator's behavior is defined as "The Abstract Equality Comparison Algorithm" in section 11.9.3 of the ES5 spec. What's listed there is a comprehensive but simple algorithm that explicitly states every possible combination of types, and how the coercions (if necessary) should happen for each combination.

**Warning:** When (*implicit*) coercion is maligned as being too complicated and too flawed to be a *useful good part*, it is these rules of "abstract equality" that are being condemned. Generally, they are said to be too complex and too unintuitive for developers to practically learn and use, and that they are prone more to causing bugs in JS programs than to enabling greater code readability. I

believe this is a flawed premise -- that you readers are competent developers who write (and read and understand!) algorithms (aka code) all day long. So, what follows is a plain exposition of the "abstract equality" in simple terms. But I implore you to also read the ES5 spec section 11.9.3. I think you'll be surprised at just how reasonable it is.

Basically, the first clause (11.9.3.1) says, if the two values being compared are of the same type, they are simply and naturally compared via Identity as you'd expect. For example, `42` is only equal to `42`, and `"abc"` is only equal to `"abc"`.

Some minor exceptions to normal expectation to be aware of:

- `Nan` is never equal to itself (see Chapter 2)
- `+0` and `-0` are equal to each other (see Chapter 2)

The final provision in clause 11.9.3.1 is for `==` loose equality comparison with `object`s (including `function`s and `array`s). Two such values are only *equal* if they are both references to *the exact same value*. No coercion occurs here.

**Note:** The `===` strict equality comparison is defined identically to 11.9.3.1, including the provision about two `object` values. It's a very little known fact that `==` and `===` behave identically in the case where two `object`s are being compared!

The rest of the algorithm in 11.9.3 specifies that if you use `==` loose equality to compare two values of different types, one or both of the values will need to be *implicitly* coerced. This coercion happens so that both values eventually end up as the same type, which can then directly be compared for equality using simple value Identity.

**Note:** The `!=` loose not-equality operation is defined exactly as you'd expect, in that it's literally the `==` operation comparison performed in its entirety, then the negation of the result. The same goes for the `!==` strict not-equality operation.

## Comparing: `string`s to `number`s

To illustrate `==` coercion, let's first build off the `string` and `number` examples earlier in this chapter:

```
var a = 42;
var b = "42";

a === b;    // false
a == b;     // true
```

As we'd expect, `a === b` fails, because no coercion is allowed, and indeed the `42` and `"42"` values are different.

However, the second comparison `a == b` uses loose equality, which means that if the types happen to be different, the comparison algorithm will perform *implicit* coercion on one or both values.

But exactly what kind of coercion happens here? Does the `a` value of `42` become a `string`, or does the `b` value of `"42"` become a `number`?

In the ES5 spec, clauses 11.9.3.4-5 say:

1. If `Type(x)` is `Number` and `Type(y)` is `String`, return the result of the comparison `x == ToNumber(y)`.
2. If `Type(x)` is `String` and `Type(y)` is `Number`, return the result of the comparison `ToNumber(x) == y`.

**Warning:** The spec uses `Number` and `String` as the formal names for the types, while this book prefers `number` and `string` for the primitive types. Do not let the capitalization of `Number` in the spec confuse you for the `Number()` native function. For our purposes, the capitalization of the type name is irrelevant -- they have basically the same meaning.

Clearly, the spec says the `"42"` value is coerced to a `number` for the comparison. The *how* of that coercion has already been covered earlier, specifically with the `ToNumber` abstract operation. In this case, it's quite obvious then that the resulting two `42` values are equal.

## Comparing: anything to `boolean`

One of the biggest gotchas with the *implicit coercion* of `==` loose equality pops up when you try to compare a value directly to `true` or `false`.

Consider:

```
var a = "42";
var b = true;

a == b; // false
```

Wait, what happened here!? We know that `"42"` is a truthy value (see earlier in this chapter). So, how come it's not `==` loose equal to `true`?

The reason is both simple and deceptively tricky. It's so easy to misunderstand, many JS developers never pay close enough attention to fully grasp it.

Let's again quote the spec, clauses 11.9.3.6-7:

1. If `Type(x)` is Boolean, return the result of the comparison `ToNumber(x) == y`.
2. If `Type(y)` is Boolean, return the result of the comparison `x == ToNumber(y)`.

Let's break that down. First:

```
var x = true;
var y = "42";

x == y; // false
```

The `Type(x)` is indeed `Boolean`, so it performs `ToNumber(x)`, which coerces `true` to `1`. Now, `1 == "42"` is evaluated. The types are still different, so (essentially recursively) we reconsult the algorithm, which just as above will coerce `"42"` to `42`, and `1 == 42` is clearly `false`.

Reverse it, and we still get the same outcome:

```
var x = "42";
var y = false;

x == y; // false
```

The `Type(y)` is `Boolean` this time, so `ToNumber(y)` yields `0`. `"42" == 0` recursively becomes `42 == 0`, which is of course `false`.

In other words, **the value `"42"` is neither `== true` nor `== false`**. At first, that statement might seem crazy. How can a value be neither truthy nor falsy?

But that's the problem! You're asking the wrong question, entirely. It's not your fault, really. Your brain is tricking you.

`"42"` is indeed truthy, but `"42" == true` **is not performing a boolean test/coercion** at all, no matter what your brain says. `"42"` is not being coerced to a `boolean` (`true`), but instead `true` is being coerced to a `1`, and then `"42"` is being coerced to `42`.

Whether we like it or not, `ToBoolean` is not even involved here, so the truthiness or falsiness of `"42"` is irrelevant to the `==` operation!

What is relevant is to understand how the `==` comparison algorithm behaves with all the different type combinations. As it regards a `boolean` value on either side of the `==`, a `boolean` always coerces to a `number` *first*.

If that seems strange to you, you're not alone. I personally would recommend to never, ever, under any circumstances, use `== true` or `== false`. Ever.

But remember, I'm only talking about `==` here. `== true` and `== false` wouldn't allow the coercion, so they're safe from this hidden `ToNumber` coercion.

Consider:

```
var a = "42";

// bad (will fail!):
if (a == true) {
    // ...
}

// also bad (will fail!):
if (a === true) {
    // ...
}

// good enough (works implicitly):
if (a) {
    // ...
}

// better (works explicitly):
if (!!a) {
    // ...
}

// also great (works explicitly):
if (Boolean(a)) {
    // ...
}
```

If you avoid ever using `== true` or `== false` (aka loose equality with `boolean`s) in your code, you'll never have to worry about this truthiness/falsiness mental gotcha.

## Comparing: `null`s to `undefined`s

Another example of *implicit* coercion can be seen with `==` loose equality between `null` and `undefined` values. Yet again quoting the ES5 spec, clauses 11.9.3.2-3:

1. If `x` is `null` and `y` is `undefined`, return `true`.
2. If `x` is `undefined` and `y` is `null`, return `true`.

`null` and `undefined`, when compared with `==` loose equality, equate to (aka coerce to) each other (as well as themselves, obviously), and no other values in the entire language.

What this means is that `null` and `undefined` can be treated as indistinguishable for comparison purposes, if you use the `==` loose equality operator to allow their mutual *implicit* coercion.

```
var a = null;
var b;

a == b;          // true
a == null;       // true
b == null;       // true

a == false;      // false
b == false;      // false
a == "";         // false
b == "";         // false
a == 0;          // false
b == 0;          // false
```

The coercion between `null` and `undefined` is safe and predictable, and no other values can give false positives in such a check. I recommend using this coercion to allow `null` and `undefined` to be indistinguishable and thus treated as the same value.

For example:

```
var a = doSomething();

if (a == null) {
    // ...
}
```

The `a == null` check will pass only if `doSomething()` returns either `null` or `undefined`, and will fail with any other value, even other falsy values like `0`, `false`, and `""`.

The *explicit* form of the check, which disallows any such coercion, is (I think) unnecessarily much uglier (and perhaps a tiny bit less performant!):

```
var a = doSomething();

if (a === undefined || a === null) {
    // ...
}
```

In my opinion, the form `a == null` is yet another example where *implicit* coercion improves code readability, but does so in a reliably safe way.

## Comparing: object s to non-object s

If an `object / function / array` is compared to a simple scalar primitive (`string`, `number`, or `boolean`), the ES5 spec says in clauses 11.9.3.8-9:

1. If `Type(x)` is either `String` or `Number` and `Type(y)` is `Object`, return the result of the comparison `x == ToPrimitive(y)`.
2. If `Type(x)` is `Object` and `Type(y)` is either `String` or `Number`, return the result of the comparison `ToPrimitive(x) == y`.

**Note:** You may notice that these clauses only mention `String` and `Number`, but not `Boolean`. That's because, as quoted earlier, clauses 11.9.3.6-7 take care of coercing any `Boolean` operand presented to a `Number` first.

Consider:

```
var a = 42;
var b = [ 42 ];

a == b;      // true
```

The `[ 42 ]` value has its `ToPrimitive` abstract operation called (see the "Abstract Value Operations" section earlier), which results in the `"42"` value. From there, it's just `42 == "42"`, which as we've already covered becomes `42 == 42`, so `a` and `b` are found to be coercively equal.

**Tip:** All the quirks of the `ToPrimitive` abstract operation that we discussed earlier in this chapter (`toString()`, `valueOf()`) apply here as you'd expect. This can be quite useful if you have a complex data structure that you want to define a custom `valueOf()` method on, to provide a simple value for equality comparison purposes.

In Chapter 3, we covered "unboxing," where an `object` wrapper around a primitive value (like from `new String("abc")`, for instance) is unwrapped, and the underlying primitive value (`"abc"`) is returned. This behavior is related to the `ToPrimitive` coercion in the `==` algorithm:

```
var a = "abc";
var b = Object( a ); // same as `new String( a )`  

  
a === b;           // false
a == b;            // true
```

`a == b` is `true` because `b` is coerced (aka "unboxed," unwrapped) via `ToPrimitive` to its underlying `"abc"` simple scalar primitive value, which is the same as the value in `a`.

There are some values where this is not the case, though, because of other overriding rules in the `==` algorithm. Consider:

```
var a = null;
var b = Object( a ); // same as `Object()`  
a == b;             // false  
  
var c = undefined;
var d = Object( c ); // same as `Object()`  
c == d;             // false  
  
var e = NaN;
var f = Object( e ); // same as `new Number( e )`  
e == f;             // false
```

The `null` and `undefined` values cannot be boxed -- they have no object wrapper equivalent -- so `Object(null)` is just like `Object()` in that both just produce a normal object.

`NaN` can be boxed to its `Number` object wrapper equivalent, but when `==` causes an unboxing, the `NaN == NaN` comparison fails because `NaN` is never equal to itself (see Chapter 2).

## Edge Cases

Now that we've thoroughly examined how the *implicit* coercion of `==` loose equality works (in both sensible and surprising ways), let's try to call out the worst, craziest corner cases so we can see what we need to avoid to not get bitten with coercion bugs.

First, let's examine how modifying the built-in native prototypes can produce crazy results:

### A Number By Any Other Value Would...

```
Number.prototype.valueOf = function() {
  return 3;
};  
  
new Number( 2 ) == 3; // true
```

**Warning:** `2 == 3` would not have fallen into this trap, because neither `2` nor `3` would have invoked the built-in `Number.prototype.valueOf()` method because both are already primitive `number` values and can be compared directly. However, `new Number(2)` must go through the `ToPrimitive` coercion, and thus invoke `valueOf()`.

Evil, huh? Of course it is. No one should ever do such a thing. The fact that you *can* do this is sometimes used as a criticism of coercion and `==`. But that's misdirected frustration. JavaScript is not *bad* because you can do such things, a developer is *bad if they do such things*. Don't fall into the "my programming language should protect me from myself" fallacy.

Next, let's consider another tricky example, which takes the evil from the previous example to another level:

```
if (a == 2 && a == 3) {  
  // ...  
}
```

You might think this would be impossible, because `a` could never be equal to both `2` and `3` *at the same time*. But "at the same time" is inaccurate, since the first expression `a == 2` happens strictly *before* `a == 3`.

So, what if we make `a.valueOf()` have side effects each time it's called, such that the first time it returns `2` and the second time it's called it returns `3`? Pretty easy:

```
var i = 2;

Number.prototype.valueOf = function() {
    return i++;
};

var a = new Number( 42 );

if (a == 2 && a == 3) {
    console.log( "Yep, this happened." );
}
```

Again, these are evil tricks. Don't do them. But also don't use them as complaints against coercion. Potential abuses of a mechanism are not sufficient evidence to condemn the mechanism. Just avoid these crazy tricks, and stick only with valid and proper usage of coercion.

## False-y Comparisons

The most common complaint against *implicit* coercion in `==` comparisons comes from how falsy values behave surprisingly when compared to each other.

To illustrate, let's look at a list of the corner-cases around falsy value comparisons, to see which ones are reasonable and which are troublesome:

```
"0" == null;           // false
"0" == undefined;     // false
"0" == false;          // true -- UH OH!
"0" == NaN;            // false
"0" == 0;              // true
"0" == "";             // false

false == null;          // false
false == undefined;     // false
false == NaN;            // false
false == 0;              // true -- UH OH!
false == "";             // true -- UH OH!
false == [];             // true -- UH OH!
false == {};             // false

"" == null;             // false
"" == undefined;        // false
"" == NaN;               // false
"" == 0;                 // true -- UH OH!
"" == [];                // true -- UH OH!
"" == {};                // false

0 == null;               // false
0 == undefined;          // false
0 == NaN;                 // false
0 == [];                  // true -- UH OH!
0 == {};                  // false
```

In this list of 24 comparisons, 17 of them are quite reasonable and predictable. For example, we know that `""` and `NaN` are not at all equatable values, and indeed they don't coerce to be loose equals, whereas `"0"` and `0` are reasonably equatable and *do* coerce as loose equals.

However, seven of the comparisons are marked with "UH OH!" because as false positives, they are much more likely gotchas that could trip you up. `""` and `0` are definitely distinctly different values, and it's rare you'd want to treat them as equatable, so their mutual coercion is troublesome. Note that there aren't any false negatives here.

## The Crazy Ones

We don't have to stop there, though. We can keep looking for even more troublesome coercions:

```
[] == ![];           // true
```

Oooo, that seems at a higher level of crazy, right?! Your brain may likely trick you that you're comparing a truthy to a falsy value, so the `true` result is surprising, as we *know* a value can never be truthy and falsy at the same time!

But that's not what's actually happening. Let's break it down. What do we know about the `!` unary operator? It explicitly coerces to a `boolean` using the `ToBoolean` rules (and it also flips the parity). So before `[] == ![]` is even processed, it's actually already translated to `[] == false`. We already saw that form in our above list (`false == []`), so its surprise result is *not new* to us.

How about other corner cases?

```
2 == [2];           // true
"" == [null];      // true
```

As we said earlier in our `ToNumber` discussion, the right-hand side `[2]` and `[null]` values will go through a `ToPrimitive` coercion so they can be more readily compared to the simple primitives (`2` and `""`, respectively) on the left-hand side. Since the `valueOf()` for `array` values just returns the `array` itself, coercion fails to stringify the `array`.

`[2]` will become `"2"`, which then is `ToNumber` coerced to `2` for the right-hand side value in the first comparison. `[null]` just straight becomes `""`.

So, `2 == 2` and `"" == ""` are completely understandable.

If your instinct is to still dislike these results, your frustration is not actually with coercion like you probably think it is. It's actually a complaint against the default `array` values' `ToPrimitive` behavior of coercing to a `string` value. More likely, you'd just wish that `[2].toString()` didn't return `"2"`, or that `[null].toString()` didn't return `""`.

But what exactly *should* these `string` coercions result in? I can't really think of any other appropriate `string` coercion of `[2]` than `"2"`, except perhaps `"[2]"` -- but that could be very strange in other contexts!

You could rightly make the case that since `String(null)` becomes `"null"`, then `String([null])` should also become `"null"`. That's a reasonable assertion. So, that's the real culprit.

*Implicit* coercion itself isn't the evil here. Even an *explicit* coercion of `[null]` to a `string` results in `""`. What's at odds is whether it's sensible at all for `array` values to stringify to the equivalent of their contents, and exactly how that happens. So, direct your frustration at the rules for `String([..])`, because that's where the craziness stems from. Perhaps there should be no stringification coercion of `array`s at all? But that would have lots of other downsides in other parts of the language.

Another famously cited gotcha:

```
0 == "\n";           // true
```

As we discussed earlier with empty `""` , `"\n"` (or `"\r\n"` or any other whitespace combination) is coerced via `ToNumber` , and the result is `0` . What other `number` value would you expect whitespace to coerce to? Does it bother you that *explicit* `Number("")` yields `0` ?

Really the only other reasonable `number` value that empty strings or whitespace strings could coerce to is the `NAN` . But would that *really* be better? The comparison `" " == NaN` would of course fail, but it's unclear that we'd have really *fixed* any of the underlying concerns.

The chances that a real-world JS program fails because `0 == "\n"` are awfully rare, and such corner cases are easy to avoid.

Type conversions **always** have corner cases, in any language -- nothing specific to coercion. The issues here are about second-guessing a certain set of corner cases (and perhaps rightly so!?), but that's not a salient argument against the overall coercion mechanism.

Bottom line: almost any crazy coercion between *normal values* that you're likely to run into (aside from intentionally tricky `valueOf()` or `toString()` hacks as earlier) will boil down to the short seven-item list of gotcha coercions we've identified above.

To contrast against these 24 likely suspects for coercion gotchas, consider another list like this:

```
42 == "43";           // false
"foo" == 42;          // false
"true" == true;       // false

42 == "42";           // true
"foo" == [ "foo" ];   // true
```

In these nonfalsy, noncorner cases (and there are literally an infinite number of comparisons we could put on this list), the coercion results are totally safe, reasonable, and explainable.

## Sanity Check

OK, we've definitely found some crazy stuff when we've looked deeply into *implicit* coercion. No wonder that most developers claim coercion is evil and should be avoided, right!?

But let's take a step back and do a sanity check.

By way of magnitude comparison, we have *a list* of seven troublesome gotcha coercions, but we have *another list* of (at least 17, but actually infinite) coercions that are totally sane and explainable.

If you're looking for a textbook example of "throwing the baby out with the bathwater," this is it: discarding the entirety of coercion (the infinitely large list of safe and useful behaviors) because of a list of literally just seven gotchas.

The more prudent reaction would be to ask, "how can I use the countless *good parts* of coercion, but avoid the few *bad parts*?"

Let's look again at the *bad* list:

```
"0" == false;        // true -- UH OH!
false == 0;          // true -- UH OH!
false == "";         // true -- UH OH!
false == [];         // true -- UH OH!
"" == 0;             // true -- UH OH!
"" == [];           // true -- UH OH!
0 == [];            // true -- UH OH!
```

Four of the seven items on this list involve `== false` comparison, which we said earlier you should **always, always** avoid. That's a pretty easy rule to remember.

Now the list is down to three.

```
"" == 0;           // true -- UH OH!
"" == [];          // true -- UH OH!
0 == [];           // true -- UH OH!
```

Are these reasonable coercions you'd do in a normal JavaScript program? Under what conditions would they really happen?

I don't think it's terribly likely that you'd literally use `== []` in a `boolean` test in your program, at least not if you know what you're doing. You'd probably instead be doing `== ""` or `== 0`, like:

```
function doSomething(a) {
  if (a == "") {
    // ...
  }
}
```

You'd have an oops if you accidentally called `doSomething(0)` or `doSomething([])`. Another scenario:

```
function doSomething(a,b) {
  if (a == b) {
    // ...
  }
}
```

Again, this could break if you did something like `doSomething("",0)` or `doSomething([], "")`.

So, while the situations *can* exist where these coercions will bite you, and you'll want to be careful around them, they're probably not super common on the whole of your code base.

## Safely Using Implicit Coercion

The most important advice I can give you: examine your program and reason about what values can show up on either side of an `==` comparison. To effectively avoid issues with such comparisons, here's some heuristic rules to follow:

1. If either side of the comparison can have `true` or `false` values, don't ever, EVER use `==`.
2. If either side of the comparison can have `[]`, `""`, or `0` values, seriously consider not using `==`.

In these scenarios, it's almost certainly better to use `===` instead of `==`, to avoid unwanted coercion. Follow those two simple rules and pretty much all the coercion gotchas that could reasonably hurt you will effectively be avoided.

**Being more explicit/verbose in these cases will save you from a lot of headaches.**

The question of `==` vs. `===` is really appropriately framed as: should you allow coercion for a comparison or not?

There's lots of cases where such coercion can be helpful, allowing you to more tersely express some comparison logic (like with `null` and `undefined`, for example).

In the overall scheme of things, there's relatively few cases where *implicit* coercion is truly dangerous. But in those places, for safety sake, definitely use `===`.

**Tip:** Another place where coercion is guaranteed *not* to bite you is with the `typeof` operator. `typeof` is always going to return you one of seven strings (see Chapter 1), and none of them are the empty `""` string. As such, there's no case where checking the type of some value is going to run afoul of *implicit* coercion. `typeof x == "function"` is 100% as safe and reliable as `typeof x === "function"`. Literally, the spec says the algorithm will be identical in this situation. So, don't just blindly use `==` everywhere simply because that's what your code tools tell you to do, or (worst of all) because you've been told in some book to **not think about it**. You own the quality of your code.

Is *implicit* coercion evil and dangerous? In a few cases, yes, but overwhelmingly, no.

Be a responsible and mature developer. Learn how to use the power of coercion (both *explicit* and *implicit*) effectively and safely. And teach those around you to do the same.

Here's a handy table made by Alex Dorey (@dorey on GitHub) to visualize a variety of comparisons:

The table illustrates the results of comparisons between various JavaScript values. The columns and rows are labeled with the values being compared. The cells contain symbols representing the type of equality or inequality found. A legend on the right side provides a key for these symbols:

- Not equal**: Represented by a light gray square.
- Loose equality**: Often gives "false" positives like "1" is true; [] is "0". Represented by a medium gray square.
- Strict equality**: Mostly evaluates as one would expect. Represented by a dark gray square.

Source: <https://github.com/dorey/JavaScript-Equality-Table>

## Abstract Relational Comparison

While this part of *implicit* coercion often gets a lot less attention, it's important nonetheless to think about what happens with `a < b` comparisons (similar to how we just examined `a == b` in depth).

The "Abstract Relational Comparison" algorithm in ES5 section 11.8.5 essentially divides itself into two parts: what to do if the comparison involves both `string` values (second half), or anything else (first half).

**Note:** The algorithm is only defined for `a < b`. So, `a > b` is handled as `b < a`.

The algorithm first calls `ToPrimitive` coercion on both values, and if the return result of either call is not a `string`, then both values are coerced to `number` values using the `ToNumber` operation rules, and compared numerically.

For example:

```
var a = [ 42 ];
var b = [ "43" ];

a < b;    // true
b < a;    // false
```

**Note:** Similar caveats for `-0` and `NAN` apply here as they did in the `==` algorithm discussed earlier.

However, if both values are `string`s for the `<` comparison, simple lexicographic (natural alphabetic) comparison on the characters is performed:

```
var a = [ "42" ];
var b = [ "043" ];

a < b; // false
```

`a` and `b` are not coerced to `number`s, because both of them end up as `string`s after the `ToPrimitive` coercion on the two array `s`. So, `"42"` is compared character by character to `"043"`, starting with the first characters `"4"` and `"0"`, respectively. Since `"0"` is lexicographically less than `"4"`, the comparison returns `false`.

The exact same behavior and reasoning goes for:

```
var a = [ 4, 2 ];
var b = [ 0, 4, 3 ];

a < b; // false
```

Here, `a` becomes `"4,2"` and `b` becomes `"0,4,3"`, and those lexicographically compare identically to the previous snippet.

What about:

```
var a = { b: 42 };
var b = { b: 43 };

a < b; // ??
```

`a < b` is also `false`, because `a` becomes `[object Object]` and `b` becomes `[object Object]`, and so clearly `a` is not lexicographically less than `b`.

But strangely:

```
var a = { b: 42 };
var b = { b: 43 };

a < b; // false
a == b; // false
a > b; // false

a <= b; // true
a >= b; // true
```

Why is `a == b` not `true`? They're the same `string` value (`"[object Object]"`), so it seems they should be equal, right? Nope. Recall the previous discussion about how `==` works with `object` references.

But then how are `a <= b` and `a >= b` resulting in `true`, if `a < b` and `a == b` and `a > b` are all `false`?

Because the spec says for `a <= b`, it will actually evaluate `b < a` first, and then negate that result. Since `b < a` is also `false`, the result of `a <= b` is `true`.

That's probably awfully contrary to how you might have explained what `<=` does up to now, which would likely have been the literal: "less than or equal to." JS more accurately considers `<=` as "not greater than" (`!(a > b)`, which JS treats as `!(b < a)`). Moreover, `a >= b` is explained by first considering it as `b <= a`, and then applying the same reasoning.

Unfortunately, there is no "strict relational comparison" as there is for equality. In other words, there's no way to prevent *implicit* coercion from occurring with relational comparisons like `a < b`, other than to ensure that `a` and `b` are of the same type explicitly before making the comparison.

Use the same reasoning from our earlier `==` vs. `===` sanity check discussion. If coercion is helpful and reasonably safe, like in a `42 < "43"` comparison, **use it**. On the other hand, if you need to be safe about a relational comparison, *explicitly coerce* the values first, before using `<` (or its counterparts).

```
var a = [ 42 ];
var b = "043";

a < b;           // false -- string comparison!
Number( a ) < Number( b ); // true -- number comparison!
```

## Review

In this chapter, we turned our attention to how JavaScript type conversions happen, called **coercion**, which can be characterized as either *explicit* or *implicit*.

Coercion gets a bad rap, but it's actually quite useful in many cases. An important task for the responsible JS developer is to take the time to learn all the ins and outs of coercion to decide which parts will help improve their code, and which parts they really should avoid.

*Explicit* coercion is code which is obvious that the intent is to convert a value from one type to another. The benefit is improvement in readability and maintainability of code by reducing confusion.

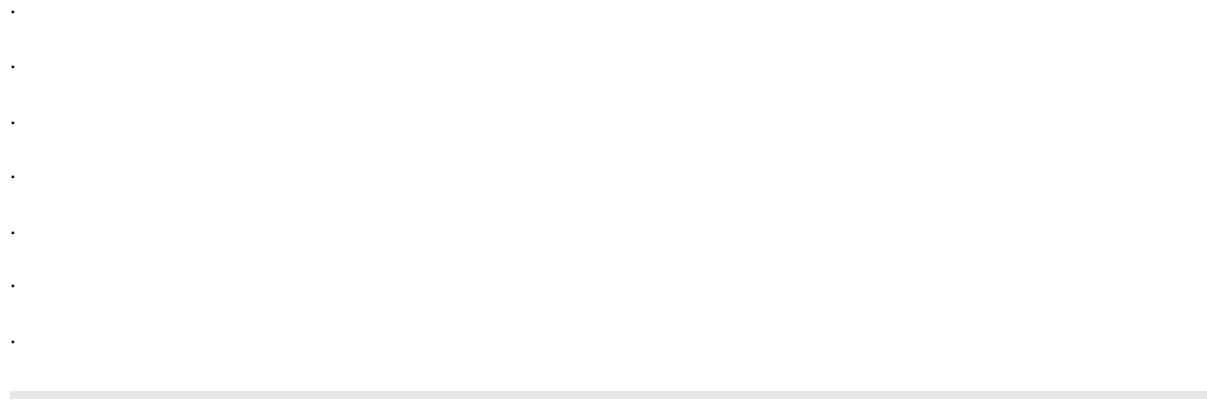
*Implicit* coercion is coercion that is "hidden" as a side-effect of some other operation, where it's not as obvious that the type conversion will occur. While it may seem that *implicit* coercion is the opposite of *explicit* and is thus bad (and indeed, many think so!), actually *implicit* coercion is also about improving the readability of code.

Especially for *implicit*, coercion must be used responsibly and consciously. Know why you're writing the code you're writing, and how it works. Strive to write code that others will easily be able to learn from and understand as well.

# You Don't Know JS Yet: Types & Grammar - 2nd Edition

## Chapter 5: Grammar

<b>NOTE:</b>
Work in progress



<b>NOTE:</b>
Everything below here is previous text from 1st edition, and is only here for reference while 2nd edition work is underway. <b>Please ignore this stuff.</b>

**Note:** The term "grammar" may be a little less familiar to readers than the term "syntax." In many ways, they are similar terms, describing the *rules* for how the language works. There are nuanced differences, but they mostly don't matter for our discussion here. The grammar for JavaScript is a structured way to describe how the syntax (operators, keywords, etc.) fits together into well-formed, valid programs. In other words, discussing syntax without grammar would leave out a lot of the important details. So our focus here in this chapter is most accurately described as *grammar*, even though the raw syntax of the language is what developers directly interact with.

## Statements & Expressions

It's fairly common for developers to assume that the term "statement" and "expression" are roughly equivalent. But here we need to distinguish between the two, because there are some very important differences in our JS programs.

To draw the distinction, let's borrow from terminology you may be more familiar with: the English language.

A "sentence" is one complete formation of words that expresses a thought. It's comprised of one or more "phrases," each of which can be connected with punctuation marks or conjunction words ("and," "or," etc). A phrase can itself be made up of smaller phrases. Some phrases are incomplete and don't accomplish much by themselves, while other phrases can stand on their own. These rules are collectively called the *grammar* of the English language.

And so it goes with JavaScript grammar. Statements are sentences, expressions are phrases, and operators are conjunctions/punctuation.

Every expression in JS can be evaluated down to a single, specific value result. For example:

```
var a = 3 * 6;
var b = a;
b;
```

In this snippet, `3 * 6` is an expression (evaluates to the value `18`). But `a` on the second line is also an expression, as is `b` on the third line. The `a` and `b` expressions both evaluate to the values stored in those variables at that moment, which also happens to be `18`.

Moreover, each of the three lines is a statement containing expressions. `var a = 3 * 6` and `var b = a` are called "declaration statements" because they each declare a variable (and optionally assign a value to it). The `a = 3 * 6` and `b = a` assignments (minus the `var`s) are called assignment expressions.

The third line contains just the expression `b`, but it's also a statement all by itself (though not a terribly interesting one!). This is generally referred to as an "expression statement."

## Statement Completion Values

It's a fairly little known fact that statements all have completion values (even if that value is just `undefined`).

How would you even go about seeing the completion value of a statement?

The most obvious answer is to type the statement into your browser's developer console, because when you execute it, the console by default reports the completion value of the most recent statement it executed.

Let's consider `var b = a`. What's the completion value of that statement?

The `b = a` assignment expression results in the value that was assigned (`18` above), but the `var` statement itself results in `undefined`. Why? Because `var` statements are defined that way in the spec. If you put `var a = 42;` into your console, you'll see `undefined` reported back instead of `42`.

**Note:** Technically, it's a little more complex than that. In the ES5 spec, section 12.2 "Variable Statement," the `variableDeclaration` algorithm actually *does* return a value (a `String` containing the name of the variable declared -- weird, huh!?), but that value is basically swallowed up (except for use by the `for...in` loop) by the `variableStatement` algorithm, which forces an empty (aka `undefined`) completion value.

In fact, if you've done much code experimenting in your console (or in a JavaScript environment REPL -- read/evaluate/print/loop tool), you've probably seen `undefined` reported after many different statements, and perhaps never realized why or what that was. Put simply, the console is just reporting the statement's completion value.

But what the console prints out for the completion value isn't something we can use inside our program. So how can we capture the completion value?

That's a much more complicated task. Before we explain *how*, let's explore *why* you would want to do that.

We need to consider other types of statement completion values. For example, any regular `{ ... }` block has a completion value of the completion value of its last contained statement/expression.

Consider:

```
var b;

if (true) {
  b = 4 + 38;
}
```

If you typed that into your console/REPL, you'd probably see `42` reported, since `42` is the completion value of the `if` block, which took on the completion value of its last assignment expression statement `b = 4 + 38`.

In other words, the completion value of a block is like an *implicit return* of the last statement value in the block.

**Note:** This is conceptually familiar in languages like CoffeeScript, which have implicit `return` values from `function`s that are the same as the last statement value in the function.

But there's an obvious problem. This kind of code doesn't work:

```
var a, b;

a = if (true) {
    b = 4 + 38;
};
```

We can't capture the completion value of a statement and assign it into another variable in any easy syntactic/grammatical way (at least not yet!).

So, what can we do?

**Warning:** For demo purposes only -- don't actually do the following in your real code!

We could use the much maligned `eval(..)` (sometimes pronounced "evil") function to capture this completion value.

```
var a, b;

a = eval( "if (true) { b = 4 + 38; }" );

a; // 42
```

Yeeeeahhhh. That's terribly ugly. But it works! And it illustrates the point that statement completion values are a real thing that can be captured not just in our console but in our programs.

There's a proposal for ES7 called "do expression." Here's how it might work:

```
var a, b;

a = do {
    if (true) {
        b = 4 + 38;
    }
};

a; // 42
```

The `do { .. }` expression executes a block (with one or many statements in it), and the final statement completion value inside the block becomes the completion value of the `do` expression, which can then be assigned to `a` as shown.

The general idea is to be able to treat statements as expressions -- they can show up inside other statements -- without needing to wrap them in an inline function expression and perform an explicit `return ...`.

For now, statement completion values are not much more than trivia. But they're probably going to take on more significance as JS evolves, and hopefully `do { .. }` expressions will reduce the temptation to use stuff like `eval(..)`.

**Warning:** Repeating my earlier admonition: avoid `eval(..)`. Seriously. See the *Scope & Closures* title of this series for more explanation.

## Expression Side Effects

Most expressions don't have side effects. For example:

```
var a = 2;
var b = a + 3;
```

The expression `a + 3` did not *itself* have a side effect, like for instance changing `a`. It had a result, which is `5`, and that result was assigned to `b` in the statement `b = a + 3`.

The most common example of an expression with (possible) side effects is a function call expression:

```
function foo() {
  a = a + 1;
}

var a = 1;
foo();           // result: `undefined`, side effect: changed `a`
```

There are other side-affecting expressions, though. For example:

```
var a = 42;
var b = a++;
```

The expression `a++` has two separate behaviors. *First*, it returns the current value of `a`, which is `42` (which then gets assigned to `b`). But *next*, it changes the value of `a` itself, incrementing it by one.

```
var a = 42;
var b = a++;

a;    // 43
b;    // 42
```

Many developers would mistakenly believe that `b` has value `43` just like `a` does. But the confusion comes from not fully considering the *when* of the side effects of the `++` operator.

The `++` increment operator and the `--` decrement operator are both unary operators (see Chapter 4), which can be used in either a postfix ("after") position or prefix ("before") position.

```
var a = 42;

a++;    // 42
a;      // 43

++a;    // 44
a;      // 44
```

When `++` is used in the prefix position as `++a`, its side effect (incrementing `a`) happens *before* the value is returned from the expression, rather than *after* as with `a++`.

**Note:** Would you think `++a++` was legal syntax? If you try it, you'll get a `ReferenceError` error, but why? Because side-affecting operators **require a variable reference** to target their side effects to. For `++a++`, the `a++` part is evaluated first (because of operator precedence -- see below), which gives back the value of `a` *before* the increment. But then it tries to evaluate `++42`, which (if you try it) gives the same `ReferenceError` error, since `++` can't have a side effect directly on a value like `42`.

It is sometimes mistakenly thought that you can encapsulate the *after* side effect of `a++` by wrapping it in a `( )` pair, like:

```
var a = 42;
```

```
var b = (a++);

a;    // 43
b;    // 42
```

Unfortunately, `( )` itself doesn't define a new wrapped expression that would be evaluated *after the after side effect* of the `a++` expression, as we might have hoped. In fact, even if it did, `a++` returns `42` first, and unless you have another expression that reevaluates `a` after the side effect of `++`, you're not going to get `43` from that expression, so `b` will not be assigned `43`.

There's an option, though: the `,` statement-series comma operator. This operator allows you to string together multiple standalone expression statements into a single statement:

```
var a = 42, b;
b = ( a++, a );

a;    // 43
b;    // 43
```

**Note:** The `( ... )` around `a++`, `a` is required here. The reason is operator precedence, which we'll cover later in this chapter.

The expression `a++, a` means that the second `a` statement expression gets evaluated *after the after side effects* of the first `a++` statement expression, which means it returns the `43` value for assignment to `b`.

Another example of a side-effecting operator is `delete`. As we showed in Chapter 2, `delete` is used to remove a property from an `object` or a slot from an `array`. But it's usually just called as a standalone statement:

```
var obj = {
  a: 42
};

obj.a;      // 42
delete obj.a; // true
obj.a;      // undefined
```

The result value of the `delete` operator is `true` if the requested operation is valid/allowable, or `false` otherwise. But the side effect of the operator is that it removes the property (or array slot).

**Note:** What do we mean by valid/allowable? Nonexistent properties, or properties that exist and are configurable (see Chapter 3 of the *this & Object Prototypes* title of this series) will return `true` from the `delete` operator. Otherwise, the result will be `false` or an error.

One last example of a side-effecting operator, which may at once be both obvious and nonobvious, is the `=` assignment operator.

Consider:

```
var a;

a = 42;      // 42
a;           // 42
```

It may not seem like `=` in `a = 42` is a side-effecting operator for the expression. But if we examine the result value of the `a = 42` statement, it's the value that was just assigned (`42`), so the assignment of that same value into `a` is essentially a side effect.

**Tip:** The same reasoning about side effects goes for the compound-assignment operators like `+=`, `-=`, etc. For example, `a = b += 2` is processed first as `b += 2` (which is `b = b + 2`), and the result of *that* `=` assignment is then assigned to `a`.

This behavior that an assignment expression (or statement) results in the assigned value is primarily useful for chained assignments, such as:

```
var a, b, c;
a = b = c = 42;
```

Here, `c = 42` is evaluated to `42` (with the side effect of assigning `42` to `c`), then `b = 42` is evaluated to `42` (with the side effect of assigning `42` to `b`), and finally `a = 42` is evaluated (with the side effect of assigning `42` to `a`).

**Warning:** A common mistake developers make with chained assignments is like `var a = b = 42`. While this looks like the same thing, it's not. If that statement were to happen without there also being a separate `var b` (somewhere in the scope) to formally declare `b`, then `var a = b = 42` would not declare `b` directly. Depending on `strict` mode, that would either throw an error or create an accidental global (see the *Scope & Closures* title of this series).

Another scenario to consider:

```
function vowels(str) {
  var matches;

  if (str) {
    // pull out all the vowels
    matches = str.match( /[aeiou]/g );
  }

  if (matches) {
    return matches;
  }
}

vowels( "Hello World" ); // ["e", "o", "o"]
```

This works, and many developers prefer such. But using an idiom where we take advantage of the assignment side effect, we can simplify by combining the two `if` statements into one:

```
function vowels(str) {
  var matches;

  // pull out all the vowels
  if (str && (matches = str.match( /[aeiou]/g ))) {
    return matches;
  }

  vowels( "Hello World" ); // ["e", "o", "o"]
```

**Note:** The `( ... )` around `matches = str.match...` is required. The reason is operator precedence, which we'll cover in the "Operator Precedence" section later in this chapter.

I prefer this shorter style, as I think it makes it clearer that the two conditionals are in fact related rather than separate. But as with most stylistic choices in JS, it's purely opinion which one is *better*.

## Contextual Rules

There are quite a few places in the JavaScript grammar rules where the same syntax means different things depending on where/how it's used. This kind of thing can, in isolation, cause quite a bit of confusion.

We won't exhaustively list all such cases here, but just call out a few of the common ones.

### { . . . } Curly Braces

There's two main places (and more coming as JS evolves!) that a pair of `{ ... }` curly braces will show up in your code. Let's take a look at each of them.

## Object Literals

First, as an `object` literal:

```
// assume there's a `bar()` function defined

var a = {
  foo: bar()
};
```

How do we know this is an `object` literal? Because the `{ ... }` pair is a value that's getting assigned to `a`.

**Note:** The `a` reference is called an "l-value" (aka left-hand value) since it's the target of an assignment. The `{ ... }` pair is an "r-value" (aka right-hand value) since it's used *just* as a value (in this case as the source of an assignment).

## Labels

What happens if we remove the `var a =` part of the above snippet?

```
// assume there's a `bar()` function defined

{
  foo: bar()
}
```

A lot of developers assume that the `{ ... }` pair is just a standalone `object` literal that doesn't get assigned anywhere. But it's actually entirely different.

Here, `{ ... }` is just a regular code block. It's not very idiomatic in JavaScript (much more so in other languages!) to have a standalone `{ ... }` block like that, but it's perfectly valid JS grammar. It can be especially helpful when combined with `let` block-scoping declarations (see the *Scope & Closures* title in this series).

The `{ ... }` code block here is functionally pretty much identical to the code block being attached to some statement, like a `for` / `while` loop, `if` conditional, etc.

But if it's a normal block of code, what's that bizarre looking `foo: bar()` syntax, and how is that legal?

It's because of a little known (and, frankly, discouraged) feature in JavaScript called "labeled statements." `foo` is a label for the statement `bar()` (which has omitted its trailing `;` -- see "Automatic Semicolons" later in this chapter). But what's the point of a labeled statement?

If JavaScript had a `goto` statement, you'd theoretically be able to say `goto foo` and have execution jump to that location in code. `goto`s are usually considered terrible coding idioms as they make code much harder to understand (aka "spaghetti code"), so it's a *very good thing* that JavaScript doesn't have a general `goto`.

However, JS *does* support a limited, special form of `goto :labeled jumps`. Both the `continue` and `break` statements can optionally accept a specified label, in which case the program flow "jumps" kind of like a `goto`. Consider:

```
// `foo` labeled-loop
foo: for (var i=0; i<4; i++) {
  for (var j=0; j<4; j++) {
    // whenever the loops meet, continue outer loop
    if (j == i) {
      // jump to the next iteration of
      // the `foo` labeled-loop
      continue foo;
    }
}
```

```

        // skip odd multiples
        if ((j * i) % 2 == 1) {
            // normal (non-labeled) `continue` of inner loop
            continue;
        }

        console.log( i, j );
    }
}

// 1 0
// 2 0
// 2 1
// 3 0
// 3 2

```

**Note:** `continue foo` does not mean "go to the 'foo' labeled position to continue", but rather, "continue the loop that is labeled 'foo' with its next iteration." So, it's not *really* an arbitrary `goto`.

As you can see, we skipped over the odd-multiple `3 1` iteration, but the labeled-loop jump also skipped iterations `1 1` and `2 2`.

Perhaps a slightly more useful form of the labeled jump is with `break __` from inside an inner loop where you want to break out of the outer loop. Without a labeled `break`, this same logic could sometimes be rather awkward to write:

```

// `foo` labeled-loop
foo: for (var i=0; i<4; i++) {
    for (var j=0; j<4; j++) {
        if ((i * j) >= 3) {
            console.log( "stopping!", i, j );
            // break out of the `foo` labeled loop
            break foo;
        }

        console.log( i, j );
    }
}

// 0 0
// 0 1
// 0 2
// 0 3
// 1 0
// 1 1
// 1 2
// stopping! 1 3

```

**Note:** `break foo` does not mean "go to the 'foo' labeled position to continue," but rather, "break out of the loop/block that is labeled 'foo' and continue *after* it." Not exactly a `goto` in the traditional sense, huh?

The nonlabeled `break` alternative to the above would probably need to involve one or more functions, shared scope variable access, etc. It would quite likely be more confusing than labeled `break`, so here using a labeled `break` is perhaps the better option.

A label can apply to a non-loop block, but only `break` can reference such a non-loop label. You can do a labeled `break __` out of any labeled block, but you cannot `continue __` a non-loop label, nor can you do a non-labeled `break` out of a block.

```

function foo() {
    // `bar` labeled-block
    bar: {
        console.log( "Hello" );
        break bar;
        console.log( "never runs" );
    }
}

```

```

    console.log( "World" );
}

foo();
// Hello
// World

```

Labeled loops/blocks are extremely uncommon, and often frowned upon. It's best to avoid them if possible; for example using function calls instead of the loop jumps. But there are perhaps some limited cases where they might be useful. If you're going to use a labeled jump, make sure to document what you're doing with plenty of comments!

It's a very common belief that JSON is a proper subset of JS, so a string of JSON (like `{"a":42}` -- notice the quotes around the property name as JSON requires!) is thought to be a valid JavaScript program. **Not true!** Try putting `{"a":42}` into your JS console, and you'll get an error.

That's because statement labels cannot have quotes around them, so `"a"` is not a valid label, and thus `:` can't come right after it.

So, JSON is truly a subset of JS syntax, but JSON is not valid JS grammar by itself.

One extremely common misconception along these lines is that if you were to load a JS file into a `<script src=..>` tag that only has JSON content in it (like from an API call), the data would be read as valid JavaScript but just be inaccessible to the program. JSON-P (the practice of wrapping the JSON data in a function call, like `foo({"a":42})`) is usually said to solve this inaccessibility by sending the value to one of your program's functions.

**Not true!** The totally valid JSON value `{"a":42}` by itself would actually throw a JS error because it'd be interpreted as a statement block with an invalid label. But `foo({"a":42})` is valid JS because in it, `{"a":42}` is an `object` literal value being passed to `foo(..)`. So, properly said, **JSON-P makes JSON into valid JS grammar!**

## Blocks

Another commonly cited JS gotcha (related to coercion -- see Chapter 4) is:

```

[] + {}; // "[object Object]"
{} + []; // 0

```

This seems to imply the `+` operator gives different results depending on whether the first operand is the `[]` or the `{}`. But that actually has nothing to do with it!

On the first line, `{}` appears in the `+` operator's expression, and is therefore interpreted as an actual value (an empty `object`). Chapter 4 explained that `[]` is coerced to `""` and thus `{}` is coerced to a `string` value as well: `"[object Object]"`.

But on the second line, `{}` is interpreted as a standalone `{}` empty block (which does nothing). Blocks don't need semicolons to terminate them, so the lack of one here isn't a problem. Finally, `+ []` is an expression that *explicitly coerces* (see Chapter 4) the `[]` to a `number`, which is the `0` value.

## Object Destructuring

Starting with ES6, another place that you'll see `{ .. }` pairs showing up is with "destructuring assignments" (see the *ES6 & Beyond* title of this series for more info), specifically `object` destructuring. Consider:

```

function getData() {
  // ...
  return {
    a: 42,
    b: "foo"
  };
}

```

```
var { a, b } = getData();

console.log( a, b ); // 42 "foo"
```

As you can probably tell, `var { a, b } = ...` is a form of ES6 destructuring assignment, which is roughly equivalent to:

```
var res = getData();
var a = res.a;
var b = res.b;
```

**Note:** `{ a, b }` is actually ES6 destructuring shorthand for `{ a: a, b: b }`, so either will work, but it's expected that the shorter `{ a, b }` will become the preferred form.

Object destructuring with a `{ ... }` pair can also be used for named function arguments, which is sugar for this same sort of implicit object property assignment:

```
function foo({ a, b, c }) {
    // no need for:
    // var a = obj.a, b = obj.b, c = obj.c
    console.log( a, b, c );
}

foo( {
    c: [1,2,3],
    a: 42,
    b: "foo"
} );    // 42 "foo" [1, 2, 3]
```

So, the context we use `{ ... }` pairs in entirely determines what they mean, which illustrates the difference between syntax and grammar. It's very important to understand these nuances to avoid unexpected interpretations by the JS engine.

## else if And Optional Blocks

It's a common misconception that JavaScript has an `else if` clause, because you can do:

```
if (a) {
    // ...
}
else if (b) {
    // ...
}
else {
    // ...
}
```

But there's a hidden characteristic of the JS grammar here: there is no `else if`. But `if` and `else` statements are allowed to omit the `{ }` around their attached block if they only contain a single statement. You've seen this many times before, undoubtedly:

```
if (a) doSomething( a );
```

Many JS style guides will insist that you always use `{ }` around a single statement block, like:

```
if (a) { doSomething( a ); }
```

However, the exact same grammar rule applies to the `else` clause, so the `else if` form you've likely always coded is *actually* parsed as:

```

if (a) {
    // ...
}
else {
    if (b) {
        // ...
    }
    else {
        // ...
    }
}

```

The `if (b) { ... } else { ... }` is a single statement that follows the `else`, so you can either put the surrounding `{ }` in or not. In other words, when you use `else if`, you're technically breaking that common style guide rule and just defining your `else` with a single `if` statement.

Of course, the `else if` idiom is extremely common and results in one less level of indentation, so it's attractive. Whichever way you do it, just call out explicitly in your own style guide/rules and don't assume things like `else if` are direct grammar rules.

## Operator Precedence

As we covered in Chapter 4, JavaScript's version of `&&` and `||` are interesting in that they select and return one of their operands, rather than just resulting in `true` or `false`. That's easy to reason about if there are only two operands and one operator.

```

var a = 42;
var b = "foo";

a && b;      // "foo"
a || b;      // 42

```

But what about when there's two operators involved, and three operands?

```

var a = 42;
var b = "foo";
var c = [1,2,3];

a && b || c; // ???
a || b && c; // ???

```

To understand what those expressions result in, we're going to need to understand what rules govern how the operators are processed when there's more than one present in an expression.

These rules are called "operator precedence."

I bet most readers feel they have a decent grasp on operator precedence. But as with everything else we've covered in this book series, we're going to poke and prod at that understanding to see just how solid it really is, and hopefully learn a few new things along the way.

Recall the example from above:

```

var a = 42, b;
b = ( a++, a );

a;      // 43
b;      // 43

```

But what would happen if we remove the `( )`?

```
var a = 42, b;
b = a++, a;

a;    // 43
b;    // 42
```

Wait! Why did that change the value assigned to `b`?

Because the `,` operator has a lower precedence than the `=` operator. So, `b = a++, a` is interpreted as `(b = a++), a`.

Because (as we explained earlier) `a++` has *after side effects*, the assigned value to `b` is the value `42` before the `++` changes `a`.

This is just a simple matter of needing to understand operator precedence. If you're going to use `,` as a statement-series operator, it's important to know that it actually has the lowest precedence. Every other operator will more tightly bind than `,` will.

Now, recall this example from above:

```
if (str && (matches = str.match( /[aeiou]/g ))) {
  // ...
}
```

We said the `( )` around the assignment is required, but why? Because `&&` has higher precedence than `=`, so without the `( )` to force the binding, the expression would instead be treated as `(str && matches) = str.match...`. But this would be an error, because the result of `(str && matches)` isn't going to be a variable, but instead a value (in this case `undefined`), and so it can't be the left-hand side of an `=` assignment!

OK, so you probably think you've got this operator precedence thing down.

Let's move on to a more complex example (which we'll carry throughout the next several sections of this chapter) to *really* test your understanding:

```
var a = 42;
var b = "foo";
var c = false;

var d = a && b || c ? c : b ? a : c && b : a;

d;      // ??
```

OK, evil, I admit it. No one would write a string of expressions like that, right? *Probably* not, but we're going to use it to examine various issues around chaining multiple operators together, which *is* a very common task.

The result above is `42`. But that's not nearly as interesting as how we can figure out that answer without just plugging it into a JS program to let JavaScript sort it out.

Let's dig in.

The first question -- it may not have even occurred to you to ask -- is, does the first part `( a && b || c )` behave like `(a && b) || c` or like `a && (b || c)`? Do you know for certain? Can you even convince yourself they are actually different?

```
(false && true) || true;    // true
false && (true || true);   // false
```

So, there's proof they're different. But still, how does `false && true || true` behave? The answer:

```
false && true || true;      // true
```

```
(false && true) || true; // true
```

So we have our answer. The `&&` operator is evaluated first and the `||` operator is evaluated second.

But is that just because of left-to-right processing? Let's reverse the order of operators:

```
true || false && false; // true
(true || false) && false; // false -- nope
true || (false && false); // true -- winner, winner!
```

Now we've proved that `&&` is evaluated first and then `||`, and in this case that was actually counter to generally expected left-to-right processing.

So what caused the behavior? **Operator precedence**.

Every language defines its own operator precedence list. It's dismaying, though, just how uncommon it is that JS developers have read JS's list.

If you knew it well, the above examples wouldn't have tripped you up in the slightest, because you'd already know that `&&` is more precedent than `||`. But I bet a fair amount of readers had to think about it a little bit.

**Note:** Unfortunately, the JS spec doesn't really have its operator precedence list in a convenient, single location. You have to parse through and understand all the grammar rules. So we'll try to lay out the more common and useful bits here in a more convenient format. For a complete list of operator precedence, see "Operator Precedence" on the MDN site (\* [https://developer.mozilla.org/en-US/docs/Web/JavaScript/Reference/Operators/Operator\\_Precedence](https://developer.mozilla.org/en-US/docs/Web/JavaScript/Reference/Operators/Operator_Precedence)).

## Short Circuited

In Chapter 4, we mentioned in a side note the "short circuiting" nature of operators like `&&` and `||`. Let's revisit that in more detail now.

For both `&&` and `||` operators, the right-hand operand will **not be evaluated** if the left-hand operand is sufficient to determine the outcome of the operation. Hence, the name "short circuited" (in that if possible, it will take an early shortcut out).

For example, with `a && b`, `b` is not evaluated if `a` is falsy, because the result of the `&&` operand is already certain, so there's no point in bothering to check `b`. Likewise, with `a || b`, if `a` is truthy, the result of the operand is already certain, so there's no reason to check `b`.

This short circuiting can be very helpful and is commonly used:

```
function doSomething(opts) {
  if (opts && opts.cool) {
    // ...
  }
}
```

The `opts` part of the `opts && opts.cool` test acts as sort of a guard, because if `opts` is unset (or is not an `object`), the expression `opts.cool` would throw an error. The `opts` test failing plus the short circuiting means that `opts.cool` won't even be evaluated, thus no error!

Similarly, you can use `||` short circuiting:

```
function doSomething(opts) {
  if (opts.cache || primeCache()) {
    // ...
  }
}
```

Here, we're checking for `opts.cache` first, and if it's present, we don't call the `primeCache()` function, thus avoiding potentially unnecessary work.

## Tighter Binding

But let's turn our attention back to that earlier complex statement example with all the chained operators, specifically the `? :` ternary operator parts. Does the `? :` operator have more or less precedence than the `&&` and `||` operators?

```
a && b || c ? c || b ? a : c && b : a
```

Is that more like this:

```
a && b || (c ? c || (b ? a : c) && b : a)
```

or this?

```
(a && b || c) ? (c || b) ? a : (c && b) : a
```

The answer is the second one. But why?

Because `&&` is more precedent than `||`, and `||` is more precedent than `? :`.

So, the expression `(a && b || c)` is evaluated *first* before the `? :` it participates in. Another way this is commonly explained is that `&&` and `||` "bind more tightly" than `? :`. If the reverse was true, then `c ? c...` would bind more tightly, and it would behave (as the first choice) like `a && b || (c ? c...)`.

## Associativity

So, the `&&` and `||` operators bind first, then the `? :` operator. But what about multiple operators of the same precedence? Do they always process left-to-right or right-to-left?

In general, operators are either left-associative or right-associative, referring to whether **grouping happens from the left or from the right**.

It's important to note that associativity is *not* the same thing as left-to-right or right-to-left processing.

But why does it matter whether processing is left-to-right or right-to-left? Because expressions can have side effects, like for instance with function calls:

```
var a = foo() && bar();
```

Here, `foo()` is evaluated first, and then possibly `bar()` depending on the result of the `foo()` expression. That definitely could result in different program behavior than if `bar()` was called before `foo()`.

But this behavior is *just* left-to-right processing (the default behavior in JavaScript!) -- it has nothing to do with the associativity of `&&`. In that example, since there's only one `&&` and thus no relevant grouping here, associativity doesn't even come into play.

But with an expression like `a && b && c`, grouping *will* happen implicitly, meaning that either `a && b` or `b && c` will be evaluated first.

Technically, `a && b && c` will be handled as `(a && b) && c`, because `&&` is left-associative (so is `||`, by the way). However, the right-associative alternative `a && (b && c)` behaves observably the same way. For the same values, the same expressions are evaluated in the same order.

**Note:** If hypothetically `&&` was right-associative, it would be processed the same as if you manually used `( )` to create grouping like `a && (b && c)`. But that still **doesn't mean** that `c` would be processed before `b`. Right-associativity does **not** mean right-to-left evaluation, it means right-to-left **grouping**. Either way, regardless of the grouping/associativity, the strict ordering of evaluation will be `a`, then `b`, then `c` (aka left-to-right).

So it doesn't really matter that much that `&&` and `||` are left-associative, other than to be accurate in how we discuss their definitions.

But that's not always the case. Some operators would behave very differently depending on left-associativity vs. right-associativity.

Consider the `? :` ("ternary" or "conditional") operator:

```
a ? b : c ? d : e;
```

`? :` is right-associative, so which grouping represents how it will be processed?

- `a ? b : (c ? d : e)`
- `(a ? b : c) ? d : e`

The answer is `a ? b : (c ? d : e)`. Unlike with `&&` and `||` above, the right-associativity here actually matters, as `(a ? b : c) ? d : e` *will* behave differently for some (but not all!) combinations of values.

One such example:

```
true ? false : true ? true : true;      // false
true ? false : (true ? true : true);    // false
(true ? false : true) ? true : true;    // true
```

Even more nuanced differences lurk with other value combinations, even if the end result is the same. Consider:

```
true ? false : true ? true : false;      // false
true ? false : (true ? true : false);    // false
(true ? false : true) ? true : false;    // false
```

From that scenario, the same end result implies that the grouping is moot. However:

```
var a = true, b = false, c = true, d = true, e = false;

a ? b : (c ? d : e); // false, evaluates only `a` and `b`
(a ? b : c) ? d : e; // false, evaluates `a`, `b` AND `e`
```

So, we've clearly proved that `? :` is right-associative, and that it actually matters with respect to how the operator behaves if chained with itself.

Another example of right-associativity (grouping) is the `=` operator. Recall the chained assignment example from earlier in the chapter:

```
var a, b, c;
a = b = c = 42;
```

We asserted earlier that `a = b = c = 42` is processed by first evaluating the `c = 42` assignment, then `b = ...`, and finally `a = ...`. Why? Because of the right-associativity, which actually treats the statement like this: `a = (b = (c = 42))`.

Remember our running complex assignment expression example from earlier in the chapter?

```
var a = 42;
var b = "foo";
var c = false;

var d = a && b || c ? c || b ? a : c && b : a;

d;      // 42
```

Armed with our knowledge of precedence and associativity, we should now be able to break down the code into its grouping behavior like this:

```
((a && b) || c) ? ((c || b) ? a : (c && b)) : a
```

Or, to present it indented if that's easier to understand:

```
(  
  (a && b)  
    ||  
    c  
)  
?  
(  
  (c || b)  
    ?  
    a  
    :  
  (c && b)  
)  
:  
a
```

Let's solve it now:

1. `(a && b)` is `"foo"`.
2. `"foo" || c` is `"foo"`.
3. For the first `?` test, `"foo"` is truthy.
4. `(c || b)` is `"foo"`.
5. For the second `?` test, `"foo"` is truthy.
6. `a` is `42`.

That's it, we're done! The answer is `42`, just as we saw earlier. That actually wasn't so hard, was it?

## Disambiguation

You should now have a much better grasp on operator precedence (and associativity) and feel much more comfortable understanding how code with multiple chained operators will behave.

But an important question remains: should we all write code understanding and perfectly relying on all the rules of operator precedence/associativity? Should we only use `( )` manual grouping when it's necessary to force a different processing binding/order?

Or, on the other hand, should we recognize that even though such rules *are in fact* learnable, there's enough gotchas to warrant ignoring automatic precedence/associativity? If so, should we thus always use `( )` manual grouping and remove all reliance on these automatic behaviors?

This debate is highly subjective, and heavily symmetrical to the debate in Chapter 4 over *implicit* coercion. Most developers feel the same way about both debates: either they accept both behaviors and code expecting them, or they discard both behaviors and stick to manual/explicit idioms.

Of course, I cannot answer this question definitively for the reader here anymore than I could in Chapter 4. But I've presented you the pros and cons, and hopefully encouraged enough deeper understanding that you can make informed rather than hype-driven decisions.

In my opinion, there's an important middle ground. We should mix both operator precedence/associativity *and* ( ) manual grouping into our programs -- I argue the same way in Chapter 4 for healthy/safe usage of *implicit* coercion, but certainly don't endorse it exclusively without bounds.

For example, `if (a && b && c) ...` is perfectly OK to me, and I wouldn't do `if ((a && b) && c) ...` just to explicitly call out the associativity, because I think it's overly verbose.

On the other hand, if I needed to chain two `? :` conditional operators together, I'd certainly use ( ) manual grouping to make it absolutely clear what my intended logic is.

Thus, my advice here is similar to that of Chapter 4: **use operator precedence/associativity where it leads to shorter and cleaner code, but use ( ) manual grouping in places where it helps create clarity and reduce confusion.**

## Automatic Semicolons

ASI (Automatic Semicolon Insertion) is when JavaScript assumes a `;` in certain places in your JS program even if you didn't put one there.

Why would it do that? Because if you omit even a single required `;` your program would fail. Not very forgiving. ASI allows JS to be tolerant of certain places where `;` aren't commonly thought to be necessary.

It's important to note that ASI will only take effect in the presence of a newline (aka line break). Semicolons are not inserted in the middle of a line.

Basically, if the JS parser parses a line where a parser error would occur (a missing expected `;`), and it can reasonably insert one, it does so. What's reasonable for insertion? Only if there's nothing but whitespace and/or comments between the end of some statement and that line's newline/line break.

Consider:

```
var a = 42, b  
c;
```

Should JS treat the `c` on the next line as part of the `var` statement? It certainly would if a `,` had come anywhere (even another line) between `b` and `c`. But since there isn't one, JS assumes instead that there's an implied `;` (at the newline) after `b`. Thus, `c;` is left as a standalone expression statement.

Similarly:

```
var a = 42, b = "foo";  
  
a  
b // "foo"
```

That's still a valid program without error, because expression statements also accept ASI.

There's certain places where ASI is helpful, like for instance:

```
var a = 42;

do {
    // ..
} while (a)    // <-- ; expected here!
a;
```

The grammar requires a `;` after a `do..while` loop, but not after `while` or `for` loops. But most developers don't remember that! So, ASI helpfully steps in and inserts one.

As we said earlier in the chapter, statement blocks do not require `;` termination, so ASI isn't necessary:

```
var a = 42;

while (a) {
    // ..
} // <-- no ; expected here
a;
```

The other major case where ASI kicks in is with the `break`, `continue`, `return`, and (ES6) `yield` keywords:

```
function foo(a) {
    if (!a) return
    a *= 2;
    // ..
}
```

The `return` statement doesn't carry across the newline to the `a *= 2` expression, as ASI assumes the `;` terminating the `return` statement. Of course, `return` statements *can* easily break across multiple lines, just not when there's nothing after `return` but the newline/line break.

```
function foo(a) {
    return (
        a * 2 + 3 / 12
    );
}
```

Identical reasoning applies to `break`, `continue`, and `yield`.

## Error Correction

One of the most hotly contested *religious wars* in the JS community (besides tabs vs. spaces) is whether to rely heavily/exclusively on ASI or not.

Most, but not all, semicolons are optional, but the two `;`s in the `for ( ... ) ...` loop header are required.

On the pro side of this debate, many developers believe that ASI is a useful mechanism that allows them to write more terse (and more "beautiful") code by omitting all but the strictly required `;`s (which are very few). It is often asserted that ASI makes many `;`s optional, so a correctly written program *without them* is no different than a correctly written program *with them*.

On the con side of the debate, many other developers will assert that there are *too many* places that can be accidental gotchas, especially for newer, less experienced developers, where unintended `;`s being magically inserted change the meaning. Similarly, some developers will argue that if they omit a semicolon, it's a flat-out mistake, and they want their tools (linters, etc.) to catch it before the JS engine *corrects* the mistake under the covers.

Let me just share my perspective. A strict reading of the spec implies that ASI is an "error correction" routine. What kind of error, you may ask? Specifically, a **parser error**. In other words, in an attempt to have the parser fail less, ASI lets it be more tolerant.

But tolerant of what? In my view, the only way a **parser error** occurs is if it's given an incorrect/errored program to parse. So, while ASI is strictly correcting parser errors, the only way it can get such errors is if there were first program authoring errors -- omitting semicolons where the grammar rules require them.

So, to put it more bluntly, when I hear someone claim that they want to omit "optional semicolons," my brain translates that claim to "I want to write the most parser-broken program I can that will still work."

I find that to be a ludicrous position to take and the arguments of saving keystrokes and having more "beautiful code" to be weak at best.

Furthermore, I don't agree that this is the same thing as the spaces vs tabs debate -- that it's purely cosmetic -- but rather I believe it's a fundamental question of writing code that adheres to grammar requirements vs. code that relies on grammar exceptions to just barely skate through.

Another way of looking at it is that relying on ASI is essentially considering newlines to be significant "whitespace." Other languages like Python have true significant whitespace. But is it really appropriate to think of JavaScript as having significant newlines as it stands today?

My take: **use semicolons wherever you know they are "required," and limit your assumptions about ASI to a minimum.**

But don't just take my word for it. Back in 2012, creator of JavaScript Brendan Eich said (<http://brendaneich.com/2012/04/the-infernal-semicolon/>) the following:

The moral of this story: ASI is (formally speaking) a syntactic error correction procedure. If you start to code as if it were a universal significant-newline rule, you will get into trouble... I wish I had made newlines more significant in JS back in those ten days in May, 1995. ... Be careful not to use ASI as if it gave JS significant newlines.

## Errors

Not only does JavaScript have different *subtypes* of errors (`TypeError`, `ReferenceError`, `SyntaxError`, etc.), but also the grammar defines certain errors to be enforced at compile time, as compared to all other errors that happen during runtime.

In particular, there have long been a number of specific conditions that should be caught and reported as "early errors" (during compilation). Any straight-up syntax error is an early error (e.g., `a = ,`), but also the grammar defines things that are syntactically valid but disallowed nonetheless.

Since execution of your code has not begun yet, these errors are not catchable with `try..catch`; they will just fail the parsing/compilation of your program.

**Tip:** There's no requirement in the spec about exactly how browsers (and developer tools) should report errors. So you may see variations across browsers in the following error examples, in what specific subtype of error is reported or what the included error message text will be.

One simple example is with syntax inside a regular expression literal. There's nothing wrong with the JS syntax here, but the invalid regex will throw an early error:

```
var a = /+foo/;      // Error!
```

The target of an assignment must be an identifier (or an ES6 destructuring expression that produces one or more identifiers), so a value like `42` in that position is illegal and can be reported right away:

```
var a;
42 = a;      // Error!
```

ES5's `strict` mode defines even more early errors. For example, in `strict` mode, function parameter names cannot be duplicated:

```
function foo(a,b,a) { }           // just fine
function bar(a,b,a) { "use strict"; } // Error!
```

Another `strict` mode early error is an object literal having more than one property of the same name:

```
(function(){
  "use strict";

  var a = {
    b: 42,
    b: 43
  };           // Error!
})();
```

**Note:** Semantically speaking, such errors aren't technically *syntax* errors but more *grammar* errors -- the above snippets are syntactically valid. But since there is no `GrammarError` type, some browsers use `SyntaxError` instead.

## Using Variables Too Early

ES6 defines a (frankly confusingly named) new concept called the TDZ ("Temporal Dead Zone").

The TDZ refers to places in code where a variable reference cannot yet be made, because it hasn't reached its required initialization.

The most clear example of this is with ES6 `let` block-scoping:

```
{
  a = 2;          // ReferenceError!
  let a;
}
```

The assignment `a = 2` is accessing the `a` variable (which is indeed block-scoped to the `{ ... }` block) before it's been initialized by the `let a` declaration, so it's in the TDZ for `a` and throws an error.

Interestingly, while `typeof` has an exception to be safe for undeclared variables (see Chapter 1), no such safety exception is made for TDZ references:

```
{
  typeof a;      // undefined
  typeof b;      // ReferenceError! (TDZ)
  let b;
}
```

## Function Arguments

Another example of a TDZ violation can be seen with ES6 default parameter values (see the *ES6 & Beyond* title of this series):

```
var b = 3;

function foo( a = 42, b = a + b + 5 ) {
  // ..
}
```

The `b` reference in the assignment would happen in the TDZ for the parameter `b` (not pull in the outer `b` reference), so it will throw an error. However, the `a` in the assignment is fine since by that time it's past the TDZ for parameter `a`.

When using ES6's default parameter values, the default value is applied to the parameter if you either omit an argument, or you pass an `undefined` value in its place:

```
function foo( a = 42, b = a + 1 ) {
  console.log( a, b );
}

foo();           // 42 43
foo( undefined ); // 42 43
foo( 5 );        // 5 6
foo( void 0, 7 ); // 42 7
foo( null );     // null 1
```

**Note:** `null` is coerced to a `0` value in the `a + 1` expression. See Chapter 4 for more info.

From the ES6 default parameter values perspective, there's no difference between omitting an argument and passing an `undefined` value. However, there is a way to detect the difference in some cases:

```
function foo( a = 42, b = a + 1 ) {
  console.log(
    arguments.length, a, b,
    arguments[0], arguments[1]
  );
}

foo();           // 0 42 43 undefined undefined
foo( 10 );       // 1 10 11 10 undefined
foo( 10, undefined ); // 2 10 11 10 undefined
foo( 10, null ); // 2 10 null 10 null
```

Even though the default parameter values are applied to the `a` and `b` parameters, if no arguments were passed in those slots, the `arguments` array will not have entries.

Conversely, if you pass an `undefined` argument explicitly, an entry will exist in the `arguments` array for that argument, but it will be `undefined` and not (necessarily) the same as the default value that was applied to the named parameter for that same slot.

While ES6 default parameter values can create divergence between the `arguments` array slot and the corresponding named parameter variable, this same disjointedness can also occur in tricky ways in ES5:

```
function foo(a) {
  a = 42;
  console.log( arguments[0] );
}

foo( 2 );      // 42 (linked)
foo();          // undefined (not linked)
```

If you pass an argument, the `arguments` slot and the named parameter are linked to always have the same value. If you omit the argument, no such linkage occurs.

But in `strict` mode, the linkage doesn't exist regardless:

```
function foo(a) {
  "use strict";
  a = 42;
  console.log( arguments[0] );
}
```

```
foo( 2 );    // 2 (not linked)
foo();        // undefined (not linked)
```

It's almost certainly a bad idea to ever rely on any such linkage, and in fact the linkage itself is a leaky abstraction that's exposing an underlying implementation detail of the engine, rather than a properly designed feature.

Use of the `arguments` array has been deprecated (especially in favor of ES6 `... rest parameters` -- see the *ES6 & Beyond* title of this series), but that doesn't mean that it's all bad.

Prior to ES6, `arguments` is the only way to get an array of all passed arguments to pass along to other functions, which turns out to be quite useful. You can also mix named parameters with the `arguments` array and be safe, as long as you follow one simple rule: **never refer to a named parameter and its corresponding `arguments` slot at the same time**. If you avoid that bad practice, you'll never expose the leaky linkage behavior.

```
function foo(a) {
  console.log( a + arguments[1] ); // safe!
}

foo( 10, 32 );    // 42
```

## try..finally

You're probably familiar with how the `try..catch` block works. But have you ever stopped to consider the `finally` clause that can be paired with it? In fact, were you aware that `try` only requires either `catch` or `finally`, though both can be present if needed.

The code in the `finally` clause *always* runs (no matter what), and it always runs right after the `try` (and `catch` if present) finish, before any other code runs. In one sense, you can kind of think of the code in a `finally` clause as being in a callback function that will always be called regardless of how the rest of the block behaves.

So what happens if there's a `return` statement inside a `try` clause? It obviously will return a value, right? But does the calling code that receives that value run before or after the `finally`?

```
function foo() {
  try {
    return 42;
  }
  finally {
    console.log( "Hello" );
  }

  console.log( "never runs" );
}

console.log( foo() );
// Hello
// 42
```

The `return 42` runs right away, which sets up the completion value from the `foo()` call. This action completes the `try` clause and the `finally` clause immediately runs next. Only then is the `foo()` function complete, so that its completion value is returned back for the `console.log(..)` statement to use.

The exact same behavior is true of a `throw` inside `try`:

```
function foo() {
  try {
    throw 42;
  }
```

```

    finally {
      console.log( "Hello" );
    }

    console.log( "never runs" );
}

console.log( foo() );
// Hello
// Uncaught Exception: 42

```

Now, if an exception is thrown (accidentally or intentionally) inside a `finally` clause, it will override as the primary completion of that function. If a previous `return` in the `try` block had set a completion value for the function, that value will be abandoned.

```

function foo() {
  try {
    return 42;
  }
  finally {
    throw "Oops!";
  }

  console.log( "never runs" );
}

console.log( foo() );
// Uncaught Exception: Oops!

```

It shouldn't be surprising that other nonlinear control statements like `continue` and `break` exhibit similar behavior to `return` and `throw`:

```

for (var i=0; i<10; i++) {
  try {
    continue;
  }
  finally {
    console.log( i );
  }
}
// 0 1 2 3 4 5 6 7 8 9

```

The `console.log(i)` statement runs at the end of the loop iteration, which is caused by the `continue` statement. However, it still runs before the `i++` iteration update statement, which is why the values printed are `0..9` instead of `1..10`.

**Note:** ES6 adds a `yield` statement, in generators (see the *Async & Performance* title of this series) which in some ways can be seen as an intermediate `return` statement. However, unlike a `return`, a `yield` isn't complete until the generator is resumed, which means a `try { .. yield .. }` has not completed. So an attached `finally` clause will not run right after the `yield` like it does with `return`.

A `return` inside a `finally` has the special ability to override a previous `return` from the `try` or `catch` clause, but only if `return` is explicitly called:

```

function foo() {
  try {
    return 42;
  }
  finally {
    // no `return ..` here, so no override
  }
}

```

```

function bar() {
  try {
    return 42;
  }
  finally {
    // override previous `return 42`
    return;
  }
}

function baz() {
  try {
    return 42;
  }
  finally {
    // override previous `return 42`
    return "Hello";
  }
}

foo();    // 42
bar();    // undefined
baz();    // "Hello"

```

Normally, the omission of `return` in a function is the same as `return;` or even `return undefined;`, but inside a `finally` block the omission of `return` does not act like an overriding `return undefined`; it just lets the previous `return` stand.

In fact, we can really up the craziness if we combine `finally` with labeled `break` (discussed earlier in the chapter):

```

function foo() {
  bar: {
    try {
      return 42;
    }
    finally {
      // break out of `bar` labeled block
      break bar;
    }
  }

  console.log( "Crazy" );

  return "Hello";
}

console.log( foo() );
// Crazy
// Hello

```

But... don't do this. Seriously. Using a `finally + labeled break` to effectively cancel a `return` is doing your best to create the most confusing code possible. I'd wager no amount of comments will redeem this code.

## switch

Let's briefly explore the `switch` statement, a sort-of syntactic shorthand for an `if..else if..else..` statement chain.

```

switch (a) {
  case 2:
    // do something
    break;
  case 42:
    // do another thing
}

```

```

        break;
    default:
        // fallback to here
}

```

As you can see, it evaluates `a` once, then matches the resulting value to each `case` expression (just simple value expressions here). If a match is found, execution will begin in that matched `case`, and will either go until a `break` is encountered or until the end of the `switch` block is found.

That much may not surprise you, but there are several quirks about `switch` you may not have noticed before.

First, the matching that occurs between the `a` expression and each `case` expression is identical to the `==` algorithm (see Chapter 4). Often times `switch` es are used with absolute values in `case` statements, as shown above, so strict matching is appropriate.

However, you may wish to allow coercive equality (aka `==`, see Chapter 4), and to do so you'll need to sort of "hack" the `switch` statement a bit:

```

var a = "42";

switch (true) {
    case a == 10:
        console.log( "10 or '10'" );
        break;
    case a == 42:
        console.log( "42 or '42'" );
        break;
    default:
        // never gets here
}
// 42 or '42'

```

This works because the `case` clause can have any expression (not just simple values), which means it will strictly match that expression's result to the test expression (`true`). Since `a == 42` results in `true` here, the match is made.

Despite `==`, the `switch` matching itself is still strict, between `true` and `true` here. If the `case` expression resulted in something that was truthy but not strictly `true` (see Chapter 4), it wouldn't work. This can bite you if you're for instance using a "logical operator" like `||` or `&&` in your expression:

```

var a = "hello world";
var b = 10;

switch (true) {
    case (a || b == 10):
        // never gets here
        break;
    default:
        console.log( "Oops" );
}
// Oops

```

Since the result of `(a || b == 10)` is `"hello world"` and not `true`, the strict match fails. In this case, the fix is to force the expression explicitly to be a `true` or `false`, such as `case !(a || b == 10):` (see Chapter 4).

Lastly, the `default` clause is optional, and it doesn't necessarily have to come at the end (although that's the strong convention). Even in the `default` clause, the same rules apply about encountering a `break` or not:

```

var a = 10;

switch (a) {

```

```

case 1:
case 2:
    // never gets here
default:
    console.log( "default" );
case 3:
    console.log( "3" );
    break;
case 4:
    console.log( "4" );
}
// default
// 3

```

**Note:** As discussed previously about labeled `break`s, the `break` inside a `case` clause can also be labeled.

The way this snippet processes is that it passes through all the `case` clause matching first, finds no match, then goes back up to the `default` clause and starts executing. Since there's no `break` there, it continues executing in the already skipped over `case 3` block, before stopping once it hits that `break`.

While this sort of round-about logic is clearly possible in JavaScript, there's almost no chance that it's going to make for reasonable or understandable code. Be very skeptical if you find yourself wanting to create such circular logic flow, and if you really do, make sure you include plenty of code comments to explain what you're up to!

## Review

JavaScript grammar has plenty of nuance that we as developers should spend a little more time paying closer attention to than we typically do. A little bit of effort goes a long way to solidifying your deeper knowledge of the language.

Statements and expressions have analogs in English language -- statements are like sentences and expressions are like phrases. Expressions can be pure/self-contained, or they can have side effects.

The JavaScript grammar layers semantic usage rules (aka context) on top of the pure syntax. For example, `{ }` pairs used in various places in your program can mean statement blocks, object literals, (ES6) destructuring assignments, or (ES6) named function arguments.

JavaScript operators all have well-defined rules for precedence (which ones bind first before others) and associativity (how multiple operator expressions are implicitly grouped). Once you learn these rules, it's up to you to decide if precedence/associativity are *too implicit* for their own good, or if they will aid in writing shorter, clearer code.

ASI (Automatic Semicolon Insertion) is a parser-error-correction mechanism built into the JS engine, which allows it under certain circumstances to insert an assumed `;` in places where it is required, was omitted, and where insertion fixes the parser error. The debate rages over whether this behavior implies that most `;` are optional (and can/should be omitted for cleaner code) or whether it means that omitting them is making mistakes that the JS engine merely cleans up for you.

JavaScript has several types of errors, but it's less known that it has two classifications for errors: "early" (compiler thrown, uncatchable) and "runtime" (`try..catch` able). All syntax errors are obviously early errors that stop the program before it runs, but there are others, too.

Function arguments have an interesting relationship to their formal declared named parameters. Specifically, the `arguments` array has a number of gotchas of leaky abstraction behavior if you're not careful. Avoid `arguments` if you can, but if you must use it, by all means avoid using the positional slot in `arguments` at the same time as using a named parameter for that same argument.

The `finally` clause attached to a `try` (or `try..catch`) offers some very interesting quirks in terms of execution processing order. Some of these quirks can be helpful, but it's possible to create lots of confusion, especially if combined with labeled blocks. As always, use `finally` to make code better and clearer, not more clever or confusing.

The `switch` offers some nice shorthand for `if..else if..` statements, but beware of many common simplifying assumptions about its behavior. There are several quirks that can trip you up if you're not careful, but there's also some neat hidden tricks that `switch` has up its sleeve!

# You Don't Know JS Yet: Types & Grammar - 2nd Edition

## Appendix A: Mixed Environment JavaScript

**NOTE:**

Work in progress

**NOTE:**

Everything below here is previous text from 1st edition, and is only here for reference while 2nd edition work is underway.  
**Please ignore this stuff.**

Beyond the core language mechanics we've fully explored in this book, there are several ways that your JS code can behave differently when it runs in the real world. If JS was executing purely inside an engine, it'd be entirely predictable based on nothing but the black-and-white of the spec. But JS pretty much always runs in the context of a hosting environment, which exposes your code to some degree of unpredictability.

For example, when your code runs alongside code from other sources, or when your code runs in different types of JS engines (not just browsers), there are some things that may behave differently.

We'll briefly explore some of these concerns.

## Annex B (ECMAScript)

It's a little known fact that the official name of the language is ECMAScript (referring to the ECMA standards body that manages it). What then is "JavaScript"? JavaScript is the common trademark of the language, of course, but more appropriately, JavaScript is basically the browser implementation of the spec.

The official ECMAScript specification includes "Annex B," which discusses specific deviations from the official spec for the purposes of JS compatibility in browsers.

The proper way to consider these deviations is that they are only reliably present/valid if your code is running in a browser. If your code always runs in browsers, you won't see any observable difference. If not (like if it can run in node.js, Rhino, etc.), or you're not sure, tread carefully.

The main compatibility differences:

- Octal number literals are allowed, such as `0123` (decimal `83`) in non-`strict mode`.

- `window.escape(..)` and `window.unescape(..)` allow you to escape or unescape strings with % -delimited hexadecimal escape sequences. For example: `window.escape( "?foo=97%&bar=3%" )` produces `"%3Ffoo%3D97%25%26bar%3D3%25"` .
- `String.prototype.substr` is quite similar to `String.prototype.substring`, except that instead of the second parameter being the ending index (noninclusive), the second parameter is the `length` (number of characters to include).

## Web ECMAScript

The Web ECMAScript specification (<http://javascript.spec.whatwg.org>) covers the differences between the official ECMAScript specification and the current JavaScript implementations in browsers.

In other words, these items are "required" of browsers (to be compatible with each other) but are not (as of the time of writing) listed in the "Annex B" section of the official spec:

- `<!--` and `-->` are valid single-line comment delimiters.
- `String.prototype` additions for returning HTML-formatted strings: `anchor(..)`, `big(..)`, `blink(..)`, `bold(..)`, `fixed(..)`, `fontcolor(..)`, `fontsize(..)`, `italics(..)`, `link(..)`, `small(..)`, `strike(..)`, and `sub(..)` . **Note:** These are very rarely used in practice, and are generally discouraged in favor of other built-in DOM APIs or user-defined utilities.
- `RegExp` extensions: `RegExp.$1 .. RegExp.$9` (match-groups) and `RegExp.lastMatch` / `RegExp["$&"]` (most recent match).
- `Function.prototype` additions: `Function.prototype.arguments` (aliases internal `arguments` object) and `Function.caller` (aliases internal `arguments.caller` ). **Note:** `arguments` and thus `arguments.caller` are deprecated, so you should avoid using them if possible. That goes doubly so for these aliases -- don't use them!

**Note:** Some other minor and rarely used deviations are not included in our list here. See the external "Annex B" and "Web ECMAScript" documents for more detailed information as needed.

Generally speaking, all these differences are rarely used, so the deviations from the specification are not significant concerns. **Just be careful** if you rely on any of them.

## Host Objects

The well-covered rules for how variables behave in JS have exceptions to them when it comes to variables that are auto-defined, or otherwise created and provided to JS by the environment that hosts your code (browser, etc.) -- so called, "host objects" (which include both built-in `object`s and `function`s).

For example:

```
var a = document.createElement( "div" );
typeof a;                                // "object" -- as expected
Object.prototype.toString.call( a );        // "[object HTMLDivElement]"
a.tagName;                                // "DIV"
```

`a` is not just an `object`, but a special host object because it's a DOM element. It has a different internal `[[Class]]` value (`"HTMLDivElement"`) and comes with predefined (and often unchangeable) properties.

Another such quirk has already been covered, in the "Falsy Objects" section in Chapter 4: some objects can exist but when coerced to `boolean` they (confoundingly) will coerce to `false` instead of the expected `true` .

Other behavior variations with host objects to be aware of can include:

- not having access to normal `object` built-ins like `toString()`
- not being overwritable

- having certain predefined read-only properties
- having methods that cannot be `this`-overridden to other objects
- and more...

Host objects are critical to making our JS code work with its surrounding environment. But it's important to note when you're interacting with a host object and be careful assuming its behaviors, as they will quite often not conform to regular JS `object`s.

One notable example of a host object that you probably interact with regularly is the `console` object and its various functions (`log(..)`, `error(..)`, etc.). The `console` object is provided by the *hosting environment* specifically so your code can interact with it for various development-related output tasks.

In browsers, `console` hooks up to the developer tools' console display, whereas in node.js and other server-side JS environments, `console` is generally connected to the standard-output (`stdout`) and standard-error (`stderr`) streams of the JavaScript environment system process.

## Global DOM Variables

You're probably aware that declaring a variable in the global scope (with or without `var`) creates not only a global variable, but also its mirror: a property of the same name on the `global` object (`window` in the browser).

But what may be less common knowledge is that (because of legacy browser behavior) creating DOM elements with `id` attributes creates global variables of those same names. For example:

```
<div id="foo"></div>
```

And:

```
if (typeof foo == "undefined") {
  foo = 42;           // will never run
}

console.log( foo );    // HTML element
```

You're perhaps used to managing global variable tests (using `typeof` or `... in window` checks) under the assumption that only JS code creates such variables, but as you can see, the contents of your hosting HTML page can also create them, which can easily throw off your existence check logic if you're not careful.

This is yet one more reason why you should, if at all possible, avoid using global variables, and if you have to, use variables with unique names that won't likely collide. But you also need to make sure not to collide with the HTML content as well as any other code.

## Native Prototypes

One of the most widely known and classic pieces of JavaScript *best practice* wisdom is: **never extend native prototypes**.

Whatever method or property name you come up with to add to `Array.prototype` that doesn't (yet) exist, if it's a useful addition and well-designed, and properly named, there's a strong chance it *could* eventually end up being added to the spec -- in which case your extension is now in conflict.

Here's a real example that actually happened to me that illustrates this point well.

I was building an embeddable widget for other websites, and my widget relied on jQuery (though pretty much any framework would have suffered this gotcha). It worked on almost every site, but we ran across one where it was totally broken.

After almost a week of analysis/debugging, I found that the site in question had, buried deep in one of its legacy files, code that looked like this:

```
// Netscape 4 doesn't have Array.push
Array.prototype.push = function(item) {
    this[this.length] = item;
};
```

Aside from the crazy comment (who cares about Netscape 4 anymore!?), this looks reasonable, right?

The problem is, `Array.prototype.push` was added to the spec sometime subsequent to this Netscape 4 era coding, but what was added is not compatible with this code. The standard `push(..)` allows multiple items to be pushed at once. This hacked one ignores the subsequent items.

Basically all JS frameworks have code that relies on `push(..)` with multiple elements. In my case, it was code around the CSS selector engine that was completely busted. But there could conceivably be dozens of other places susceptible.

The developer who originally wrote that `push(..)` hack had the right instinct to call it `push`, but didn't foresee pushing multiple elements. They were certainly acting in good faith, but they created a landmine that didn't go off until almost 10 years later when I unwittingly came along.

There's multiple lessons to take away on all sides.

First, don't extend the natives unless you're absolutely sure your code is the only code that will ever run in that environment. If you can't say that 100%, then extending the natives is dangerous. You must weigh the risks.

Next, don't unconditionally define extensions (because you can overwrite natives accidentally). In this particular example, had the code said this:

```
if (!Array.prototype.push) {
    // Netscape 4 doesn't have Array.push
    Array.prototype.push = function(item) {
        this[this.length] = item;
    };
}
```

The `if` statement guard would have only defined this hacked `push()` for JS environments where it didn't exist. In my case, that probably would have been OK. But even this approach is not without risk:

1. If the site's code (for some crazy reason!) was relying on a `push(..)` that ignored multiple items, that code would have been broken years ago when the standard `push(..)` was rolled out.
2. If any other library had come in and hacked in a `push(..)` ahead of this `if` guard, and it did so in an incompatible way, that would have broken the site at that time.

What that highlights is an interesting question that, frankly, doesn't get enough attention from JS developers: **Should you EVER rely on native built-in behavior** if your code is running in any environment where it's not the only code present?

The strict answer is **no**, but that's awfully impractical. Your code usually can't redefine its own private untouchable versions of all built-in behavior relied on. Even if you *could*, that's pretty wasteful.

So, should you feature-test for the built-in behavior as well as compliance-testing that it does what you expect? And what if that test fails -- should your code just refuse to run?

```
// don't trust Array.prototype.push
(function(){
    if (Array.prototype.push) {
        var a = [];
        a.push(1,2);
        if (a[0] === 1 && a[1] === 2) {
```

```

        // tests passed, safe to use!
        return;
    }

    throw Error(
        "Array#push() is missing/broken!"
    );
})();
}

```

In theory, that sounds plausible, but it's also pretty impractical to design tests for every single built-in method.

So, what should we do? Should we *trust but verify* (feature- and compliance-test) **everything**? Should we just assume existence is compliance and let breakage (caused by others) bubble up as it will?

There's no great answer. The only fact that can be observed is that extending native prototypes is the only way these things bite you.

If you don't do it, and no one else does in the code in your application, you're safe. Otherwise, you should build in at least a little bit of skepticism, pessimism, and expectation of possible breakage.

Having a full set of unit/regression tests of your code that runs in all known environments is one way to surface some of these issues earlier, but it doesn't do anything to actually protect you from these conflicts.

## Shims/Polyfills

It's usually said that the only safe place to extend a native is in an older (non-spec-compliant) environment, since that's unlikely to ever change -- new browsers with new spec features replace older browsers rather than amending them.

If you could see into the future, and know for sure what a future standard was going to be, like for `Array.prototype.foobar`, it'd be totally safe to make your own compatible version of it to use now, right?

```

if (!Array.prototype.foobar) {
    // silly, silly
    Array.prototype.foobar = function() {
        this.push( "foo", "bar" );
    };
}

```

If there's already a spec for `Array.prototype.foobar`, and the specified behavior is equal to this logic, you're pretty safe in defining such a snippet, and in that case it's generally called a "polyfill" (or "shim").

Such code is **very** useful to include in your code base to "patch" older browser environments that aren't updated to the newest specs. Using polyfills is a great way to create predictable code across all your supported environments.

**Tip:** ES5-Shim (<https://github.com/es-shims/es5-shim>) is a comprehensive collection of shims/polyfills for bringing a project up to ES5 baseline, and similarly, ES6-Shim (<https://github.com/es-shims/es6-shim>) provides shims for new APIs added as of ES6. While APIs can be shimmed/polyfilled, new syntax generally cannot. To bridge the syntactic divide, you'll want to also use an ES6-to-ES5 transpiler like Traceur (<https://github.com/google/traceur-compiler/wiki/Getting-Started>).

If there's likely a coming standard, and most discussions agree what it's going to be called and how it will operate, creating the ahead-of-time polyfill for future-facing standards compliance is called "prollyfill" (probably-fill).

The real catch is if some new standard behavior can't be (fully) polyfilled/prollyfilled.

There's debate in the community if a partial-polyfill for the common cases is acceptable (documenting the parts that cannot be polyfilled), or if a polyfill should be avoided if it purely can't be 100% compliant to the spec.

Many developers at least accept some common partial polyfills (like for instance `Object.create(...)`), because the parts that aren't covered are not parts they intend to use anyway.

Some developers believe that the `if` guard around a polyfill/shim should include some form of conformance test, replacing the existing method either if it's absent or fails the tests. This extra layer of compliance testing is sometimes used to distinguish "shim" (compliance tested) from "polyfill" (existence checked).

The only absolute take-away is that there is no absolute *right* answer here. Extending natives, even when done "safely" in older environments, is not 100% safe. The same goes for relying upon (possibly extended) natives in the presence of others' code.

Either should always be done with caution, defensive code, and lots of obvious documentation about the risks.

## <script> s

Most browser-viewed websites/applications have more than one file that contains their code, and it's common to have a few or several `<script src=..></script>` elements in the page that load these files separately, and even a few inline-code `<script> ... </script>` elements as well.

But do these separate files/code snippets constitute separate programs or are they collectively one JS program?

The (perhaps surprising) reality is they act more like independent JS programs in most, but not all, respects.

The one thing they *share* is the single `global` object (`window` in the browser), which means multiple files can append their code to that shared namespace and they can all interact.

So, if one `script` element defines a global function `foo()`, when a second `script` later runs, it can access and call `foo()` just as if it had defined the function itself.

But global variable scope *hoisting* (see the *Scope & Closures* title of this series) does not occur across these boundaries, so the following code would not work (because `foo()`'s declaration isn't yet declared), regardless of if they are (as shown) inline `<script> ... </script>` elements or externally loaded `<script src=..></script>` files:

```
<script>foo();</script>

<script>
  function foo() { ... }
</script>
```

But either of these *would* work instead:

```
<script>
  foo();
  function foo() { ... }
</script>
```

Or:

```
<script>
  function foo() { ... }
</script>

<script>foo();</script>
```

Also, if an error occurs in a `script` element (inline or external), as a separate standalone JS program it will fail and stop, but any subsequent `script`s will run (still with the shared `global`) unimpeded.

You can create `script` elements dynamically from your code, and inject them into the DOM of the page, and the code in them will behave basically as if loaded normally in a separate file:

```
var greeting = "Hello World";
```

```
var el = document.createElement( "script" );

el.text = "function foo(){ alert( greeting );\n} setTimeout( foo, 1000 );";

document.body.appendChild( el );
```

**Note:** Of course, if you tried the above snippet but set `el.src` to some file URL instead of setting `el.text` to the code contents, you'd be dynamically creating an externally loaded `<script src=..></script>` element.

One difference between code in an inline code block and that same code in an external file is that in the inline code block, the sequence of characters `</script>` cannot appear together, as (regardless of where it appears) it would be interpreted as the end of the code block. So, beware of code like:

```
<script>
  var code = "<script>alert( 'Hello World' )</script>";
</script>
```

It looks harmless, but the `</script>` appearing inside the `string` literal will terminate the script block abnormally, causing an error. The most common workaround is:

```
"</sc" + "ript>;
```

Also, beware that code inside an external file will be interpreted in the character set (UTF-8, ISO-8859-8, etc.) the file is served with (or the default), but that same code in an inline `script` element in your HTML page will be interpreted by the character set of the page (or its default).

**Warning:** The `charset` attribute will not work on inline script elements.

Another deprecated practice with inline `script` elements is including HTML-style or X(HT)ML-style comments around inline code, like:

```
<script>
<!--
alert( "Hello" );
//-->
</script>

<script>
<!--//--><![CDATA[//><!--
alert( "World" );
//--><!]]>
</script>
```

Both of these are totally unnecessary now, so if you're still doing that, stop it!

**Note:** Both `<!--` and `-->` (HTML-style comments) are actually specified as valid single-line comment delimiters (`var x = 2; <!-- valid comment and --> another valid line comment`) in JavaScript (see the "Web ECMAScript" section earlier), purely because of this old technique. But never use them.

## Reserved Words

The ES5 spec defines a set of "reserved words" in Section 7.6.1 that cannot be used as standalone variable names. Technically, there are four categories: "keywords", "future reserved words", the `null` literal, and the `true` / `false` boolean literals.

Keywords are the obvious ones like `function` and `switch`. Future reserved words include things like `enum`, though many of the rest of them (`class`, `extends`, etc.) are all now actually used by ES6; there are other strict-mode only reserved words like `interface`.

StackOverflow user "art4theSould" creatively worked all these reserved words into a fun little poem (<http://stackoverflow.com/questions/26255/reserved-keywords-in-javascript/12114140#12114140>):

Let this long package float, Goto private class if short. While protected with debugger case, Continue volatile interface.  
Instanceof super synchronized throw, Extends final export throws.

Try import double enum?

- False, boolean, abstract function, Implements typeof transient break! Void static, default do, Switch int native new.  
Else, delete null public var In return for const, true, char ...Finally catch byte.

**Note:** This poem includes words that were reserved in ES3 (`byte`, `long`, etc.) that are no longer reserved as of ES5.

Prior to ES5, the reserved words also could not be property names or keys in object literals, but that restriction no longer exists.

So, this is not allowed:

```
var import = "42";
```

But this is allowed:

```
var obj = { import: "42" };
console.log( obj.import );
```

You should be aware though that some older browser versions (mainly older IE) weren't completely consistent on applying these rules, so there are places where using reserved words in object property name locations can still cause issues. Carefully test all supported browser environments.

## Implementation Limits

The JavaScript spec does not place arbitrary limits on things such as the number of arguments to a function or the length of a string literal, but these limits exist nonetheless, because of implementation details in different engines.

For example:

```
function addAll() {
    var sum = 0;
    for (var i=0; i < arguments.length; i++) {
        sum += arguments[i];
    }
    return sum;
}

var nums = [];
for (var i=1; i < 1000000; i++) {
    nums.push(i);
}

addAll( 2, 4, 6 );           // 12
addAll.apply( null, nums ); // should be: 499950000
```

In some JS engines, you'll get the correct `499950000` answer, but in others (like Safari 6.x), you'll get the error: "RangeError: Maximum call stack size exceeded."

Examples of other limits known to exist:

- maximum number of characters allowed in a string literal (not just a string value)
- size (bytes) of data that can be sent in arguments to a function call (aka stack size)
- number of parameters in a function declaration
- maximum depth of non-optimized call stack (i.e., with recursion): how long a chain of function calls from one to the other can be
- number of seconds a JS program can run continuously blocking the browser
- maximum length allowed for a variable name
- ...

It's not very common at all to run into these limits, but you should be aware that limits can and do exist, and importantly that they vary between engines.

## Review

We know and can rely upon the fact that the JS language itself has one standard and is predictably implemented by all the modern browsers/engines. This is a very good thing!

But JavaScript rarely runs in isolation. It runs in an environment mixed in with code from third-party libraries, and sometimes it even runs in engines/environments that differ from those found in browsers.

Paying close attention to these issues improves the reliability and robustness of your code.

# You Don't Know JS Yet: Async & Performance - 2nd Edition

<b>NOTE:</b>
Work in progress

## Table of Contents

- [Foreword \(by TBA\)](#)
- [Preface](#)
- [Chapter 1: Asynchrony: Now & Later](#)
- [Chapter 2: Callbacks](#)
- [Chapter 3: Promises](#)
- [Chapter 4: Generators](#)
- [Chapter 5: Program Performance](#)
- [Chapter 6: Benchmarking & Tuning](#)
- [Appendix A: Library: asynquence](#)
- [Appendix B: Advanced Async Patterns](#)

# You Don't Know JS Yet: Sync & Async - 2nd Edition

## Chapter 1: Asynchrony: Now & Later

<b>NOTE:</b>
Work in progress

<b>NOTE:</b>
Everything below here is previous text from 1st edition, and is only here for reference while 2nd edition work is underway. <b>Please ignore this stuff.</b>

Everything below here is previous text from 1st edition, and is only here for reference while 2nd edition work is underway.  
**Please ignore this stuff.**

One of the most important and yet often misunderstood parts of programming in a language like JavaScript is how to express and manipulate program behavior spread out over a period of time.

This is not just about what happens from the beginning of a `for` loop to the end of a `for` loop, which of course takes *some time* (microseconds to milliseconds) to complete. It's about what happens when part of your program runs *now*, and another part of your program runs *later* -- there's a gap between *now* and *later* where your program isn't actively executing.

Practically all nontrivial programs ever written (especially in JS) have in some way or another had to manage this gap, whether that be in waiting for user input, requesting data from a database or file system, sending data across the network and waiting for a response, or performing a repeated task at a fixed interval of time (like animation). In all these various ways, your program has to manage state across the gap in time. As they famously say in London (of the chasm between the subway door and the platform): "mind the gap."

In fact, the relationship between the *now* and *later* parts of your program is at the heart of asynchronous programming.

Asynchronous programming has been around since the beginning of JS, for sure. But most JS developers have never really carefully considered exactly how and why it crops up in their programs, or explored various *other* ways to handle it. The *good enough* approach has always been the humble callback function. Many to this day will insist that callbacks are more than sufficient.

But as JS continues to grow in both scope and complexity, to meet the ever-widening demands of a first-class programming language that runs in browsers and servers and every conceivable device in between, the pains by which we manage asynchrony are becoming increasingly crippling, and they cry out for approaches that are both more capable and more reasonable.

While this all may seem rather abstract right now, I assure you we'll tackle it more completely and concretely as we go on through this book. We'll explore a variety of emerging techniques for async JavaScript programming over the next several chapters.

But before we can get there, we're going to have to understand much more deeply what asynchrony is and how it operates in JS.

## A Program in Chunks

You may write your JS program in one `.js` file, but your program is almost certainly comprised of several chunks, only one of which is going to execute *now*, and the rest of which will execute *later*. The most common unit of *chunk* is the `function`.

The problem most developers new to JS seem to have is that *later* doesn't happen strictly and immediately after *now*. In other words, tasks that cannot complete *now* are, by definition, going to complete asynchronously, and thus we will not have blocking behavior as you might intuitively expect or want.

Consider:

```
// ajax(..) is some arbitrary Ajax function given by a library
var data = ajax( "http://some.url.1" );

console.log( data );
// Oops! `data` generally won't have the Ajax results
```

You're probably aware that standard Ajax requests don't complete synchronously, which means the `ajax(..)` function does not yet have any value to return back to be assigned to `data` variable. If `ajax(..)` could block until the response came back, then the `data = ..` assignment would work fine.

But that's not how we do Ajax. We make an asynchronous Ajax request *now*, and we won't get the results back until *later*.

The simplest (but definitely not only, or necessarily even best!) way of "waiting" from *now* until *later* is to use a function, commonly called a **callback** function:

```
// ajax(..) is some arbitrary Ajax function given by a library
ajax( "http://some.url.1", function myCallbackFunction(data){

    console.log( data ); // Yay, I gots me some `data`!

});
```

**Warning:** You may have heard that it's possible to make synchronous Ajax requests. While that's technically true, you should never, ever do it, under any circumstances, because it locks the browser UI (buttons, menus, scrolling, etc.) and prevents any user interaction whatsoever. This is a terrible idea, and should always be avoided.

Before you protest in disagreement, no, your desire to avoid the mess of callbacks is *not* justification for blocking, synchronous Ajax.

For example, consider this code:

```
function now() {
    return 21;
}

function later() {
    answer = answer * 2;
    console.log( "Meaning of life:", answer );
}

var answer = now();

setTimeout( later, 1000 ); // Meaning of life: 42
```

There are two chunks to this program: the stuff that will run *now*, and the stuff that will run *later*. It should be fairly obvious what those two chunks are, but let's be super explicit:

Now:

```
function now() {
  return 21;
}

function later() { ... }

var answer = now();

setTimeout( later, 1000 );
```

Later:

```
answer = answer * 2;
console.log( "Meaning of life:", answer );
```

The `now` chunk runs right away, as soon as you execute your program. But `setTimeout(...)` also sets up an event (a timeout) to happen *later*, so the contents of the `later()` function will be executed at a later time (1,000 milliseconds from now).

Any time you wrap a portion of code into a `function` and specify that it should be executed in response to some event (timer, mouse click, Ajax response, etc.), you are creating a *later* chunk of your code, and thus introducing asynchrony to your program.

## Async Console

There is no specification or set of requirements around how the `console.*` methods work -- they are not officially part of JavaScript, but are instead added to JS by the *hosting environment* (see the *Types & Grammar* title of this book series).

So, different browsers and JS environments do as they please, which can sometimes lead to confusing behavior.

In particular, there are some browsers and some conditions that `console.log(...)` does not actually immediately output what it's given. The main reason this may happen is because I/O is a very slow and blocking part of many programs (not just JS). So, it may perform better (from the page/UI perspective) for a browser to handle `console` I/O asynchronously in the background, without you perhaps even knowing that occurred.

A not terribly common, but possible, scenario where this could be *observable* (not from code itself but from the outside):

```
var a = {
  index: 1
};

// later
console.log( a ); // ??

// even later
a.index++;
```

We'd normally expect to see the `a` object be snapshotted at the exact moment of the `console.log(...)` statement, printing something like `{ index: 1 }`, such that in the next statement when `a.index++` happens, it's modifying something different than, or just strictly after, the output of `a`.

Most of the time, the preceding code will probably produce an object representation in your developer tools' console that's what you'd expect. But it's possible this same code could run in a situation where the browser felt it needed to defer the `console` I/O to the background, in which case it's *possible* that by the time the object is represented in the browser console, the `a.index++` has already happened, and it shows `{ index: 2 }`.

It's a moving target under what conditions exactly `console` I/O will be deferred, or even whether it will be observable. Just be aware of this possible asynchronicity in I/O in case you ever run into issues in debugging where objects have been modified *after* a `console.log(...)` statement and yet you see the unexpected modifications show up.

**Note:** If you run into this rare scenario, the best option is to use breakpoints in your JS debugger instead of relying on `console` output. The next best option would be to force a "snapshot" of the object in question by serializing it to a `string`, like with `JSON.stringify(...)`.

## Event Loop

Let's make a (perhaps shocking) claim: despite clearly allowing asynchronous JS code (like the timeout we just looked at), up until recently (ES6), JavaScript itself has actually never had any direct notion of asynchrony built into it.

**What!?** That seems like a crazy claim, right? In fact, it's quite true. The JS engine itself has never done anything more than execute a single chunk of your program at any given moment, when asked to.

"Asked to." By whom? That's the important part!

The JS engine doesn't run in isolation. It runs inside a *hosting environment*, which is for most developers the typical web browser. Over the last several years (but by no means exclusively), JS has expanded beyond the browser into other environments, such as servers, via things like Node.js. In fact, JavaScript gets embedded into all kinds of devices these days, from robots to lightbulbs.

But the one common "thread" (that's a not-so-subtle asynchronous joke, for what it's worth) of all these environments is that they have a mechanism in them that handles executing multiple chunks of your program *over time*, at each moment invoking the JS engine, called the "event loop."

In other words, the JS engine has had no innate sense of *time*, but has instead been an on-demand execution environment for any arbitrary snippet of JS. It's the surrounding environment that has always *scheduled* "events" (JS code executions).

So, for example, when your JS program makes an Ajax request to fetch some data from a server, you set up the "response" code in a function (commonly called a "callback"), and the JS engine tells the hosting environment, "Hey, I'm going to suspend execution for now, but whenever you finish with that network request, and you have some data, please *call* this function *back*."

The browser is then set up to listen for the response from the network, and when it has something to give you, it schedules the callback function to be executed by inserting it into the *event loop*.

So what is the *event loop*?

Let's conceptualize it first through some fake-ish code:

```
// `eventLoop` is an array that acts as a queue (first-in, first-out)
var eventLoop = [ ];
var event;

// keep going "forever"
while (true) {
    // perform a "tick"
    if (eventLoop.length > 0) {
        // get the next event in the queue
        event = eventLoop.shift();

        // now, execute the next event
        try {
            event();
        }
        catch (err) {
            reportError(err);
        }
    }
}
```

This is, of course, vastly simplified pseudocode to illustrate the concepts. But it should be enough to help get a better understanding.

As you can see, there's a continuously running loop represented by the `while` loop, and each iteration of this loop is called a "tick." For each tick, if an event is waiting on the queue, it's taken off and executed. These events are your function callbacks.

It's important to note that `setTimeout(...)` doesn't put your callback on the event loop queue. What it does is set up a timer; when the timer expires, the environment places your callback into the event loop, such that some future tick will pick it up and execute it.

What if there are already 20 items in the event loop at that moment? Your callback waits. It gets in line behind the others -- there's not normally a path for preempting the queue and skipping ahead in line. This explains why `setTimeout(...)` timers may not fire with perfect temporal accuracy. You're guaranteed (roughly speaking) that your callback won't fire *before* the time interval you specify, but it can happen at or after that time, depending on the state of the event queue.

So, in other words, your program is generally broken up into lots of small chunks, which happen one after the other in the event loop queue. And technically, other events not related directly to your program can be interleaved within the queue as well.

**Note:** We mentioned "up until recently" in relation to ES6 changing the nature of where the event loop queue is managed. It's mostly a formal technicality, but ES6 now specifies how the event loop works, which means technically it's within the purview of the JS engine, rather than just the *hosting environment*. One main reason for this change is the introduction of ES6 Promises, which we'll discuss in Chapter 3, because they require the ability to have direct, fine-grained control over scheduling operations on the event loop queue (see the discussion of `setTimeout(..0)` in the "Cooperation" section).

## Parallel Threading

It's very common to conflate the terms "async" and "parallel," but they are actually quite different. Remember, async is about the gap between *now* and *later*. But parallel is about things being able to occur simultaneously.

The most common tools for parallel computing are processes and threads. Processes and threads execute independently and may execute simultaneously: on separate processors, or even separate computers, but multiple threads can share the memory of a single process.

An event loop, by contrast, breaks its work into tasks and executes them in serial, disallowing parallel access and changes to shared memory. Parallelism and "serialism" can coexist in the form of cooperating event loops in separate threads.

The interleaving of parallel threads of execution and the interleaving of asynchronous events occur at very different levels of granularity.

For example:

```
function later() {
  answer = answer * 2;
  console.log( "Meaning of life:", answer );
}
```

While the entire contents of `later()` would be regarded as a single event loop queue entry, when thinking about a thread this code would run on, there's actually perhaps a dozen different low-level operations. For example, `answer = answer * 2` requires first loading the current value of `answer`, then putting `2` somewhere, then performing the multiplication, then taking the result and storing it back into `answer`.

In a single-threaded environment, it really doesn't matter that the items in the thread queue are low-level operations, because nothing can interrupt the thread. But if you have a parallel system, where two different threads are operating in the same program, you could very likely have unpredictable behavior.

Consider:

```

var a = 20;

function foo() {
    a = a + 1;
}

function bar() {
    a = a * 2;
}

// ajax(..) is some arbitrary Ajax function given by a library
ajax( "http://some.url.1", foo );
ajax( "http://some.url.2", bar );

```

In JavaScript's single-threaded behavior, if `foo()` runs before `bar()`, the result is that `a` has `42`, but if `bar()` runs before `foo()` the result in `a` will be `41`.

If JS events sharing the same data executed in parallel, though, the problems would be much more subtle. Consider these two lists of pseudocode tasks as the threads that could respectively run the code in `foo()` and `bar()`, and consider what happens if they are running at exactly the same time:

Thread 1 (`x` and `y` are temporary memory locations):

```

foo():
    a. load value of `a` in `X`
    b. store `1` in `Y`
    c. add `X` and `Y`, store result in `X`
    d. store value of `X` in `a`

```

Thread 2 (`x` and `y` are temporary memory locations):

```

bar():
    a. load value of `a` in `X`
    b. store `2` in `Y`
    c. multiply `X` and `Y`, store result in `X`
    d. store value of `X` in `a`

```

Now, let's say that the two threads are running truly in parallel. You can probably spot the problem, right? They use shared memory locations `x` and `y` for their temporary steps.

What's the end result in `a` if the steps happen like this?

```

1a (load value of `a` in `X` ==> `20`)
2a (load value of `a` in `X` ==> `20`)
1b (store `1` in `Y` ==> `1`)
2b (store `2` in `Y` ==> `2`)
1c (add `X` and `Y`, store result in `X` ==> `22`)
1d (store value of `X` in `a` ==> `22`)
2c (multiply `X` and `Y`, store result in `X` ==> `44`)
2d (store value of `X` in `a` ==> `44`)

```

The result in `a` will be `44`. But what about this ordering?

```

1a (load value of `a` in `X` ==> `20`)
2a (load value of `a` in `X` ==> `20`)
2b (store `2` in `Y` ==> `2`)
1b (store `1` in `Y` ==> `1`)
2c (multiply `X` and `Y`, store result in `X` ==> `20`)
1c (add `X` and `Y`, store result in `X` ==> `21`)
1d (store value of `X` in `a` ==> `21`)
2d (store value of `X` in `a` ==> `21`)

```

The result in `a` will be `21`.

So, threaded programming is very tricky, because if you don't take special steps to prevent this kind of interruption/interleaving from happening, you can get very surprising, nondeterministic behavior that frequently leads to headaches.

JavaScript never shares data across threads, which means *that* level of nondeterminism isn't a concern. But that doesn't mean JS is always deterministic. Remember earlier, where the relative ordering of `foo()` and `bar()` produces two different results (`41` or `42`)?

**Note:** It may not be obvious yet, but not all nondeterminism is bad. Sometimes it's irrelevant, and sometimes it's intentional. We'll see more examples of that throughout this and the next few chapters.

## Run-to-Completion

Because of JavaScript's single-threading, the code inside of `foo()` (and `bar()`) is atomic, which means that once `foo()` starts running, the entirety of its code will finish before any of the code in `bar()` can run, or vice versa. This is called "run-to-completion" behavior.

In fact, the run-to-completion semantics are more obvious when `foo()` and `bar()` have more code in them, such as:

```
var a = 1;
var b = 2;

function foo() {
    a++;
    b = b * a;
    a = b + 3;
}

function bar() {
    b--;
    a = 8 + b;
    b = a * 2;
}

// ajax(..) is some arbitrary Ajax function given by a library
ajax( "http://some.url.1", foo );
ajax( "http://some.url.2", bar );
```

Because `foo()` can't be interrupted by `bar()`, and `bar()` can't be interrupted by `foo()`, this program only has two possible outcomes depending on which starts running first -- if threading were present, and the individual statements in `foo()` and `bar()` could be interleaved, the number of possible outcomes would be greatly increased!

Chunk 1 is synchronous (happens *now*), but chunks 2 and 3 are asynchronous (happen *later*), which means their execution will be separated by a gap of time.

Chunk 1:

```
var a = 1;
var b = 2;
```

Chunk 2 (`foo()`):

```
a++;
b = b * a;
a = b + 3;
```

Chunk 3 (`bar()`):

```
b--;
a = 8 + b;
b = a * 2;
```

Chunks 2 and 3 may happen in either-first order, so there are two possible outcomes for this program, as illustrated here:

Outcome 1:

```
var a = 1;
var b = 2;

// foo()
a++;
b = b * a;
a = b + 3;

// bar()
b--;
a = 8 + b;
b = a * 2;

a; // 11
b; // 22
```

Outcome 2:

```
var a = 1;
var b = 2;

// bar()
b--;
a = 8 + b;
b = a * 2;

// foo()
a++;
b = b * a;
a = b + 3;

a; // 183
b; // 180
```

Two outcomes from the same code means we still have nondeterminism! But it's at the function (event) ordering level, rather than at the statement ordering level (or, in fact, the expression operation ordering level) as it is with threads. In other words, it's *more deterministic* than threads would have been.

As applied to JavaScript's behavior, this function-ordering nondeterminism is the common term "race condition," as `foo()` and `bar()` are racing against each other to see which runs first. Specifically, it's a "race condition" because you cannot predict reliably how `a` and `b` will turn out.

**Note:** If there was a function in JS that somehow did not have run-to-completion behavior, we could have many more possible outcomes, right? It turns out ES6 introduces just such a thing (see Chapter 4 "Generators"), but don't worry right now, we'll come back to that!

## Concurrency

Let's imagine a site that displays a list of status updates (like a social network news feed) that progressively loads as the user scrolls down the list. To make such a feature work correctly, (at least) two separate "processes" will need to be executing *simultaneously* (i.e., during the same window of time, but not necessarily at the same instant).

**Note:** We're using "process" in quotes here because they aren't true operating system-level processes in the computer science sense. They're virtual processes, or tasks, that represent a logically connected, sequential series of operations. We'll simply prefer "process" over "task" because terminology-wise, it will match the definitions of the concepts we're exploring.

The first "process" will respond to `onscroll` events (making Ajax requests for new content) as they fire when the user has scrolled the page further down. The second "process" will receive Ajax responses back (to render content onto the page).

Obviously, if a user scrolls fast enough, you may see two or more `onscroll` events fired during the time it takes to get the first response back and process, and thus you're going to have `onscroll` events and Ajax response events firing rapidly, interleaved with each other.

Concurrency is when two or more "processes" are executing simultaneously over the same period, regardless of whether their individual constituent operations happen *in parallel* (at the same instant on separate processors or cores) or not. You can think of concurrency then as "process"-level (or task-level) parallelism, as opposed to operation-level parallelism (separate-processor threads).

**Note:** Concurrency also introduces an optional notion of these "processes" interacting with each other. We'll come back to that later.

For a given window of time (a few seconds worth of a user scrolling), let's visualize each independent "process" as a series of events/operations:

"Process" 1 (`onscroll` events):

```
onscroll, request 1
onscroll, request 2
onscroll, request 3
onscroll, request 4
onscroll, request 5
onscroll, request 6
onscroll, request 7
```

"Process" 2 (Ajax response events):

```
response 1
response 2
response 3
response 4
response 5
response 6
response 7
```

It's quite possible that an `onscroll` event and an Ajax response event could be ready to be processed at exactly the same *moment*. For example, let's visualize these events in a timeline:

```
onscroll, request 1
onscroll, request 2      response 1
onscroll, request 3      response 2
response 3
onscroll, request 4
onscroll, request 5
onscroll, request 6      response 4
onscroll, request 7
response 6
response 5
response 7
```

But, going back to our notion of the event loop from earlier in the chapter, JS is only going to be able to handle one event at a time, so either `onscroll, request 2` is going to happen first or `response 1` is going to happen first, but they cannot happen at literally the same moment. Just like kids at a school cafeteria, no matter what crowd they form outside the doors, they'll have to merge into a single line to get their lunch!

Let's visualize the interleaving of all these events onto the event loop queue.

Event Loop Queue:

```
onscroll, request 1    <--- Process 1 starts
onscroll, request 2
response 1            <--- Process 2 starts
onscroll, request 3
response 2
response 3
onscroll, request 4
onscroll, request 5
onscroll, request 6
response 4
onscroll, request 7    <--- Process 1 finishes
response 6
response 5
response 7            <--- Process 2 finishes
```

"Process 1" and "Process 2" run concurrently (task-level parallel), but their individual events run sequentially on the event loop queue.

By the way, notice how `response 6` and `response 5` came back out of expected order?

The single-threaded event loop is one expression of concurrency (there are certainly others, which we'll come back to later).

## Noninteracting

As two or more "processes" are interleaving their steps/events concurrently within the same program, they don't necessarily need to interact with each other if the tasks are unrelated. **If they don't interact, nondeterminism is perfectly acceptable.**

For example:

```
var res = {};

function foo(results) {
  res.foo = results;
}

function bar(results) {
  res.bar = results;
}

// ajax(..) is some arbitrary Ajax function given by a library
ajax( "http://some.url.1", foo );
ajax( "http://some.url.2", bar );
```

`foo()` and `bar()` are two concurrent "processes," and it's nondeterminate which order they will be fired in. But we've constructed the program so it doesn't matter what order they fire in, because they act independently and as such don't need to interact.

This is not a "race condition" bug, as the code will always work correctly, regardless of the ordering.

## Interaction

More commonly, concurrent "processes" will by necessity interact, indirectly through scope and/or the DOM. When such interaction will occur, you need to coordinate these interactions to prevent "race conditions," as described earlier.

Here's a simple example of two concurrent "processes" that interact because of implied ordering, which is only *sometimes broken*:

```
var res = [];

function response(data) {
    res.push( data );
}

// ajax(..) is some arbitrary Ajax function given by a library
ajax( "http://some.url.1", response );
ajax( "http://some.url.2", response );
```

The concurrent "processes" are the two `response()` calls that will be made to handle the Ajax responses. They can happen in either-first order.

Let's assume the expected behavior is that `res[0]` has the results of the `"http://some.url.1"` call, and `res[1]` has the results of the `"http://some.url.2"` call. Sometimes that will be the case, but sometimes they'll be flipped, depending on which call finishes first. There's a pretty good likelihood that this nondeterminism is a "race condition" bug.

**Note:** Be extremely wary of assumptions you might tend to make in these situations. For example, it's not uncommon for a developer to observe that `"http://some.url.2"` is "always" much slower to respond than `"http://some.url.1"`, perhaps by virtue of what tasks they're doing (e.g., one performing a database task and the other just fetching a static file), so the observed ordering seems to always be as expected. Even if both requests go to the same server, and *it* intentionally responds in a certain order, there's no *real* guarantee of what order the responses will arrive back in the browser.

So, to address such a race condition, you can coordinate ordering interaction:

```
var res = [];

function response(data) {
    if (data.url == "http://some.url.1") {
        res[0] = data;
    }
    else if (data.url == "http://some.url.2") {
        res[1] = data;
    }
}

// ajax(..) is some arbitrary Ajax function given by a library
ajax( "http://some.url.1", response );
ajax( "http://some.url.2", response );
```

Regardless of which Ajax response comes back first, we inspect the `data.url` (assuming one is returned from the server, of course!) to figure out which position the response data should occupy in the `res` array. `res[0]` will always hold the `"http://some.url.1"` results and `res[1]` will always hold the `"http://some.url.2"` results. Through simple coordination, we eliminated the "race condition" nondeterminism.

The same reasoning from this scenario would apply if multiple concurrent function calls were interacting with each other through the shared DOM, like one updating the contents of a `<div>` and the other updating the style or attributes of the `<div>` (e.g., to make the DOM element visible once it has content). You probably wouldn't want to show the DOM element before it had content, so the coordination must ensure proper ordering interaction.

Some concurrency scenarios are *always broken* (not just *sometimes*) without coordinated interaction. Consider:

```
var a, b;

function foo(x) {
```

```

    a = x * 2;
    baz();
}

function bar(y) {
    b = y * 2;
    baz();
}

function baz() {
    console.log(a + b);
}

// ajax(..) is some arbitrary Ajax function given by a library
ajax( "http://some.url.1", foo );
ajax( "http://some.url.2", bar );

```

In this example, whether `foo()` or `bar()` fires first, it will always cause `baz()` to run too early (either `a` or `b` will still be `undefined`), but the second invocation of `baz()` will work, as both `a` and `b` will be available.

There are different ways to address such a condition. Here's one simple way:

```

var a, b;

function foo(x) {
    a = x * 2;
    if (a && b) {
        baz();
    }
}

function bar(y) {
    b = y * 2;
    if (a && b) {
        baz();
    }
}

function baz() {
    console.log( a + b );
}

// ajax(..) is some arbitrary Ajax function given by a library
ajax( "http://some.url.1", foo );
ajax( "http://some.url.2", bar );

```

The `if (a && b)` conditional around the `baz()` call is traditionally called a "gate," because we're not sure what order `a` and `b` will arrive, but we wait for both of them to get there before we proceed to open the gate (call `baz()`).

Another concurrency interaction condition you may run into is sometimes called a "race," but more correctly called a "latch." It's characterized by "only the first one wins" behavior. Here, nondeterminism is acceptable, in that you are explicitly saying it's OK for the "race" to the finish line to have only one winner.

Consider this broken code:

```

var a;

function foo(x) {
    a = x * 2;
    baz();
}

function bar(x) {
    a = x / 2;
}

```

```

        baz();
    }

    function baz() {
        console.log( a );
    }

// ajax(..) is some arbitrary Ajax function given by a library
ajax( "http://some.url.1", foo );
ajax( "http://some.url.2", bar );

```

Whichever one (`foo()` or `bar()`) fires last will not only overwrite the assigned `a` value from the other, but it will also duplicate the call to `baz()` (likely undesired).

So, we can coordinate the interaction with a simple latch, to let only the first one through:

```

var a;

function foo(x) {
    if (a == undefined) {
        a = x * 2;
        baz();
    }
}

function bar(x) {
    if (a == undefined) {
        a = x / 2;
        baz();
    }
}

function baz() {
    console.log( a );
}

// ajax(..) is some arbitrary Ajax function given by a library
ajax( "http://some.url.1", foo );
ajax( "http://some.url.2", bar );

```

The `if (a == undefined)` conditional allows only the first of `foo()` or `bar()` through, and the second (and indeed any subsequent) calls would just be ignored. There's just no virtue in coming in second place!

**Note:** In all these scenarios, we've been using global variables for simplistic illustration purposes, but there's nothing about our reasoning here that requires it. As long as the functions in question can access the variables (via scope), they'll work as intended. Relying on lexically scoped variables (see the *Scope & Closures* title of this book series), and in fact global variables as in these examples, is one obvious downside to these forms of concurrency coordination. As we go through the next few chapters, we'll see other ways of coordination that are much cleaner in that respect.

## Cooperation

Another expression of concurrency coordination is called "cooperative concurrency." Here, the focus isn't so much on interacting via value sharing in scopes (though that's obviously still allowed!). The goal is to take a long-running "process" and break it up into steps or batches so that other concurrent "processes" have a chance to interleave their operations into the event loop queue.

For example, consider an Ajax response handler that needs to run through a long list of results to transform the values. We'll use `Array#map(..)` to keep the code shorter:

```

var res = [];

// `response(..)` receives array of results from the Ajax call

```

```

function response(data) {
    // add onto existing `res` array
    res = res.concat(
        // make a new transformed array with all `data` values doubled
        data.map( function(val){
            return val * 2;
        })
    );
}

// ajax(..) is some arbitrary Ajax function given by a library
ajax( "http://some.url.1", response );
ajax( "http://some.url.2", response );

```

If `"http://some.url.1"` gets its results back first, the entire list will be mapped into `res` all at once. If it's a few thousand or less records, this is not generally a big deal. But if it's say 10 million records, that can take a while to run (several seconds on a powerful laptop, much longer on a mobile device, etc.).

While such a "process" is running, nothing else in the page can happen, including no other `response(..)` calls, no UI updates, not even user events like scrolling, typing, button clicking, and the like. That's pretty painful.

So, to make a more cooperatively concurrent system, one that's friendlier and doesn't hog the event loop queue, you can process these results in asynchronous batches, after each one "yielding" back to the event loop to let other waiting events happen.

Here's a very simple approach:

```

var res = [];

// `response(..)` receives array of results from the Ajax call
function response(data) {
    // let's just do 1000 at a time
    var chunk = data.splice( 0, 1000 );

    // add onto existing `res` array
    res = res.concat(
        // make a new transformed array with all `chunk` values doubled
        chunk.map( function(val){
            return val * 2;
        })
    );

    // anything left to process?
    if (data.length > 0) {
        // async schedule next batch
        setTimeout( function(){
            response( data );
        }, 0 );
    }
}

// ajax(..) is some arbitrary Ajax function given by a library
ajax( "http://some.url.1", response );
ajax( "http://some.url.2", response );

```

We process the data set in maximum-sized chunks of 1,000 items. By doing so, we ensure a short-running "process," even if that means many more subsequent "processes," as the interleaving onto the event loop queue will give us a much more responsive (performant) site/app.

Of course, we're not interaction-coordinating the ordering of any of these "processes," so the order of results in `res` won't be predictable. If ordering was required, you'd need to use interaction techniques like those we discussed earlier, or ones we will cover in later chapters of this book.

We use the `setTimeout(..0)` (hack) for async scheduling, which basically just means "stick this function at the end of the current event loop queue."

**Note:** `setTimeout(..0)` is not technically inserting an item directly onto the event loop queue. The timer will insert the event at its next opportunity. For example, two subsequent `setTimeout(..0)` calls would not be strictly guaranteed to be processed in call order, so it is possible to see various conditions like timer drift where the ordering of such events isn't predictable. In Node.js, a similar approach is `process.nextTick(..)`. Despite how convenient (and usually more performant) it would be, there's not a single direct way (at least yet) across all environments to ensure async event ordering. We cover this topic in more detail in the next section.

## Jobs

As of ES6, there's a new concept layered on top of the event loop queue, called the "Job queue." The most likely exposure you'll have to it is with the asynchronous behavior of Promises (see Chapter 3).

Unfortunately, at the moment it's a mechanism without an exposed API, and thus demonstrating it is a bit more convoluted. So we're going to have to just describe it conceptually, such that when we discuss async behavior with Promises in Chapter 3, you'll understand how those actions are being scheduled and processed.

So, the best way to think about this that I've found is that the "Job queue" is a queue hanging off the end of every tick in the event loop queue. Certain async-implied actions that may occur during a tick of the event loop will not cause a whole new event to be added to the event loop queue, but will instead add an item (aka Job) to the end of the current tick's Job queue.

It's kinda like saying, "oh, here's this other thing I need to do *later*, but make sure it happens right away before anything else can happen."

Or, to use a metaphor: the event loop queue is like an amusement park ride, where once you finish the ride, you have to go to the back of the line to ride again. But the Job queue is like finishing the ride, but then cutting in line and getting right back on.

A Job can also cause more Jobs to be added to the end of the same queue. So, it's theoretically possible that a Job "loop" (a Job that keeps adding another Job, etc.) could spin indefinitely, thus starving the program of the ability to move on to the next event loop tick. This would conceptually be almost the same as just expressing a long-running or infinite loop (like `while (true) ..`) in your code.

Jobs are kind of like the spirit of the `setTimeout(..0)` hack, but implemented in such a way as to have a much more well-defined and guaranteed ordering: **later, but as soon as possible**.

Let's imagine an API for scheduling Jobs (directly, without hacks), and call it `schedule(..)`. Consider:

```
console.log( "A" );

setTimeout( function(){
    console.log( "B" );
}, 0 );

// theoretical "Job API"
schedule( function(){
    console.log( "C" );

    schedule( function(){
        console.log( "D" );
    } );
} );
```

You might expect this to print out `A B C D`, but instead it would print out `A C D B`, because the Jobs happen at the end of the current event loop tick, and the timer fires to schedule for the *next* event loop tick (if available!).

In Chapter 3, we'll see that the asynchronous behavior of Promises is based on Jobs, so it's important to keep clear how that relates to event loop behavior.

## Statement Ordering

The order in which we express statements in our code is not necessarily the same order as the JS engine will execute them. That may seem like quite a strange assertion to make, so we'll just briefly explore it.

But before we do, we should be crystal clear on something: the rules/grammar of the language (see the *Types & Grammar* title of this book series) dictate a very predictable and reliable behavior for statement ordering from the program point of view. So what we're about to discuss are **not things you should ever be able to observe** in your JS program.

**Warning:** If you are ever able to *observe* compiler statement reordering like we're about to illustrate, that'd be a clear violation of the specification, and it would unquestionably be due to a bug in the JS engine in question -- one which should promptly be reported and fixed! But it's vastly more common that you *suspect* something crazy is happening in the JS engine, when in fact it's just a bug (probably a "race condition"!) in your own code -- so look there first, and again and again. The JS debugger, using breakpoints and stepping through code line by line, will be your most powerful tool for sniffing out such bugs in *your code*.

Consider:

```
var a, b;

a = 10;
b = 30;

a = a + 1;
b = b + 1;

console.log( a + b ); // 42
```

This code has no expressed asynchrony to it (other than the rare `console.async` I/O discussed earlier!), so the most likely assumption is that it would process line by line in top-down fashion.

But it's *possible* that the JS engine, after compiling this code (yes, JS is compiled -- see the *Scope & Closures* title of this book series!) might find opportunities to run your code faster by rearranging (safely) the order of these statements. Essentially, as long as you can't observe the reordering, anything's fair game.

For example, the engine might find it's faster to actually execute the code like this:

```
var a, b;

a = 10;
a++;

b = 30;
b++;

console.log( a + b ); // 42
```

Or this:

```
var a, b;

a = 11;
b = 31;

console.log( a + b ); // 42
```

Or even:

```
// because `a` and `b` aren't used anymore, we can
// inline and don't even need them!
console.log( 42 ); // 42
```

In all these cases, the JS engine is performing safe optimizations during its compilation, as the end *observable* result will be the same.

But here's a scenario where these specific optimizations would be unsafe and thus couldn't be allowed (of course, not to say that it's not optimized at all):

```
var a, b;

a = 10;
b = 30;

// we need `a` and `b` in their preincremented state!
console.log( a * b ); // 300

a = a + 1;
b = b + 1;

console.log( a + b ); // 42
```

Other examples where the compiler reordering could create observable side effects (and thus must be disallowed) would include things like any function call with side effects (even and especially getter functions), or ES6 Proxy objects (see the *ES6 & Beyond* title of this book series).

Consider:

```
function foo() {
    console.log( b );
    return 1;
}

var a, b, c;

// ES5.1 getter literal syntax
c = {
    get bar() {
        console.log( a );
        return 1;
    }
};

a = 10;
b = 30;

a += foo();           // 30
b += c.bar;          // 11

console.log( a + b ); // 42
```

If it weren't for the `console.log(..)` statements in this snippet (just used as a convenient form of observable side effect for the illustration), the JS engine would likely have been free, if it wanted to (who knows if it would!?), to reorder the code to:

```
// ...
a = 10 + foo();
b = 30 + c.bar;
```

```
// ...
```

While JS semantics thankfully protect us from the *observable* nightmares that compiler statement reordering would seem to be in danger of, it's still important to understand just how tenuous a link there is between the way source code is authored (in top-down fashion) and the way it runs after compilation.

Compiler statement reordering is almost a micro-metaphor for concurrency and interaction. As a general concept, such awareness can help you understand async JS code flow issues better.

## Review

A JavaScript program is (practically) always broken up into two or more chunks, where the first chunk runs *now* and the next chunk runs *later*, in response to an event. Even though the program is executed chunk-by-chunk, all of them share the same access to the program scope and state, so each modification to state is made on top of the previous state.

Whenever there are events to run, the *event loop* runs until the queue is empty. Each iteration of the event loop is a "tick." User interaction, IO, and timers enqueue events on the event queue.

At any given moment, only one event can be processed from the queue at a time. While an event is executing, it can directly or indirectly cause one or more subsequent events.

Concurrency is when two or more chains of events interleave over time, such that from a high-level perspective, they appear to be running *simultaneously* (even though at any given moment only one event is being processed).

It's often necessary to do some form of interaction coordination between these concurrent "processes" (as distinct from operating system processes), for instance to ensure ordering or to prevent "race conditions." These "processes" can also *cooperate* by breaking themselves into smaller chunks and to allow other "process" interleaving.

# You Don't Know JS Yet: Sync & Async - 2nd Edition

## Chapter 2: Callbacks

**NOTE:**

Work in progress

**NOTE:**

Everything below here is previous text from 1st edition, and is only here for reference while 2nd edition work is underway.  
**Please ignore this stuff.**

In Chapter 1, we explored the terminology and concepts around asynchronous programming in JavaScript. Our focus is on understanding the single-threaded (one-at-a-time) event loop queue that drives all "events" (async function invocations). We also explored various ways that concurrency patterns explain the relationships (if any!) between *simultaneously* running chains of events, or "processes" (tasks, function calls, etc.).

All our examples in Chapter 1 used the function as the individual, indivisible unit of operations, whereby inside the function, statements run in predictable order (above the compiler level!), but at the function-ordering level, events (aka `async` function invocations) can happen in a variety of orders.

In all these cases, the function is acting as a "callback," because it serves as the target for the event loop to "call back into" the program, whenever that item in the queue is processed.

As you no doubt have observed, callbacks are by far the most common way that asynchrony in JS programs is expressed and managed. Indeed, the callback is the most fundamental `async` pattern in the language.

Countless JS programs, even very sophisticated and complex ones, have been written upon no other `async` foundation than the callback (with of course the concurrency interaction patterns we explored in Chapter 1). The callback function is the `async` work horse for JavaScript, and it does its job respectably.

Except... callbacks are not without their shortcomings. Many developers are excited by the *promise* (pun intended!) of better `async` patterns. But it's impossible to effectively use any abstraction if you don't understand what it's abstracting, and why.

In this chapter, we will explore a couple of those in depth, as motivation for why more sophisticated `async` patterns (explored in subsequent chapters of this book) are necessary and desired.

## Continuations

Let's go back to the async callback example we started with in Chapter 1, but let me slightly modify it to illustrate a point:

```
// A
ajax( "...", function(..){
  // C
} );
// B
```

`// A` and `// B` represent the first half of the program (aka the *now*), and `// c` marks the second half of the program (aka the *later*). The first half executes right away, and then there's a "pause" of indeterminate length. At some future moment, if the Ajax call completes, then the program will pick up where it left off, and *continue* with the second half.

In other words, the callback function wraps or encapsulates the *continuation* of the program.

Let's make the code even simpler:

```
// A
setTimeout( function(){
  // C
}, 1000 );
// B
```

Stop for a moment and ask yourself how you'd describe (to someone else less informed about how JS works) the way that program behaves. Go ahead, try it out loud. It's a good exercise that will help my next points make more sense.

Most readers just now probably thought or said something to the effect of: "Do A, then set up a timeout to wait 1,000 milliseconds, then once that fires, do C." How close was your rendition?

You might have caught yourself and self-edited to: "Do A, setup the timeout for 1,000 milliseconds, then do B, then after the timeout fires, do C." That's more accurate than the first version. Can you spot the difference?

Even though the second version is more accurate, both versions are deficient in explaining this code in a way that matches our brains to the code, and the code to the JS engine. The disconnect is both subtle and monumental, and is at the very heart of understanding the shortcomings of callbacks as async expression and management.

As soon as we introduce a single continuation (or several dozen as many programs do!) in the form of a callback function, we have allowed a divergence to form between how our brains work and the way the code will operate. Any time these two diverge (and this is by far not the only place that happens, as I'm sure you know!), we run into the inevitable fact that our code becomes harder to understand, reason about, debug, and maintain.

## Sequential Brain

I'm pretty sure most of you readers have heard someone say (even made the claim yourself), "I'm a multitasker." The effects of trying to act as a multitasker range from humorous (e.g., the silly patting-head-rubbing-stomach kids' game) to mundane (chewing gum while walking) to downright dangerous (texting while driving).

But are we multitaskers? Can we really do two conscious, intentional actions at once and think/reason about both of them at exactly the same moment? Does our highest level of brain functionality have parallel multithreading going on?

The answer may surprise you: **probably not.**

That's just not really how our brains appear to be set up. We're much more single taskers than many of us (especially A-type personalities!) would like to admit. We can really only think about one thing at any given instant.

I'm not talking about all our involuntary, subconscious, automatic brain functions, such as heart beating, breathing, and eyelid blinking. Those are all vital tasks to our sustained life, but we don't intentionally allocate any brain power to them. Thankfully, while we obsess about checking social network feeds for the 15th time in three minutes, our brain carries on in the background

(threads!) with all those important tasks.

We're instead talking about whatever task is at the forefront of our minds at the moment. For me, it's writing the text in this book right now. Am I doing any other higher level brain function at exactly this same moment? Nope, not really. I get distracted quickly and easily -- a few dozen times in these last couple of paragraphs!

When we *fake* multitasking, such as trying to type something at the same time we're talking to a friend or family member on the phone, what we're actually most likely doing is acting as fast context switchers. In other words, we switch back and forth between two or more tasks in rapid succession, *simultaneously* progressing on each task in tiny, fast little chunks. We do it so fast that to the outside world it appears as if we're doing these things *in parallel*.

Does that sound suspiciously like async evented concurrency (like the sort that happens in JS) to you?! If not, go back and read Chapter 1 again!

In fact, one way of simplifying (i.e., abusing) the massively complex world of neurology into something I can remotely hope to discuss here is that our brains work kinda like the event loop queue.

If you think about every single letter (or word) I type as a single async event, in just this sentence alone there are several dozen opportunities for my brain to be interrupted by some other event, such as from my senses, or even just my random thoughts.

I don't get interrupted and pulled to another "process" at every opportunity that I could be (thankfully -- or this book would never be written!). But it happens often enough that I feel my own brain is nearly constantly switching to various different contexts (aka "processes"). And that's an awful lot like how the JS engine would probably feel.

## Doing Versus Planning

OK, so our brains can be thought of as operating in single-threaded event loop queue like ways, as can the JS engine. That sounds like a good match.

But we need to be more nuanced than that in our analysis. There's a big, observable difference between how we plan various tasks, and how our brains actually operate those tasks.

Again, back to the writing of this text as my metaphor. My rough mental outline plan here is to keep writing and writing, going sequentially through a set of points I have ordered in my thoughts. I don't plan to have any interruptions or nonlinear activity in this writing. But yet, my brain is nevertheless switching around all the time.

Even though at an operational level our brains are async evented, we seem to plan out tasks in a sequential, synchronous way. "I need to go to the store, then buy some milk, then drop off my dry cleaning."

You'll notice that this higher level thinking (planning) doesn't seem very async evented in its formulation. In fact, it's kind of rare for us to deliberately think solely in terms of events. Instead, we plan things out carefully, sequentially (A then B then C), and we assume to an extent a sort of temporal blocking that forces B to wait on A, and C to wait on B.

When a developer writes code, they are planning out a set of actions to occur. If they're any good at being a developer, they're **carefully planning** it out. "I need to set `z` to the value of `x`, and then `x` to the value of `y`," and so forth.

When we write out synchronous code, statement by statement, it works a lot like our errands to-do list:

```
// swap `x` and `y` (via temp variable `z`)
z = x;
x = y;
y = z;
```

These three assignment statements are synchronous, so `x = y` waits for `z = x` to finish, and `y = z` in turn waits for `x = y` to finish. Another way of saying it is that these three statements are temporally bound to execute in a certain order, one right after the other. Thankfully, we don't need to be bothered with any async evented details here. If we did, the code gets a lot more complex, quickly!

So if synchronous brain planning maps well to synchronous code statements, how well do our brains do at planning out asynchronous code?

It turns out that how we express asynchrony (with callbacks) in our code doesn't map very well at all to that synchronous brain planning behavior.

Can you actually imagine having a line of thinking that plans out your to-do errands like this?

"I need to go to the store, but on the way I'm sure I'll get a phone call, so 'Hi, Mom', and while she starts talking, I'll be looking up the store address on GPS, but that'll take a second to load, so I'll turn down the radio so I can hear Mom better, then I'll realize I forgot to put on a jacket and it's cold outside, but no matter, keep driving and talking to Mom, and then the seatbelt ding reminds me to buckle up, so 'Yes, Mom, I am wearing my seatbelt, I always do!'. Ah, finally the GPS got the directions, now..."

As ridiculous as that sounds as a formulation for how we plan our day out and think about what to do and in what order, nonetheless it's exactly how our brains operate at a functional level. Remember, that's not multitasking, it's just fast context switching.

The reason it's difficult for us as developers to write async evented code, especially when all we have is the callback to do it, is that stream of consciousness thinking/planning is unnatural for most of us.

We think in step-by-step terms, but the tools (callbacks) available to us in code are not expressed in a step-by-step fashion once we move from synchronous to asynchronous.

And **that** is why it's so hard to accurately author and reason about async JS code with callbacks: because it's not how our brain planning works.

**Note:** The only thing worse than not knowing why some code breaks is not knowing why it worked in the first place! It's the classic "house of cards" mentality: "it works, but not sure why, so nobody touch it!" You may have heard, "Hell is other people" (Sartre), and the programmer meme twist, "Hell is other people's code." I believe truly: "Hell is not understanding my own code." And callbacks are one main culprit.

## Nested/Chained Callbacks

Consider:

```
listen( "click", function handler(evt){
    setTimeout( function request(){
        ajax( "http://some.url.1", function response(text){
            if (text == "hello") {
                handler();
            }
            else if (text == "world") {
                request();
            }
        });
    }, 500);
});
```

There's a good chance code like that is recognizable to you. We've got a chain of three functions nested together, each one representing a step in an asynchronous series (task, "process").

This kind of code is often called "callback hell," and sometimes also referred to as the "pyramid of doom" (for its sideways-facing triangular shape due to the nested indentation).

But "callback hell" actually has almost nothing to do with the nesting/indentation. It's a far deeper problem than that. We'll see how and why as we continue through the rest of this chapter.

First, we're waiting for the "click" event, then we're waiting for the timer to fire, then we're waiting for the Ajax response to come back, at which point it might do it all again.

At first glance, this code may seem to map its asynchrony naturally to sequential brain planning.

First (*now*), we:

```
listen( "...", function handler(..){
    // ...
} );
```

Then *later*, we:

```
setTimeout( function request(..){
    // ...
}, 500) ;
```

Then still *later*, we:

```
ajax( "...", function response(..){
    // ...
} );
```

And finally (most *later*), we:

```
if ( ... ) {
    // ...
}
else ..
```

But there's several problems with reasoning about this code linearly in such a fashion.

First, it's an accident of the example that our steps are on subsequent lines (1, 2, 3, and 4...). In real async JS programs, there's often a lot more noise cluttering things up, noise that we have to deftly maneuver past in our brains as we jump from one function to the next. Understanding the async flow in such callback-laden code is not impossible, but it's certainly not natural or easy, even with lots of practice.

But also, there's something deeper wrong, which isn't evident just in that code example. Let me make up another scenario (pseudocode-ish) to illustrate it:

```
doA( function(){
    doB();

    doC( function(){
        doD();
    } )

    doE();
} );

doF();
```

While the experienced among you will correctly identify the true order of operations here, I'm betting it is more than a little confusing at first glance, and takes some concerted mental cycles to arrive at. The operations will happen in this order:

- doA()
- doF()
- doB()

- doC()
- doE()
- doD()

Did you get that right the very first time you glanced at the code?

OK, some of you are thinking I was unfair in my function naming, to intentionally lead you astray. I swear I was just naming in top-down appearance order. But let me try again:

```
doA( function(){
    doC();
}

doD( function(){
    doF();
} )

doE();
} );

doB();
```

Now, I've named them alphabetically in order of actual execution. But I still bet, even with experience now in this scenario, tracing through the A -> B -> C -> D -> E -> F order doesn't come natural to many of you readers. Certainly, your eyes do an awful lot of jumping up and down the code snippet, right?

But even if that all comes natural to you, there's still one more hazard that could wreak havoc. Can you spot what it is?

What if doA(..) or doD(..) aren't actually async, the way we obviously assumed them to be? Uh oh, now the order is different. If they're both sync (and maybe only sometimes, depending on the conditions of the program at the time), the order is now A -> C -> D -> F -> E -> B .

That sound you just heard faintly in the background is the sighs of thousands of JS developers who just had a face-in-hands moment.

Is nesting the problem? Is that what makes it so hard to trace the async flow? That's part of it, certainly.

But let me rewrite the previous nested event/timeout/Ajax example without using nesting:

```
listen( "click", handler );

function handler() {
    setTimeout( request, 500 );
}

function request(){
    ajax( "http://some.url.1", response );
}

function response(text){
    if (text == "hello") {
        handler();
    }
    else if (text == "world") {
        request();
    }
}
```

This formulation of the code is not hardly as recognizable as having the nesting/indentation woes of its previous form, and yet it's every bit as susceptible to "callback hell." Why?

As we go to linearly (sequentially) reason about this code, we have to skip from one function, to the next, to the next, and bounce all around the code base to "see" the sequence flow. And remember, this is simplified code in sort of best-case fashion. We all know that real async JS program code bases are often fantastically more jumbled, which makes such reasoning orders of magnitude more difficult.

Another thing to notice: to get steps 2, 3, and 4 linked together so they happen in succession, the only affordance callbacks alone gives us is to hardcode step 2 into step 1, step 3 into step 2, step 4 into step 3, and so on. The hardcoded isn't necessarily a bad thing, if it really is a fixed condition that step 2 should always lead to step 3.

But the hardcoded definitely makes the code a bit more brittle, as it doesn't account for anything going wrong that might cause a deviation in the progression of steps. For example, if step 2 fails, step 3 never gets reached, nor does step 2 retry, or move to an alternate error handling flow, and so on.

All of these issues are things you *can* manually hardcode into each step, but that code is often very repetitive and not reusable in other steps or in other async flows in your program.

Even though our brains might plan out a series of tasks in a sequential type of way (this, then this, then this), the evented nature of our brain operation makes recovery/retry/forking of flow control almost effortless. If you're out running errands, and you realize you left a shopping list at home, it doesn't end the day because you didn't plan that ahead of time. Your brain routes around this hiccup easily: you go home, get the list, then head right back out to the store.

But the brittle nature of manually hardcoded callbacks (even with hardcoded error handling) is often far less graceful. Once you end up specifying (aka pre-planning) all the various eventualities/paths, the code becomes so convoluted that it's hard to ever maintain or update it.

**That** is what "callback hell" is all about! The nesting/indentation are basically a side show, a red herring.

And as if all that's not enough, we haven't even touched what happens when two or more chains of these callback continuations are happening *simultaneously*, or when the third step branches out into "parallel" callbacks with gates or latches, or... OMG, my brain hurts, how about yours!?

Are you catching the notion here that our sequential, blocking brain planning behaviors just don't map well onto callback-oriented async code? That's the first major deficiency to articulate about callbacks: they express asynchrony in code in ways our brains have to fight just to keep in sync with (pun intended!).

## Trust Issues

The mismatch between sequential brain planning and callback-driven async JS code is only part of the problem with callbacks. There's something much deeper to be concerned about.

Let's once again revisit the notion of a callback function as the continuation (aka the second half) of our program:

```
// A
ajax( "...", function(...){
  // C
} );
// B
```

// A and // B happen *now*, under the direct control of the main JS program. But // c gets deferred to happen *later*, and under the control of another party -- in this case, the ajax(...) function. In a basic sense, that sort of hand-off of control doesn't regularly cause lots of problems for programs.

But don't be fooled by its infrequency that this control switch isn't a big deal. In fact, it's one of the worst (and yet most subtle) problems about callback-driven design. It revolves around the idea that sometimes ajax(..) (i.e., the "party" you hand your callback continuation to) is not a function that you wrote, or that you directly control. Many times, it's a utility provided by some third party.

We call this "inversion of control," when you take part of your program and give over control of its execution to another third party. There's an unspoken "contract" that exists between your code and the third-party utility -- a set of things you expect to be maintained.

## Tale of Five Callbacks

It might not be terribly obvious why this is such a big deal. Let me construct an exaggerated scenario to illustrate the hazards of trust at play.

Imagine you're a developer tasked with building out an ecommerce checkout system for a site that sells expensive TVs. You already have all the various pages of the checkout system built out just fine. On the last page, when the user clicks "confirm" to buy the TV, you need to call a third-party function (provided say by some analytics tracking company) so that the sale can be tracked.

You notice that they've provided what looks like an async tracking utility, probably for the sake of performance best practices, which means you need to pass in a callback function. In this continuation that you pass in, you will have the final code that charges the customer's credit card and displays the thank you page.

This code might look like:

```
analytics.trackPurchase( purchaseData, function(){
    chargeCreditCard();
    displayThankyouPage();
});
```

Easy enough, right? You write the code, test it, everything works, and you deploy to production. Everyone's happy!

Six months go by and no issues. You've almost forgotten you even wrote that code. One morning, you're at a coffee shop before work, casually enjoying your latte, when you get a panicked call from your boss insisting you drop the coffee and rush into work right away.

When you arrive, you find out that a high-profile customer has had his credit card charged five times for the same TV, and he's understandably upset. Customer service has already issued an apology and processed a refund. But your boss demands to know how this could possibly have happened. "Don't we have tests for stuff like this!?"

You don't even remember the code you wrote. But you dig back in and start trying to find out what could have gone awry.

After digging through some logs, you come to the conclusion that the only explanation is that the analytics utility somehow, for some reason, called your callback five times instead of once. Nothing in their documentation mentions anything about this.

Frustrated, you contact customer support, who of course is as astonished as you are. They agree to escalate it to their developers, and promise to get back to you. The next day, you receive a lengthy email explaining what they found, which you promptly forward to your boss.

Apparently, the developers at the analytics company had been working on some experimental code that, under certain conditions, would retry the provided callback once per second, for five seconds, before failing with a timeout. They had never intended to push that into production, but somehow they did, and they're totally embarrassed and apologetic. They go into plenty of detail about how they've identified the breakdown and what they'll do to ensure it never happens again. Yadda, yadda.

What's next?

You talk it over with your boss, but he's not feeling particularly comfortable with the state of things. He insists, and you reluctantly agree, that you can't trust *them* anymore (that's what bit you), and that you'll need to figure out how to protect the checkout code from such a vulnerability again.

After some tinkering, you implement some simple ad hoc code like the following, which the team seems happy with:

```
var tracked = false;
```

```
analytics.trackPurchase( purchaseData, function(){
  if (!tracked) {
    tracked = true;
    chargeCreditCard();
    displayThankyouPage();
  }
});
```

**Note:** This should look familiar to you from Chapter 1, because we're essentially creating a latch to handle if there happen to be multiple concurrent invocations of our callback.

But then one of your QA engineers asks, "what happens if they never call the callback?" Oops. Neither of you had thought about that.

You begin to chase down the rabbit hole, and think of all the possible things that could go wrong with them calling your callback. Here's roughly the list you come up with of ways the analytics utility could misbehave:

- Call the callback too early (before it's been tracked)
- Call the callback too late (or never)
- Call the callback too few or too many times (like the problem you encountered!)
- Fail to pass along any necessary environment/parameters to your callback
- Swallow any errors/exceptions that may happen
- ...

That should feel like a troubling list, because it is. You're probably slowly starting to realize that you're going to have to invent an awful lot of ad hoc logic **in each and every single callback** that's passed to a utility you're not positive you can trust.

Now you realize a bit more completely just how hellish "callback hell" is.

## Not Just Others' Code

Some of you may be skeptical at this point whether this is as big a deal as I'm making it out to be. Perhaps you don't interact with truly third-party utilities much if at all. Perhaps you use versioned APIs or self-host such libraries, so that its behavior can't be changed out from underneath you.

So, contemplate this: can you even *really* trust utilities that you do theoretically control (in your own code base)?

Think of it this way: most of us agree that at least to some extent we should build our own internal functions with some defensive checks on the input parameters, to reduce/prevent unexpected issues.

Overly trusting of input:

```
function addNumbers(x,y) {
  // + is overloaded with coercion to also be
  // string concatenation, so this operation
  // isn't strictly safe depending on what's
  // passed in.
  return x + y;
}

addNumbers( 21, 21 );      // 42
addNumbers( 21, "21" );    // "2121"
```

Defensive against untrusted input:

```
function addNumbers(x,y) {
  // ensure numerical input
  if (typeof x != "number" || typeof y != "number") {
    throw Error( "Bad parameters" );
  }
}
```

```

    }

    // if we get here, + will safely do numeric addition
    return x + y;
}

addNumbers( 21, 21 );      // 42
addNumbers( 21, "21" );    // Error: "Bad parameters"

```

Or perhaps still safe but friendlier:

```

function addNumbers(x,y) {
    // ensure numerical input
    x = Number( x );
    y = Number( y );

    // + will safely do numeric addition
    return x + y;
}

addNumbers( 21, 21 );      // 42
addNumbers( 21, "21" );    // 42

```

However you go about it, these sorts of checks/normalizations are fairly common on function inputs, even with code we theoretically entirely trust. In a crude sort of way, it's like the programming equivalent of the geopolitical principle of "Trust But Verify."

So, doesn't it stand to reason that we should do the same thing about composition of async function callbacks, not just with truly external code but even with code we know is generally "under our own control"? **Of course we should.**

But callbacks don't really offer anything to assist us. We have to construct all that machinery ourselves, and it often ends up being a lot of boilerplate/overhead that we repeat for every single async callback.

The most troublesome problem with callbacks is *inversion of control* leading to a complete breakdown along all those trust lines.

If you have code that uses callbacks, especially but not exclusively with third-party utilities, and you're not already applying some sort of mitigation logic for all these *inversion of control* trust issues, your code *has* bugs in it right now even though they may not have bitten you yet. Latent bugs are still bugs.

Hell indeed.

## Trying to Save Callbacks

There are several variations of callback design that have attempted to address some (not all!) of the trust issues we've just looked at. It's a valiant, but doomed, effort to save the callback pattern from imploding on itself.

For example, regarding more graceful error handling, some API designs provide for split callbacks (one for the success notification, one for the error notification):

```

function success(data) {
    console.log( data );
}

function failure(err) {
    console.error( err );
}

ajax( "http://some.url.1", success, failure );

```

In APIs of this design, often the `failure()` error handler is optional, and if not provided it will be assumed you want the errors swallowed. Ugh.

**Note:** This split-callback design is what the ES6 Promise API uses. We'll cover ES6 Promises in much more detail in the next chapter.

Another common callback pattern is called "error-first style" (sometimes called "Node style," as it's also the convention used across nearly all Node.js APIs), where the first argument of a single callback is reserved for an error object (if any). If success, this argument will be empty/falsy (and any subsequent arguments will be the success data), but if an error result is being signaled, the first argument is set/truthy (and usually nothing else is passed):

```
function response(err, data) {
  // error?
  if (err) {
    console.error( err );
  }
  // otherwise, assume success
  else {
    console.log( data );
  }
}

ajax( "http://some.url.1", response );
```

In both of these cases, several things should be observed.

First, it has not really resolved the majority of trust issues like it may appear. There's nothing about either callback that prevents or filters unwanted repeated invocations. Moreover, things are worse now, because you may get both success and error signals, or neither, and you still have to code around either of those conditions.

Also, don't miss the fact that while it's a standard pattern you can employ, it's definitely more verbose and boilerplate-ish without much reuse, so you're going to get weary of typing all that out for every single callback in your application.

What about the trust issue of never being called? If this is a concern (and it probably should be!), you likely will need to set up a timeout that cancels the event. You could make a utility (proof-of-concept only shown) to help you with that:

```
function timeoutify(fn, delay) {
  var intv = setTimeout( function(){
    intv = null;
    fn( new Error( "Timeout!" ) );
  }, delay )
;

  return function() {
    // timeout hasn't happened yet?
    if (intv) {
      clearTimeout( intv );
      fn.apply( this, [ null ].concat( [].slice.call( arguments ) ) );
    }
  };
}
```

Here's how you use it:

```
// using "error-first style" callback design
function foo(err,data) {
  if (err) {
    console.error( err );
  }
  else {
    console.log( data );
  }
}
```

```

}

ajax( "http://some.url.1", timeoutify( foo, 500 ) );

```

Another trust issue is being called "too early." In application-specific terms, this may actually involve being called before some critical task is complete. But more generally, the problem is evident in utilities that can either invoke the callback you provide *now* (synchronously), or *later* (asynchronously).

This nondeterminism around the sync-or-async behavior is almost always going to lead to very difficult to track down bugs. In some circles, the fictional insanity-inducing monster named Zalgo is used to describe the sync/async nightmares. "Don't release Zalgo!" is a common cry, and it leads to very sound advice: always invoke callbacks asynchronously, even if that's "right away" on the next turn of the event loop, so that all callbacks are predictably async.

**Note:** For more information on Zalgo, see Oren Golan's "Don't Release Zalgo!"

(<https://github.com/oren/oren.github.io/blob/master/posts/zalgo.md>) and Isaac Z. Schlueter's "Designing APIs for Asynchrony" (<http://blog.izs.me/post/59142742143/designing-apis-for-asynchrony>).

Consider:

```

function result(data) {
  console.log( a );
}

var a = 0;

ajax( "...pre-cached-url...", result );
a++;

```

Will this code print `0` (sync callback invocation) or `1` (async callback invocation)? Depends... on the conditions.

You can see just how quickly the unpredictability of Zalgo can threaten any JS program. So the silly-sounding "never release Zalgo" is actually incredibly common and solid advice. Always be asyncing.

What if you don't know whether the API in question will always execute async? You could invent a utility like this `asyncify(..)` proof-of-concept:

```

function asyncify(fn) {
  var orig_fn = fn,
    intv = setTimeout( function(){
      intv = null;
      if (fn) fn();
    }, 0 )
  ;

  fn = null;

  return function() {
    // firing too quickly, before `intv` timer has fired to
    // indicate async turn has passed?
    if (intv) {
      fn = orig_fn.bind.apply(
        orig_fn,
        // add the wrapper's `this` to the `bind(..)`
        // call parameters, as well as currying any
        // passed in parameters
        [this].concat( [].slice.call( arguments ) )
      );
    }
    // already async
    else {
      // invoke original function
      orig_fn.apply( this, arguments );
    }
  }
}

```

```

        }
    };
}

```

You use `asyncify(..)` like this:

```

function result(data) {
    console.log( a );
}

var a = 0;

ajax( "...pre-cached-url...", asyncify( result ) );
a++;

```

Whether the Ajax request is in the cache and resolves to try to call the callback right away, or must be fetched over the wire and thus complete later asynchronously, this code will always output `1` instead of `0` -- `result(..)` cannot help but be invoked asynchronously, which means the `a++` has a chance to run before `result(..)` does.

Yay, another trust issued "solved"! But it's inefficient, and yet again more bloated boilerplate to weigh your project down.

That's just the story, over and over again, with callbacks. They can do pretty much anything you want, but you have to be willing to work hard to get it, and oftentimes this effort is much more than you can or should spend on such code reasoning.

You might find yourself wishing for built-in APIs or other language mechanics to address these issues. Finally ES6 has arrived on the scene with some great answers, so keep reading!

## Review

Callbacks are the fundamental unit of asynchrony in JS. But they're not enough for the evolving landscape of async programming as JS matures.

First, our brains plan things out in sequential, blocking, single-threaded semantic ways, but callbacks express asynchronous flow in a rather nonlinear, nonsequential way, which makes reasoning properly about such code much harder. Bad to reason about code is bad code that leads to bad bugs.

We need a way to express asynchrony in a more synchronous, sequential, blocking manner, just like our brains do.

Second, and more importantly, callbacks suffer from *inversion of control* in that they implicitly give control over to another party (often a third-party utility not in your control!) to invoke the *continuation* of your program. This control transfer leads us to a troubling list of trust issues, such as whether the callback is called more times than we expect.

Inventing ad hoc logic to solve these trust issues is possible, but it's more difficult than it should be, and it produces clunkier and harder to maintain code, as well as code that is likely insufficiently protected from these hazards until you get visibly bitten by the bugs.

We need a generalized solution to **all of the trust issues**, one that can be reused for as many callbacks as we create without all the extra boilerplate overhead.

We need something better than callbacks. They've served us well to this point, but the *future* of JavaScript demands more sophisticated and capable async patterns. The subsequent chapters in this book will dive into those emerging evolutions.

# You Don't Know JS Yet: Sync & Async - 2nd Edition

## Chapter 3: Promises

<b>NOTE:</b>
Work in progress

<b>NOTE:</b>
Everything below here is previous text from 1st edition, and is only here for reference while 2nd edition work is underway. <b>Please ignore this stuff.</b>

In Chapter 2, we identified two major categories of deficiencies with using callbacks to express program asynchrony and manage concurrency: lack of sequentiality and lack of trustability. Now that we understand the problems more intimately, it's time we turn our attention to patterns that can address them.

The issue we want to address first is the *inversion of control*, the trust that is so fragilely held and so easily lost.

Recall that we wrap up the *continuation* of our program in a callback function, and hand that callback over to another party (potentially even external code) and just cross our fingers that it will do the right thing with the invocation of the callback.

We do this because we want to say, "here's what happens *later*, after the current step finishes."

But what if we could uninvert that *inversion of control*? What if instead of handing the continuation of our program to another party, we could expect it to return us a capability to know when its task finishes, and then our code could decide what to do next?

This paradigm is called **Promises**.

Promises are starting to take the JS world by storm, as developers and specification writers alike desperately seek to untangle the insanity of callback hell in their code/design. In fact, most new async APIs being added to JS/DOM platform are being built on Promises. So it's probably a good idea to dig in and learn them, don't you think!?

**Note:** The word "immediately" will be used frequently in this chapter, generally to refer to some Promise resolution action. However, in essentially all cases, "immediately" means in terms of the Job queue behavior (see Chapter 1), not in the strictly synchronous *now* sense.

## What Is a Promise?

When developers decide to learn a new technology or pattern, usually their first step is "Show me the code!" It's quite natural for

When developers decide to learn a new technology or pattern, usually their first step is "Show me the code!" It's quite natural for us to just jump in feet first and learn as we go.

But it turns out that some abstractions get lost on the APIs alone. Promises are one of those tools where it can be painfully obvious from how someone uses it whether they understand what it's for and about versus just learning and using the API.

So before I show the Promise code, I want to fully explain what a Promise really is conceptually. I hope this will then guide you better as you explore integrating Promise theory into your own async flow.

With that in mind, let's look at two different analogies for what a Promise *is*.

## Future Value

Imagine this scenario: I walk up to the counter at a fast-food restaurant, and place an order for a cheeseburger. I hand the cashier \$1.47. By placing my order and paying for it, I've made a request for a *value* back (the cheeseburger). I've started a transaction.

But often, the cheeseburger is not immediately available for me. The cashier hands me something in place of my cheeseburger: a receipt with an order number on it. This order number is an IOU ("I owe you") *promise* that ensures that eventually, I should receive my cheeseburger.

So I hold onto my receipt and order number. I know it represents my *future cheeseburger*, so I don't need to worry about it anymore -- aside from being hungry!

While I wait, I can do other things, like send a text message to a friend that says, "Hey, can you come join me for lunch? I'm going to eat a cheeseburger."

I am reasoning about my *future cheeseburger* already, even though I don't have it in my hands yet. My brain is able to do this because it's treating the order number as a placeholder for the cheeseburger. The placeholder essentially makes the value *time independent*. It's a **future value**.

Eventually, I hear, "Order 113!" and I gleefully walk back up to the counter with receipt in hand. I hand my receipt to the cashier, and I take my cheeseburger in return.

In other words, once my *future value* was ready, I exchanged my value-promise for the value itself.

But there's another possible outcome. They call my order number, but when I go to retrieve my cheeseburger, the cashier regretfully informs me, "I'm sorry, but we appear to be all out of cheeseburgers." Setting aside the customer frustration of this scenario for a moment, we can see an important characteristic of *future values*: they can either indicate a success or failure.

Every time I order a cheeseburger, I know that I'll either get a cheeseburger eventually, or I'll get the sad news of the cheeseburger shortage, and I'll have to figure out something else to eat for lunch.

**Note:** In code, things are not quite as simple, because metaphorically the order number may never be called, in which case we're left indefinitely in an unresolved state. We'll come back to dealing with that case later.

## Values Now and Later

This all might sound too mentally abstract to apply to your code. So let's be more concrete.

However, before we can introduce how Promises work in this fashion, we're going to derive in code that we already understand -- callbacks! -- how to handle these *future values*.

When you write code to reason about a value, such as performing math on a `number`, whether you realize it or not, you've been assuming something very fundamental about that value, which is that it's a concrete *now value* already:

```
var x, y = 2;

console.log( x + y ); // NaN <-- because `x` isn't set yet
```

The `x + y` operation assumes both `x` and `y` are already set. In terms we'll expound on shortly, we assume the `x` and `y` values are already *resolved*.

It would be nonsense to expect that the `+` operator by itself would somehow be magically capable of detecting and waiting around until both `x` and `y` are resolved (aka ready), only then to do the operation. That would cause chaos in the program if different statements finished *now* and others finished *later*, right?

How could you possibly reason about the relationships between two statements if either one (or both) of them might not be finished yet? If statement 2 relies on statement 1 being finished, there are just two outcomes: either statement 1 finished right *now* and everything proceeds fine, or statement 1 didn't finish yet, and thus statement 2 is going to fail.

If this sort of thing sounds familiar from Chapter 1, good!

Let's go back to our `x + y` math operation. Imagine if there was a way to say, "Add `x` and `y`, but if either of them isn't ready yet, just wait until they are. Add them as soon as you can."

Your brain might have just jumped to callbacks. OK, so...

```
function add(getX, getY, cb) {
  var x, y;
  getX( function(xVal){
    x = xVal;
    // both are ready?
    if (y != undefined) {
      cb( x + y ); // send along sum
    }
  } );
  getY( function(yVal){
    y = yVal;
    // both are ready?
    if (x != undefined) {
      cb( x + y ); // send along sum
    }
  } );
}

// `fetchX()` and `fetchY()` are sync or async
// functions
add( fetchX, fetchY, function(sum){
  console.log( sum ); // that was easy, huh?
} );
```

Take just a moment to let the beauty (or lack thereof) of that snippet sink in (whistles patiently).

While the ugliness is undeniable, there's something very important to notice about this async pattern.

In that snippet, we treated `x` and `y` as future values, and we express an operation `add(..)` that (from the outside) does not care whether `x` or `y` or both are available right away or not. In other words, it normalizes the *now* and *later*, such that we can rely on a predictable outcome of the `add(..)` operation.

By using an `add(..)` that is temporally consistent -- it behaves the same across *now* and *later* times -- the async code is much easier to reason about.

To put it more plainly: to consistently handle both *now* and *later*, we make both of them *later*: all operations become async.

Of course, this rough callbacks-based approach leaves much to be desired. It's just a first tiny step toward realizing the benefits of reasoning about *future values* without worrying about the time aspect of when it's available or not.

## Promise Value

We'll definitely go into a lot more detail about Promises later in the chapter -- so don't worry if some of this is confusing -- but let's just briefly glimpse at how we can express the `x + y` example via `Promise s`:

```
function add(xPromise, yPromise) {
  // `Promise.all([ ... ])` takes an array of promises,
  // and returns a new promise that waits on them
  // all to finish
  return Promise.all( [xPromise, yPromise] )

  // when that promise is resolved, let's take the
  // received `X` and `Y` values and add them together.
  .then( function(values){
    // `values` is an array of the messages from the
    // previously resolved promises
    return values[0] + values[1];
  } );
}

// `fetchX()` and `fetchY()` return promises for
// their respective values, which may be ready
// *now* or *later*.
add( fetchX(), fetchY() )

// we get a promise back for the sum of those
// two numbers.
// now we chain-call `then(..)` to wait for the
// resolution of that returned promise.
.then( function(sum){
  console.log( sum ); // that was easier!
} );
```

There are two layers of Promises in this snippet.

`fetchX()` and `fetchY()` are called directly, and the values they return (promises!) are passed into `add(..)`. The underlying values those promises represent may be ready *now* or *later*, but each promise normalizes the behavior to be the same regardless. We reason about `x` and `y` values in a time-independent way. They are *future values*.

The second layer is the promise that `add(..)` creates (via `Promise.all([ ... ])`) and returns, which we wait on by calling `then(..)`. When the `add(..)` operation completes, our `sum` *future value* is ready and we can print it out. We hide inside of `add(..)` the logic for waiting on the `x` and `y` *future values*.

**Note:** Inside `add(..)`, the `Promise.all([ ... ])` call creates a promise (which is waiting on `promiseX` and `promiseY` to resolve). The chained call to `.then(..)` creates another promise, which the `return values[0] + values[1]` line immediately resolves (with the result of the addition). Thus, the `then(..)` call we chain off the end of the `add(..)` call -- at the end of the snippet -- is actually operating on that second promise returned, rather than the first one created by `Promise.all([ ... ])`. Also, though we are not chaining off the end of that second `then(..)`, it too has created another promise, had we chosen to observe/use it. This Promise chaining stuff will be explained in much greater detail later in this chapter.

Just like with cheeseburger orders, it's possible that the resolution of a Promise is rejection instead of fulfillment. Unlike a fulfilled Promise, where the value is always programmatic, a rejection value -- commonly called a "rejection reason" -- can either be set directly by the program logic, or it can result implicitly from a runtime exception.

With Promises, the `then(..)` call can actually take two functions, the first for fulfillment (as shown earlier), and the second for rejection:

```
add( fetchX(), fetchY() )
.then(
  // fulfillment handler
  function(sum) {
    console.log( sum );
  },
  // rejection handler
```

```
function(err) {
  console.error( err ); // bummer!
}
);
```

If something went wrong getting `x` or `y`, or something somehow failed during the addition, the promise that `add(..)` returns is rejected, and the second callback error handler passed to `then(..)` will receive the rejection value from the promise.

Because Promises encapsulate the time-dependent state -- waiting on the fulfillment or rejection of the underlying value -- from the outside, the Promise itself is time-independent, and thus Promises can be composed (combined) in predictable ways regardless of the timing or outcome underneath.

Moreover, once a Promise is resolved, it stays that way forever -- it becomes an *immutable value* at that point -- and can then be *observed* as many times as necessary.

**Note:** Because a Promise is externally immutable once resolved, it's now safe to pass that value around to any party and know that it cannot be modified accidentally or maliciously. This is especially true in relation to multiple parties observing the resolution of a Promise. It is not possible for one party to affect another party's ability to observe Promise resolution. Immutability may sound like an academic topic, but it's actually one of the most fundamental and important aspects of Promise design, and shouldn't be casually passed over.

That's one of the most powerful and important concepts to understand about Promises. With a fair amount of work, you could ad hoc create the same effects with nothing but ugly callback composition, but that's not really an effective strategy, especially because you have to do it over and over again.

Promises are an easily repeatable mechanism for encapsulating and composing *future values*.

## Completion Event

As we just saw, an individual Promise behaves as a *future value*. But there's another way to think of the resolution of a Promise: as a flow-control mechanism -- a temporal this-then-that -- for two or more steps in an asynchronous task.

Let's imagine calling a function `foo(..)` to perform some task. We don't know about any of its details, nor do we care. It may complete the task right away, or it may take a while.

We just simply need to know when `foo(..)` finishes so that we can move on to our next task. In other words, we'd like a way to be notified of `foo(..)`'s completion so that we can *continue*.

In typical JavaScript fashion, if you need to listen for a notification, you'd likely think of that in terms of events. So we could reframe our need for notification as a need to listen for a *completion* (or *continuation*) event emitted by `foo(..)`.

**Note:** Whether you call it a "completion event" or a "continuation event" depends on your perspective. Is the focus more on what happens with `foo(..)`, or what happens *after* `foo(..)` finishes? Both perspectives are accurate and useful. The event notification tells us that `foo(..)` has *completed*, but also that it's OK to *continue* with the next step. Indeed, the callback you pass to be called for the event notification is itself what we've previously called a *continuation*. Because *completion event* is a bit more focused on the `foo(..)`, which more has our attention at present, we slightly favor *completion event* for the rest of this text.

With callbacks, the "notification" would be our callback invoked by the task (`foo(..)`). But with Promises, we turn the relationship around, and expect that we can listen for an event from `foo(..)`, and when notified, proceed accordingly.

First, consider some pseudocode:

```
foo(x) {
  // start doing something that could take a while
}

foo( 42 )
```

```

on (foo "completion") {
    // now we can do the next step!
}

on (foo "error") {
    // oops, something went wrong in `foo(..)`
}

```

We call `foo(..)` and then we set up two event listeners, one for `"completion"` and one for `"error"` -- the two possible *final* outcomes of the `foo(..)` call. In essence, `foo(..)` doesn't even appear to be aware that the calling code has subscribed to these events, which makes for a very nice *separation of concerns*.

Unfortunately, such code would require some "magic" of the JS environment that doesn't exist (and would likely be a bit impractical). Here's the more natural way we could express that in JS:

```

function foo(x) {
    // start doing something that could take a while

    // make a `listener` event notification
    // capability to return

    return listener;
}

var evt = foo( 42 );

evt.on( "completion", function(){
    // now we can do the next step!
} );

evt.on( "failure", function(err){
    // oops, something went wrong in `foo(..)`
} );

```

`foo(..)` expressly creates an event subscription capability to return back, and the calling code receives and registers the two event handlers against it.

The inversion from normal callback-oriented code should be obvious, and it's intentional. Instead of passing the callbacks to `foo(..)`, it returns an event capability we call `evt`, which receives the callbacks.

But if you recall from Chapter 2, callbacks themselves represent an *inversion of control*. So inverting the callback pattern is actually an *inversion of inversion*, or an *uninversion of control* -- restoring control back to the calling code where we wanted it to be in the first place.

One important benefit is that multiple separate parts of the code can be given the event listening capability, and they can all independently be notified of when `foo(..)` completes to perform subsequent steps after its completion:

```

var evt = foo( 42 );

// let `bar(..)` listen to `foo(..)`'s completion
bar( evt );

// also, let `baz(..)` listen to `foo(..)`'s completion
baz( evt );

```

*Uninversion of control* enables a nicer *separation of concerns*, where `bar(..)` and `baz(..)` don't need to be involved in how `foo(..)` is called. Similarly, `foo(..)` doesn't need to know or care that `bar(..)` and `baz(..)` exist or are waiting to be notified when `foo(..)` completes.

Essentially, this `evt` object is a neutral third-party negotiation between the separate concerns.

## Promise "Events"

As you may have guessed by now, the `evt` event listening capability is an analogy for a Promise.

In a Promise-based approach, the previous snippet would have `foo(..)` creating and returning a `Promise` instance, and that promise would then be passed to `bar(..)` and `baz(..)`.

**Note:** The Promise resolution "events" we listen for aren't strictly events (though they certainly behave like events for these purposes), and they're not typically called "completion" or "error". Instead, we use `then(..)` to register a "then" event. Or perhaps more precisely, `then(..)` registers "fulfillment" and/or "rejection" event(s), though we don't see those terms used explicitly in the code.

Consider:

```
function foo(x) {
    // start doing something that could take a while

    // construct and return a promise
    return new Promise( function(resolve,reject){
        // eventually, call `resolve(..)` or `reject(..)`,
        // which are the resolution callbacks for
        // the promise.
    } );
}

var p = foo( 42 );

bar( p );
baz( p );
```

**Note:** The pattern shown with `new Promise( function(){ .. } )` is generally called the "[revealing constructor](#)". The function passed in is executed immediately (not async deferred, as callbacks to `then(..)` are), and it's provided two parameters, which in this case we've named `resolve` and `reject`. These are the resolution functions for the promise. `resolve(..)` generally signals fulfillment, and `reject(..)` signals rejection.

You can probably guess what the internals of `bar(..)` and `baz(..)` might look like:

```
function bar(fooPromise) {
    // listen for `foo(..)` to complete
    fooPromise.then(
        function(){
            // `foo(..)` has now finished, so
            // do `bar(..)`'s task
        },
        function(){
            // oops, something went wrong in `foo(..)`
        }
    );
}

// ditto for `baz(..)`
```

Promise resolution doesn't necessarily need to involve sending along a message, as it did when we were examining Promises as *future values*. It can just be a flow-control signal, as used in the previous snippet.

Another way to approach this is:

```
function bar() {
    // `foo(..)` has definitely finished, so
    // do `bar(..)`'s task
}
```

```

function oopsBar() {
  // oops, something went wrong in `foo(..)`,
  // so `bar(..)` didn't run
}

// ditto for `baz()` and `oopsBaz()`

var p = foo( 42 );

p.then( bar, oopsBar );

p.then( baz, oopsBaz );

```

**Note:** If you've seen Promise-based coding before, you might be tempted to believe that the last two lines of that code could be written as `p.then( ... ).then( ... )`, using chaining, rather than `p.then(...); p.then(...)`. That would have an entirely different behavior, so be careful! The difference might not be clear right now, but it's actually a different async pattern than we've seen thus far: splitting/forking. Don't worry! We'll come back to this point later in this chapter.

Instead of passing the `p` promise to `bar(..)` and `baz(..)`, we use the promise to control when `bar(..)` and `baz(..)` will get executed, if ever. The primary difference is in the error handling.

In the first snippet's approach, `bar(..)` is called regardless of whether `foo(..)` succeeds or fails, and it handles its own fallback logic if it's notified that `foo(..)` failed. The same is true for `baz(..)`, obviously.

In the second snippet, `bar(..)` only gets called if `foo(..)` succeeds, and otherwise `oopsBar(..)` gets called. Ditto for `baz(..)`.

Neither approach is *correct* per se. There will be cases where one is preferred over the other.

In either case, the promise `p` that comes back from `foo(..)` is used to control what happens next.

Moreover, the fact that both snippets end up calling `then(..)` twice against the same promise `p` illustrates the point made earlier, which is that Promises (once resolved) retain their same resolution (fulfillment or rejection) forever, and can subsequently be observed as many times as necessary.

Whenever `p` is resolved, the next step will always be the same, both *now* and *later*.

## Thenable Duck Typing

In Promises-land, an important detail is how to know for sure if some value is a genuine Promise or not. Or more directly, is it a value that will behave like a Promise?

Given that Promises are constructed by the `new Promise(..)` syntax, you might think that `p instanceof Promise` would be an acceptable check. But unfortunately, there are a number of reasons that's not totally sufficient.

Mainly, you can receive a Promise value from another browser window (iframe, etc.), which would have its own Promise different from the one in the current window/frame, and that check would fail to identify the Promise instance.

Moreover, a library or framework may choose to vend its own Promises and not use the native ES6 `Promise` implementation to do so. In fact, you may very well be using Promises with libraries in older browsers that have no Promise at all.

When we discuss Promise resolution processes later in this chapter, it will become more obvious why a non-genuine-but-Promise-like value would still be very important to be able to recognize and assimilate. But for now, just take my word for it that it's a critical piece of the puzzle.

As such, it was decided that the way to recognize a Promise (or something that behaves like a Promise) would be to define something called a "thenable" as any object or function which has a `then(..)` method on it. It is assumed that any such value is a Promise-conforming thenable.

The general term for "type checks" that make assumptions about a value's "type" based on its shape (what properties are present) is called "duck typing" -- "If it looks like a duck, and quacks like a duck, it must be a duck" (see the *Types & Grammar* title of this book series). So the duck typing check for a thenable would roughly be:

```
if (
  p !== null &&
  (
    typeof p === "object" ||
    typeof p === "function"
  ) &&
  typeof p.then === "function"
) {
  // assume it's a thenable!
}
else {
  // not a thenable
}
```

Yuck! Setting aside the fact that this logic is a bit ugly to implement in various places, there's something deeper and more troubling going on.

If you try to fulfill a Promise with any object/function value that happens to have a `then(..)` function on it, but you weren't intending it to be treated as a Promise/thenable, you're out of luck, because it will automatically be recognized as thenable and treated with special rules (see later in the chapter).

This is even true if you didn't realize the value has a `then(..)` on it. For example:

```
var o = { then: function(){ } };

// make `v` be `[[Prototype]]`-linked to `o`
var v = Object.create( o );

v.someStuff = "cool";
v.otherStuff = "not so cool";

v.hasOwnProperty( "then" );           // false
```

`v` doesn't look like a Promise or thenable at all. It's just a plain object with some properties on it. You're probably just intending to send that value around like any other object.

But unknown to you, `v` is also `[[Prototype]]`-linked (see the *this & Object Prototypes* title of this book series) to another object `o`, which happens to have a `then(..)` on it. So the thenable duck typing checks will think and assume `v` is a thenable. Uh oh.

It doesn't even need to be something as directly intentional as that:

```
Object.prototype.then = function(){};
Array.prototype.then = function(){};

var v1 = { hello: "world" };
var v2 = [ "Hello", "World" ];
```

Both `v1` and `v2` will be assumed to be thenables. You can't control or predict if any other code accidentally or maliciously adds `then(..)` to `Object.prototype`, `Array.prototype`, or any of the other native prototypes. And if what's specified is a function that doesn't call either of its parameters as callbacks, then any Promise resolved with such a value will just silently hang forever! Crazy.

Sound implausible or unlikely? Perhaps.

But keep in mind that there were several well-known non-Promise libraries preexisting in the community prior to ES6 that happened to already have a method on them called `then()`. Some of those libraries chose to rename their own methods to avoid collision (that sucks!). Others have simply been relegated to the unfortunate status of "incompatible with Promise-based coding" in reward for their inability to change to get out of the way.

The standards decision to hijack the previously nonreserved -- and completely general-purpose sounding -- `then` property name means that no value (or any of its delegates), either past, present, or future, can have a `then()` function present, either on purpose or by accident, or that value will be confused for a thenable in Promises systems, which will probably create bugs that are really hard to track down.

**Warning:** I do not like how we ended up with duck typing of thenables for Promise recognition. There were other options, such as "branding" or even "anti-branding"; what we got seems like a worst-case compromise. But it's not all doom and gloom. Thenable duck typing can be helpful, as we'll see later. Just beware that thenable duck typing can be hazardous if it incorrectly identifies something as a Promise that isn't.

## Promise Trust

We've now seen two strong analogies that explain different aspects of what Promises can do for our async code. But if we stop there, we've missed perhaps the single most important characteristic that the Promise pattern establishes: trust.

Whereas the *future values* and *completion events* analogies play out explicitly in the code patterns we've explored, it won't be entirely obvious why or how Promises are designed to solve all of the *inversion of control* trust issues we laid out in the "Trust Issues" section of Chapter 2. But with a little digging, we can uncover some important guarantees that restore the confidence in async coding that Chapter 2 tore down!

Let's start by reviewing the trust issues with callbacks-only coding. When you pass a callback to a utility `foo(..)`, it might:

- Call the callback too early
- Call the callback too late (or never)
- Call the callback too few or too many times
- Fail to pass along any necessary environment/parameters
- Swallow any errors/exceptions that may happen

The characteristics of Promises are intentionally designed to provide useful, repeatable answers to all these concerns.

### Calling Too Early

Primarily, this is a concern of whether code can introduce Zalgo-like effects (see Chapter 2), where sometimes a task finishes synchronously and sometimes asynchronously, which can lead to race conditions.

Promises by definition cannot be susceptible to this concern, because even an immediately fulfilled Promise (like `new Promise(function(resolve){ resolve(42); })`) cannot be *observed* synchronously.

That is, when you call `then()` on a Promise, even if that Promise was already resolved, the callback you provide to `then()` will **always** be called asynchronously (for more on this, refer back to "Jobs" in Chapter 1).

No more need to insert your own `setTimeout(..,0)` hacks. Promises prevent Zalgo automatically.

### Calling Too Late

Similar to the previous point, a Promise's `then()` registered observation callbacks are automatically scheduled when either `resolve(..)` or `reject(..)` are called by the Promise creation capability. Those scheduled callbacks will predictably be fired at the next asynchronous moment (see "Jobs" in Chapter 1).

It's not possible for synchronous observation, so it's not possible for a synchronous chain of tasks to run in such a way to in effect "delay" another callback from happening as expected. That is, when a Promise is resolved, all `then(...)` registered callbacks on it will be called, in order, immediately at the next asynchronous opportunity (again, see "Jobs" in Chapter 1), and nothing that happens inside of one of those callbacks can affect/delay the calling of the other callbacks.

For example:

```
p.then( function(){
  p.then( function(){
    console.log( "C" );
  } );
  console.log( "A" );
} );
p.then( function(){
  console.log( "B" );
} );
// A B C
```

Here, `"C"` cannot interrupt and precede `"B"`, by virtue of how Promises are defined to operate.

## Promises Scheduling Quirks

It's important to note, though, that there are lots of nuances of scheduling where the relative ordering between callbacks chained off two separate Promises is not reliably predictable.

If two promises `p1` and `p2` are both already resolved, it should be true that `p1.then(...); p2.then(...)` would end up calling the callback(s) for `p1` before the ones for `p2`. But there are subtle cases where that might not be true, such as the following:

```
var p3 = new Promise( function(resolve,reject){
  resolve( "B" );
} );

var p1 = new Promise( function(resolve,reject){
  resolve( p3 );
} );

var p2 = new Promise( function(resolve,reject){
  resolve( "A" );
} );

p1.then( function(v){
  console.log( v );
} );

p2.then( function(v){
  console.log( v );
} );

// A B <-- not B A as you might expect
```

We'll cover this more later, but as you can see, `p1` is resolved not with an immediate value, but with another promise `p3` which is itself resolved with the value `"B"`. The specified behavior is to *unwrap* `p3` into `p1`, but asynchronously, so `p1`'s callback(s) are *behind* `p2`'s callback(s) in the asynchronous Job queue (see Chapter 1).

To avoid such nuanced nightmares, you should never rely on anything about the ordering/scheduling of callbacks across Promises. In fact, a good practice is not to code in such a way where the ordering of multiple callbacks matters at all. Avoid that if you can.

## Never Calling the Callback

This is a very common concern. It's addressable in several ways with Promises.

First, nothing (not even a JS error) can prevent a Promise from notifying you of its resolution (if it's resolved). If you register both fulfillment and rejection callbacks for a Promise, and the Promise gets resolved, one of the two callbacks will always be called.

Of course, if your callbacks themselves have JS errors, you may not see the outcome you expect, but the callback will in fact have been called. We'll cover later how to be notified of an error in your callback, because even those don't get swallowed.

But what if the Promise itself never gets resolved either way? Even that is a condition that Promises provide an answer for, using a higher level abstraction called a "race":

```
// a utility for timing out a Promise
function timeoutPromise(delay) {
  return new Promise( function(resolve,reject){
    setTimeout( function(){
      reject( "Timeout!" );
    }, delay );
  } );
}

// setup a timeout for `foo()`
Promise.race( [
  foo(),           // attempt `foo()`
  timeoutPromise( 3000 ) // give it 3 seconds
] )
.then(
  function(){
    // `foo(..)` fulfilled in time!
  },
  function(err){
    // either `foo()` rejected, or it just
    // didn't finish in time, so inspect
    // `err` to know which
  }
);

```

There are more details to consider with this Promise timeout pattern, but we'll come back to it later.

Importantly, we can ensure a signal as to the outcome of `foo()`, to prevent it from hanging our program indefinitely.

## Calling Too Few or Too Many Times

By definition, *one* is the appropriate number of times for the callback to be called. The "too few" case would be zero calls, which is the same as the "never" case we just examined.

The "too many" case is easy to explain. Promises are defined so that they can only be resolved once. If for some reason the Promise creation code tries to call `resolve(..)` or `reject(..)` multiple times, or tries to call both, the Promise will accept only the first resolution, and will silently ignore any subsequent attempts.

Because a Promise can only be resolved once, any `then(..)` registered callbacks will only ever be called once (each).

Of course, if you register the same callback more than once, (e.g., `p.then(f); p.then(f);` ), it'll be called as many times as it was registered. The guarantee that a response function is called only once does not prevent you from shooting yourself in the foot.

## Failing to Pass Along Any Parameters/Environment

Promises can have, at most, one resolution value (fulfillment or rejection).

If you don't explicitly resolve with a value either way, the value is `undefined`, as is typical in JS. But whatever the value, it will always be passed to all registered (and appropriate: fulfillment or rejection) callbacks, either *now* or in the future.

Something to be aware of: If you call `resolve(..)` or `reject(..)` with multiple parameters, all subsequent parameters beyond the first will be silently ignored. Although that might seem a violation of the guarantee we just described, it's not exactly, because it constitutes an invalid usage of the Promise mechanism. Other invalid usages of the API (such as calling `resolve(..)` multiple times) are similarly *protected*, so the Promise behavior here is consistent (if not a tiny bit frustrating).

If you want to pass along multiple values, you must wrap them in another single value that you pass, such as an `array` or an `object`.

As for environment, functions in JS always retain their closure of the scope in which they're defined (see the *Scope & Closures* title of this series), so they of course would continue to have access to whatever surrounding state you provide. Of course, the same is true of callbacks-only design, so this isn't a specific augmentation of benefit from Promises -- but it's a guarantee we can rely on nonetheless.

## Swallowing Any Errors/Exceptions

In the base sense, this is a restatement of the previous point. If you reject a Promise with a *reason* (aka error message), that value is passed to the rejection callback(s).

But there's something much bigger at play here. If at any point in the creation of a Promise, or in the observation of its resolution, a JS exception error occurs, such as a `TypeError` or `ReferenceError`, that exception will be caught, and it will force the Promise in question to become rejected.

For example:

```
var p = new Promise( function(resolve,reject){
  foo.bar();    // `foo` is not defined, so error!
  resolve( 42 );    // never gets here :(
}

p.then(
  function fulfilled(){
    // never gets here :(
  },
  function rejected(err){
    // `err` will be a `TypeError` exception object
    // from the `foo.bar()` line.
  }
);
```

The JS exception that occurs from `foo.bar()` becomes a Promise rejection that you can catch and respond to.

This is an important detail, because it effectively solves another potential Zalgo moment, which is that errors could create a synchronous reaction whereas nonerrors would be asynchronous. Promises turn even JS exceptions into asynchronous behavior, thereby reducing the race condition chances greatly.

But what happens if a Promise is fulfilled, but there's a JS exception error during the observation (in a `then(..)` registered callback)? Even those aren't lost, but you may find how they're handled a bit surprising, until you dig in a little deeper:

```
var p = new Promise( function(resolve,reject){
  resolve( 42 );
}

p.then(
  function fulfilled(msg){
    foo.bar();
    console.log( msg );    // never gets here :(
  },
  function rejected(err){
    // never gets here either :(
  }
);
```

```
});
```

Wait, that makes it seem like the exception from `foo.bar()` really did get swallowed. Never fear, it didn't. But something deeper is wrong, which is that we've failed to listen for it. The `p.then(..)` call itself returns another promise, and it's *that* promise that will be rejected with the `TypeError` exception.

Why couldn't it just call the error handler we have defined there? Seems like a logical behavior on the surface. But it would violate the fundamental principle that Promises are **immutable** once resolved. `p` was already fulfilled to the value `42`, so it can't later be changed to a rejection just because there's an error in observing `p`'s resolution.

Besides the principle violation, such behavior could wreak havoc, if say there were multiple `then(..)` registered callbacks on the promise `p`, because some would get called and others wouldn't, and it would be very opaque as to why.

## Trustable Promise?

There's one last detail to examine to establish trust based on the Promise pattern.

You've no doubt noticed that Promises don't get rid of callbacks at all. They just change where the callback is passed to. Instead of passing a callback to `foo(..)`, we get *something* (ostensibly a genuine Promise) back from `foo(..)`, and we pass the callback to that *something* instead.

But why would this be any more trustable than just callbacks alone? How can we be sure the *something* we get back is in fact a trustable Promise? Isn't it basically all just a house of cards where we can trust only because we already trusted?

One of the most important, but often overlooked, details of Promises is that they have a solution to this issue as well. Included with the native ES6 `Promise` implementation is `Promise.resolve(..)`.

If you pass an immediate, non-Promise, non-thenable value to `Promise.resolve(..)`, you get a promise that's fulfilled with that value. In other words, these two promises `p1` and `p2` will behave basically identically:

```
var p1 = new Promise( function(resolve,reject){
  resolve( 42 );
} );
var p2 = Promise.resolve( 42 );
```

But if you pass a genuine Promise to `Promise.resolve(..)`, you just get the same promise back:

```
var p1 = Promise.resolve( 42 );
var p2 = Promise.resolve( p1 );
p1 === p2; // true
```

Even more importantly, if you pass a non-Promise thenable value to `Promise.resolve(..)`, it will attempt to unwrap that value, and the unwrapping will keep going until a concrete final non-Promise-like value is extracted.

Recall our previous discussion of thenables?

Consider:

```
var p = {
  then: function(cb) {
    cb( 42 );
  }
};

// this works OK, but only by good fortune
p
```

```
.then(
  function fulfilled(val){
    console.log( val ); // 42
  },
  function rejected(err){
    // never gets here
  }
);
```

This `p` is a thenable, but it's not a genuine Promise. Luckily, it's reasonable, as most will be. But what if you got back instead something that looked like:

```
var p = {
  then: function(cb,errcb) {
    cb( 42 );
    errcb( "evil laugh" );
  }
};

p
.then(
  function fulfilled(val){
    console.log( val ); // 42
  },
  function rejected(err){
    // oops, shouldn't have run
    console.log( err ); // evil laugh
  }
);
```

This `p` is a thenable but it's not so well behaved of a promise. Is it malicious? Or is it just ignorant of how Promises should work? It doesn't really matter, to be honest. In either case, it's not trustable as is.

Nonetheless, we can pass either of these versions of `p` to `Promise.resolve(..)`, and we'll get the normalized, safe result we'd expect:

```
Promise.resolve( p )
.then(
  function fulfilled(val){
    console.log( val ); // 42
  },
  function rejected(err){
    // never gets here
  }
);
```

`Promise.resolve(..)` will accept any thenable, and will unwrap it to its non-thenable value. But you get back from `Promise.resolve(..)` a real, genuine Promise in its place, **one that you can trust**. If what you passed in is already a genuine Promise, you just get it right back, so there's no downside at all to filtering through `Promise.resolve(..)` to gain trust.

So let's say we're calling a `foo(..)` utility and we're not sure we can trust its return value to be a well-behaving Promise, but we know it's at least a thenable. `Promise.resolve(..)` will give us a trustable Promise wrapper to chain off of:

```
// don't just do this:
foo( 42 )
.then( function(v){
  console.log( v );
} );

// instead, do this:
Promise.resolve( foo( 42 ) )
.then( function(v){
```

```
    console.log( v );
} );
```

**Note:** Another beneficial side effect of wrapping `Promise.resolve(..)` around any function's return value (thenable or not) is that it's an easy way to normalize that function call into a well-behaving async task. If `foo(42)` returns an immediate value sometimes, or a Promise other times, `Promise.resolve( foo(42) )` makes sure it's always a Promise result. And avoiding Zalgo makes for much better code.

## Trust Built

Hopefully the previous discussion now fully "resolves" (pun intended) in your mind why the Promise is trustable, and more importantly, why that trust is so critical in building robust, maintainable software.

Can you write async code in JS without trust? Of course you can. We JS developers have been coding async with nothing but callbacks for nearly two decades.

But once you start questioning just how much you can trust the mechanisms you build upon to actually be predictable and reliable, you start to realize callbacks have a pretty shaky trust foundation.

Promises are a pattern that augments callbacks with trustable semantics, so that the behavior is more reason-able and more reliable. By uninverting the *inversion of control* of callbacks, we place the control with a trustable system (Promises) that was designed specifically to bring sanity to our async.

## Chain Flow

We've hinted at this a couple of times already, but Promises are not just a mechanism for a single-step *this-then-that* sort of operation. That's the building block, of course, but it turns out we can string multiple Promises together to represent a sequence of async steps.

The key to making this work is built on two behaviors intrinsic to Promises:

- Every time you call `then(..)` on a Promise, it creates and returns a new Promise, which we can *chain* with.
- Whatever value you return from the `then(..)` call's fulfillment callback (the first parameter) is automatically set as the fulfillment of the *chained* Promise (from the first point).

Let's first illustrate what that means, and *then* we'll derive how that helps us create async sequences of flow control. Consider the following:

```
var p = Promise.resolve( 21 );

var p2 = p.then( function(v){
  console.log( v );      // 21

  // fulfill `p2` with value `42`
  return v * 2;
} );

// chain off `p2`
p2.then( function(v){
  console.log( v );      // 42
} );
```

By returning `v * 2` (i.e., `42`), we fulfill the `p2` promise that the first `then(..)` call created and returned. When `p2`'s `then(..)` call runs, it's receiving the fulfillment from the `return v * 2` statement. Of course, `p2.then(..)` creates yet another promise, which we could have stored in a `p3` variable.

But it's a little annoying to have to create an intermediate variable `p2` (or `p3`, etc.). Thankfully, we can easily just chain these together:

```
var p = Promise.resolve( 21 );

p
.then( function(v){
  console.log( v ); // 21

  // fulfill the chained promise with value `42`
  return v * 2;
} )
// here's the chained promise
.then( function(v){
  console.log( v ); // 42
} );
```

So now the first `then(..)` is the first step in an async sequence, and the second `then(..)` is the second step. This could keep going for as long as you needed it to extend. Just keep chaining off a previous `then(..)` with each automatically created Promise.

But there's something missing here. What if we want step 2 to wait for step 1 to do something asynchronous? We're using an immediate `return` statement, which immediately fulfills the chained promise.

The key to making a Promise sequence truly async capable at every step is to recall how `Promise.resolve(..)` operates when what you pass to it is a Promise or thenable instead of a final value. `Promise.resolve(..)` directly returns a received genuine Promise, or it unwraps the value of a received thenable -- and keeps going recursively while it keeps unwrapping thenables.

The same sort of unwrapping happens if you `return` a thenable or Promise from the fulfillment (or rejection) handler. Consider:

```
var p = Promise.resolve( 21 );

p.then( function(v){
  console.log( v ); // 21

  // create a promise and return it
  return new Promise( function(resolve,reject){
    // fulfill with value `42`
    resolve( v * 2 );
  } );
} )
.then( function(v){
  console.log( v ); // 42
} );
```

Even though we wrapped `42` up in a promise that we returned, it still got unwrapped and ended up as the resolution of the chained promise, such that the second `then(..)` still received `42`. If we introduce asynchrony to that wrapping promise, everything still nicely works the same:

```
var p = Promise.resolve( 21 );

p.then( function(v){
  console.log( v ); // 21

  // create a promise to return
  return new Promise( function(resolve,reject){
    // introduce asynchrony!
    setTimeout( function(){
      // fulfill with value `42`
      resolve( v * 2 );
    }, 100 );
  } );
} );
```

```

} )
.then( function(v){
    // runs after the 100ms delay in the previous step
    console.log( v );    // 42
} );

```

That's incredibly powerful! Now we can construct a sequence of however many async steps we want, and each step can delay the next step (or not!), as necessary.

Of course, the value passing from step to step in these examples is optional. If you don't return an explicit value, an implicit `undefined` is assumed, and the promises still chain together the same way. Each Promise resolution is thus just a signal to proceed to the next step.

To further the chain illustration, let's generalize a delay-Promise creation (without resolution messages) into a utility we can reuse for multiple steps:

```

function delay(time) {
    return new Promise( function(resolve,reject){
        setTimeout( resolve, time );
    } );
}

delay( 100 ) // step 1
.then( function STEP2(){
    console.log( "step 2 (after 100ms)" );
    return delay( 200 );
} )
.then( function STEP3(){
    console.log( "step 3 (after another 200ms)" );
} )
.then( function STEP4(){
    console.log( "step 4 (next Job)" );
    return delay( 50 );
} )
.then( function STEP5(){
    console.log( "step 5 (after another 50ms)" );
} )
...

```

Calling `delay(200)` creates a promise that will fulfill in 200ms, and then we return that from the first `then(..)` fulfillment callback, which causes the second `then(..)`'s promise to wait on that 200ms promise.

**Note:** As described, technically there are two promises in that interchange: the 200ms-delay promise and the chained promise that the second `then(..)` chains from. But you may find it easier to mentally combine these two promises together, because the Promise mechanism automatically merges their states for you. In that respect, you could think of `return delay(200)` as creating a promise that replaces the earlier-returned chained promise.

To be honest, though, sequences of delays with no message passing isn't a terribly useful example of Promise flow control. Let's look at a scenario that's a little more practical.

Instead of timers, let's consider making Ajax requests:

```

// assume an `ajax( {url}, {callback} )` utility

// Promise-aware ajax
function request(url) {
    return new Promise( function(resolve,reject){
        // the `ajax(..)` callback should be our
        // promise's `resolve(..)` function
        ajax( url, resolve );
    } );
}

```

We first define a `request(...)` utility that constructs a promise to represent the completion of the `ajax(...)` call:

```
request( "http://some.url.1/" )
  .then( function(response1){
    return request( "http://some.url.2/?v=" + response1 );
  })
  .then( function(response2){
    console.log( response2 );
  });
}
```

**Note:** Developers commonly encounter situations in which they want to do Promise-aware async flow control with utilities that are not themselves Promise-enabled (like `ajax(...)` here, which expects a callback). Although the native ES6 `Promise` mechanism doesn't automatically solve this pattern for us, practically all Promise libraries *do*. They usually call this process "lifting" or "promisifying" or some variation thereof. We'll come back to this technique later.

Using the Promise-returning `request(...)`, we create the first step in our chain implicitly by calling it with the first URL, and chain off that returned promise with the first `then(...)`.

Once `response1` comes back, we use that value to construct a second URL, and make a second `request(...)` call. That second `request(...)` promise is `return`ed so that the third step in our async flow control waits for that Ajax call to complete. Finally, we print `response2` once it returns.

The Promise chain we construct is not only a flow control that expresses a multistep async sequence, but it also acts as a message channel to propagate messages from step to step.

What if something went wrong in one of the steps of the Promise chain? An error/exception is on a per-Promise basis, which means it's possible to catch such an error at any point in the chain, and that catching acts to sort of "reset" the chain back to normal operation at that point:

```
// step 1:
request( "http://some.url.1/" )

// step 2:
.then( function(response1){
  foo.bar(); // undefined, error!

  // never gets here
  return request( "http://some.url.2/?v=" + response1 );
})

// step 3:
.then(
  function fulfilled(response2){
    // never gets here
  },
  // rejection handler to catch the error
  function rejected(err){
    console.log( err ); // `TypeError` from `foo.bar()` error
    return 42;
  }
)

// step 4:
.then( function(msg){
  console.log( msg ); // 42
});
```

When the error occurs in step 2, the rejection handler in step 3 catches it. The return value (`42` in this snippet), if any, from that rejection handler fulfills the promise for the next step (4), such that the chain is now back in a fulfillment state.

**Note:** As we discussed earlier, when returning a promise from a fulfillment handler, it's unwrapped and can delay the next step. That's also true for returning promises from rejection handlers, such that if the `return 42` in step 3 instead returned a promise, that promise could delay step 4. A thrown exception inside either the fulfillment or rejection handler of a `then(..)` call causes the next (chained) promise to be immediately rejected with that exception.

If you call `then(..)` on a promise, and you only pass a fulfillment handler to it, an assumed rejection handler is substituted:

```
var p = new Promise( function(resolve,reject){
  reject( "Oops" );
}

var p2 = p.then(
  function fulfilled(){
    // never gets here
  }
  // assumed rejection handler, if omitted or
  // any other non-function value passed
  // function(err) {
  //   throw err;
  // }
);

```

As you can see, the assumed rejection handler simply rethrows the error, which ends up forcing `p2` (the chained promise) to reject with the same error reason. In essence, this allows the error to continue propagating along a Promise chain until an explicitly defined rejection handler is encountered.

**Note:** We'll cover more details of error handling with Promises a little later, because there are other nuanced details to be concerned about.

If a proper valid function is not passed as the fulfillment handler parameter to `then(..)`, there's also a default handler substituted:

```
var p = Promise.resolve( 42 );

p.then(
  // assumed fulfillment handler, if omitted or
  // any other non-function value passed
  // function(v) {
  //   return v;
  // }
  null,
  function rejected(err){
    // never gets here
  }
);

```

As you can see, the default fulfillment handler simply passes whatever value it receives along to the next step (Promise).

**Note:** The `then(null,function(err){ .. })` pattern -- only handling rejections (if any) but letting fulfillments pass through -- has a shortcut in the API: `catch(function(err){ .. })`. We'll cover `catch(..)` more fully in the next section.

Let's review briefly the intrinsic behaviors of Promises that enable chaining flow control:

- A `then(..)` call against one Promise automatically produces a new Promise to return from the call.
- Inside the fulfillment/rejection handlers, if you return a value or an exception is thrown, the new returned (chainable) Promise is resolved accordingly.
- If the fulfillment or rejection handler returns a Promise, it is unwrapped, so that whatever its resolution is will become the resolution of the chained Promise returned from the current `then(..)`.

While chaining flow control is helpful, it's probably most accurate to think of it as a side benefit of how Promises compose (combine) together, rather than the main intent. As we've discussed in detail several times already, Promises normalize asynchrony and encapsulate time-dependent value state, and *that* is what lets us chain them together in this useful way.

Certainly, the sequential expressiveness of the chain (this-then-this-then-this...) is a big improvement over the tangled mess of callbacks as we identified in Chapter 2. But there's still a fair amount of boilerplate (`then(..)` and `function(){ .. }`) to wade through. In the next chapter, we'll see a significantly nicer pattern for sequential flow control expressivity, with generators.

## Terminology: Resolve, Fulfill, and Reject

There's some slight confusion around the terms "resolve," "fulfill," and "reject" that we need to clear up, before you get too much deeper into learning about Promises. Let's first consider the `Promise(..)` constructor:

```
var p = new Promise( function(X,Y){
  // X() for fulfillment
  // Y() for rejection
});
```

As you can see, two callbacks (here labeled `X` and `Y`) are provided. The first is *usually* used to mark the Promise as fulfilled, and the second *always* marks the Promise as rejected. But what's the "usually" about, and what does that imply about accurately naming those parameters?

Ultimately, it's just your user code and the identifier names aren't interpreted by the engine to mean anything, so it doesn't *technically* matter; `foo(..)` and `bar(..)` are equally functional. But the words you use can affect not only how you are thinking about the code, but how other developers on your team will think about it. Thinking wrongly about carefully orchestrated async code is almost surely going to be worse than the spaghetti-callback alternatives.

So it actually does kind of matter what you call them.

The second parameter is easy to decide. Almost all literature uses `reject(..)` as its name, and because that's exactly (and only!) what it does, that's a very good choice for the name. I'd strongly recommend you always use `reject(..)`.

But there's a little more ambiguity around the first parameter, which in Promise literature is often labeled `resolve(..)`. That word is obviously related to "resolution," which is what's used across the literature (including this book) to describe setting a final value/state to a Promise. We've already used "resolve the Promise" several times to mean either fulfilling or rejecting the Promise.

But if this parameter seems to be used to specifically fulfill the Promise, why shouldn't we call it `fulfill(..)` instead of `resolve(..)` to be more accurate? To answer that question, let's also take a look at two of the `Promise` API methods:

```
var fulfilledPr = Promise.resolve( 42 );
var rejectedPr = Promise.reject( "Oops" );
```

`Promise.resolve(..)` creates a Promise that's resolved to the value given to it. In this example, `42` is a normal, non-Promise, non-thenable value, so the fulfilled promise `fulfilledPr` is created for the value `42`. `Promise.reject("Oops")` creates the rejected promise `rejectedPr` for the reason `"Oops"`.

Let's now illustrate why the word "resolve" (such as in `Promise.resolve(..)`) is unambiguous and indeed more accurate, if used explicitly in a context that could result in either fulfillment or rejection:

```
var rejectedTh = {
  then: function(resolved,rejected) {
    rejected( "Oops" );
  }
};

var rejectedPr = Promise.resolve( rejectedTh );
```

As we discussed earlier in this chapter, `Promise.resolve(..)` will return a received genuine Promise directly, or unwrap a received thenable. If that thenable unwrapping reveals a rejected state, the Promise returned from `Promise.resolve(..)` is in fact in that same rejected state.

So `Promise.resolve(..)` is a good, accurate name for the API method, because it can actually result in either fulfillment or rejection.

The first callback parameter of the `Promise(..)` constructor will unwrap either a thenable (identically to `Promise.resolve(..)`) or a genuine Promise:

```
var rejectedPr = new Promise( function(resolve,reject){
    // resolve this promise with a rejected promise
    resolve( Promise.reject( "Oops" ) );
} );

rejectedPr.then(
    function fulfilled(){
        // never gets here
    },
    function rejected(err){
        console.log( err );      // "Oops"
    }
);
```

It should be clear now that `resolve(..)` is the appropriate name for the first callback parameter of the `Promise(..)` constructor.

**Warning:** The previously mentioned `reject(..)` does **not** do the unwrapping that `resolve(..)` does. If you pass a Promise/thenable value to `reject(..)`, that untouched value will be set as the rejection reason. A subsequent rejection handler would receive the actual Promise/thenable you passed to `reject(..)`, not its underlying immediate value.

But now let's turn our attention to the callbacks provided to `then(..)`. What should they be called (both in literature and in code)? I would suggest `fulfilled(..)` and `rejected(..)`:

```
function fulfilled(msg) {
    console.log( msg );
}

function rejected(err) {
    console.error( err );
}

p.then(
    fulfilled,
    rejected
);
```

In the case of the first parameter to `then(..)`, it's unambiguously always the fulfillment case, so there's no need for the duality of "resolve" terminology. As a side note, the ES6 specification uses `onFulfilled(..)` and `onRejected(..)` to label these two callbacks, so they are accurate terms.

## Error Handling

We've already seen several examples of how Promise rejection -- either intentional through calling `reject(..)` or accidental through JS exceptions -- allows saner error handling in asynchronous programming. Let's circle back though and be explicit about some of the details that we glossed over.

The most natural form of error handling for most developers is the synchronous `try..catch` construct. Unfortunately, it's synchronous-only, so it fails to help in async code patterns:

```
function foo() {
  setTimeout( function(){
    baz.bar();
  }, 100 );
}

try {
  foo();
  // later throws global error from `baz.bar()`
}
catch (err) {
  // never gets here
}
```

`try..catch` would certainly be nice to have, but it doesn't work across async operations. That is, unless there's some additional environmental support, which we'll come back to with generators in Chapter 4.

In callbacks, some standards have emerged for patterned error handling, most notably the "error-first callback" style:

```
function foo(cb) {
  setTimeout( function(){
    try {
      var x = baz.bar();
      cb( null, x ); // success!
    }
    catch (err) {
      cb( err );
    }
  }, 100 );
}

foo( function(err, val){
  if (err) {
    console.error( err ); // bummer :(
  }
  else {
    console.log( val );
  }
});
```

**Note:** The `try..catch` here works only from the perspective that the `baz.bar()` call will either succeed or fail immediately, synchronously. If `baz.bar()` was itself its own async completing function, any async errors inside it would not be catchable.

The callback we pass to `foo(..)` expects to receive a signal of an error by the reserved first parameter `err`. If present, error is assumed. If not, success is assumed.

This sort of error handling is technically *async capable*, but it doesn't compose well at all. Multiple levels of error-first callbacks woven together with these ubiquitous `if` statement checks inevitably will lead you to the perils of callback hell (see Chapter 2).

So we come back to error handling in Promises, with the rejection handler passed to `then(..)`. Promises don't use the popular "error-first callback" design style, but instead use "split callbacks" style; there's one callback for fulfillment and one for rejection:

```
var p = Promise.reject( "Oops" );

p.then(
  function fulfilled(){
    // never gets here
  },
  function rejected(err){
    console.log( err ); // "Oops"
```

```

    }
);

```

While this pattern of error handling makes fine sense on the surface, the nuances of Promise error handling are often a fair bit more difficult to fully grasp.

Consider:

```

var p = Promise.resolve( 42 );

p.then(
  function fulfilled(msg){
    // numbers don't have string functions,
    // so will throw an error
    console.log( msg.toLowerCase() );
  },
  function rejected(err){
    // never gets here
  }
);

```

If the `msg.toLowerCase()` legitimately throws an error (it does!), why doesn't our error handler get notified? As we explained earlier, it's because *that* error handler is for the `p` promise, which has already been fulfilled with value `42`. The `p` promise is immutable, so the only promise that can be notified of the error is the one returned from `p.then(...)`, which in this case we don't capture.

That should paint a clear picture of why error handling with Promises is error-prone (pun intended). It's far too easy to have errors swallowed, as this is very rarely what you'd intend.

**Warning:** If you use the Promise API in an invalid way and an error occurs that prevents proper Promise construction, the result will be an immediately thrown exception, **not a rejected Promise**. Some examples of incorrect usage that fail Promise construction: `new Promise(null)`, `Promise.all()`, `Promise.race(42)`, and so on. You can't get a rejected Promise if you don't use the Promise API validly enough to actually construct a Promise in the first place!

## Pit of Despair

Jeff Atwood noted years ago: programming languages are often set up in such a way that by default, developers fall into the "pit of despair" (<http://blog.codinghorror.com/falling-into-the-pit-of-success/>) -- where accidents are punished -- and that you have to try harder to get it right. He implored us to instead create a "pit of success," where by default you fall into expected (successful) action, and thus would have to try hard to fail.

Promise error handling is unquestionably "pit of despair" design. By default, it assumes that you want any error to be swallowed by the Promise state, and if you forget to observe that state, the error silently languishes/dies in obscurity -- usually despair.

To avoid losing an error to the silence of a forgotten/discarded Promise, some developers have claimed that a "best practice" for Promise chains is to always end your chain with a final `catch(...)`, like:

```

var p = Promise.resolve( 42 );

p.then(
  function fulfilled(msg){
    // numbers don't have string functions,
    // so will throw an error
    console.log( msg.toLowerCase() );
  }
)
.catch( handleErrors );

```

Because we didn't pass a rejection handler to the `then(..)`, the default handler was substituted, which simply propagates the error to the next promise in the chain. As such, both errors that come into `p`, and errors that come *after* `p` in its resolution (like the `msg.toLowerCase()` one) will filter down to the final `handleErrors(..)`.

Problem solved, right? Not so fast!

What happens if `handleErrors(..)` itself also has an error in it? Who catches that? There's still yet another unattended promise: the one `catch(..)` returns, which we don't capture and don't register a rejection handler for.

You can't just stick another `catch(..)` on the end of that chain, because it too could fail. The last step in any Promise chain, whatever it is, always has the possibility, even decreasingly so, of dangling with an uncaught error stuck inside an unobserved Promise.

Sound like an impossible conundrum yet?

## Uncaught Handling

It's not exactly an easy problem to solve completely. There are other ways to approach it which many would say are *better*.

Some Promise libraries have added methods for registering something like a "global unhandled rejection" handler, which would be called instead of a globally thrown error. But their solution for how to identify an error as "uncaught" is to have an arbitrary-length timer, say 3 seconds, running from time of rejection. If a Promise is rejected but no error handler is registered before the timer fires, then it's assumed that you won't ever be registering a handler, so it's "uncaught."

In practice, this has worked well for many libraries, as most usage patterns don't typically call for significant delay between Promise rejection and observation of that rejection. But this pattern is troublesome because 3 seconds is so arbitrary (even if empirical), and also because there are indeed some cases where you want a Promise to hold on to its rejectedness for some indefinite period of time, and you don't really want to have your "uncaught" handler called for all those false positives (not-yet-handled "uncaught errors").

Another more common suggestion is that Promises should have a `done(..)` added to them, which essentially marks the Promise chain as "done." `done(..)` doesn't create and return a Promise, so the callbacks passed to `done(..)` are obviously not wired up to report problems to a chained Promise that doesn't exist.

So what happens instead? It's treated as you might usually expect in uncaught error conditions: any exception inside a `done(..)` rejection handler would be thrown as a global uncaught error (in the developer console, basically):

```
var p = Promise.resolve( 42 );

p.then(
  function fulfilled(msg){
    // numbers don't have string functions,
    // so will throw an error
    console.log( msg.toLowerCase() );
  }
)
.done( null, handleErrors );

// if `handleErrors(..)` caused its own exception, it would
// be thrown globally here
```

This might sound more attractive than the never-ending chain or the arbitrary timeouts. But the biggest problem is that it's not part of the ES6 standard, so no matter how good it sounds, at best it's a lot longer way off from being a reliable and ubiquitous solution.

Are we just stuck, then? Not entirely.

Browsers have a unique capability that our code does not have: they can track and know for sure when any object gets thrown away and garbage collected. So, browsers can track Promise objects, and whenever they get garbage collected, if they have a rejection in them, the browser knows for sure this was a legitimate "uncaught error," and can thus confidently know it should report it to the developer console.

**Note:** At the time of this writing, both Chrome and Firefox have early attempts at that sort of "uncaught rejection" capability, though support is incomplete at best.

However, if a Promise doesn't get garbage collected -- it's exceedingly easy for that to accidentally happen through lots of different coding patterns -- the browser's garbage collection sniffing won't help you know and diagnose that you have a silently rejected Promise laying around.

Is there any other alternative? Yes.

## Pit of Success

The following is just theoretical, how Promises *could* be someday changed to behave. I believe it would be far superior to what we currently have. And I think this change would be possible even post-ES6 because I don't think it would break web compatibility with ES6 Promises. Moreover, it can be polyfilled/prollyfilled in, if you're careful. Let's take a look:

- Promises could default to reporting (to the developer console) any rejection, on the next Job or event loop tick, if at that exact moment no error handler has been registered for the Promise.
- For the cases where you want a rejected Promise to hold onto its rejected state for an indefinite amount of time before observing, you could call `defer()`, which suppresses automatic error reporting on that Promise.

If a Promise is rejected, it defaults to noisily reporting that fact to the developer console (instead of defaulting to silence). You can opt out of that reporting either implicitly (by registering an error handler before rejection), or explicitly (with `defer()`). In either case, *you* control the false positives.

Consider:

```
var p = Promise.reject("Oops").defer();

// `foo(..)` is Promise-aware
foo(42)
.then(
  function fulfilled(){
    return p;
  },
  function rejected(err){
    // handle `foo(..)` error
  }
);
...
...
```

When we create `p`, we know we're going to wait a while to use/observe its rejection, so we call `defer()` -- thus no global reporting. `defer()` simply returns the same promise, for chaining purposes.

The promise returned from `foo(..)` gets an error handler attached *right away*, so it's implicitly opted out and no global reporting for it occurs either.

But the promise returned from the `then(..)` call has no `defer()` or error handler attached, so if it rejects (from inside either resolution handler), then *it will* be reported to the developer console as an uncaught error.

**This design is a pit of success.** By default, all errors are either handled or reported -- what almost all developers in almost all cases would expect. You either have to register a handler or you have to intentionally opt out, and indicate you intend to defer error handling until *later*; you're opting for the extra responsibility in just that specific case.

The only real danger in this approach is if you `defer()` a Promise but then fail to actually ever observe/handle its rejection.

But you had to intentionally call `defer()` to opt into that pit of despair -- the default was the pit of success -- so there's not much else we could do to save you from your own mistakes.

I think there's still hope for Promise error handling (post-ES6). I hope the powers that be will rethink the situation and consider this alternative. In the meantime, you can implement this yourself (a challenging exercise for the reader!), or use a *smarter* Promise library that does so for you!

**Note:** This exact model for error handling/reporting is implemented in my *asyncquence* Promise abstraction library, which will be discussed in Appendix A of this book.

## Promise Patterns

We've already implicitly seen the sequence pattern with Promise chains (this-then-this-then-that flow control) but there are lots of variations on asynchronous patterns that we can build as abstractions on top of Promises. These patterns serve to simplify the expression of async flow control -- which helps make our code more reasonable and more maintainable -- even in the most complex parts of our programs.

Two such patterns are codified directly into the native ES6 `Promise` implementation, so we get them for free, to use as building blocks for other patterns.

### Promise.all([ .. ])

In an async sequence (Promise chain), only one async task is being coordinated at any given moment -- step 2 strictly follows step 1, and step 3 strictly follows step 2. But what about doing two or more steps concurrently (aka "in parallel")?

In classic programming terminology, a "gate" is a mechanism that waits on two or more parallel/concurrent tasks to complete before continuing. It doesn't matter what order they finish in, just that all of them have to complete for the gate to open and let the flow control through.

In the Promise API, we call this pattern `all([ .. ])`.

Say you wanted to make two Ajax requests at the same time, and wait for both to finish, regardless of their order, before making a third Ajax request. Consider:

```
// `request(..)` is a Promise-aware Ajax utility,
// like we defined earlier in the chapter

var p1 = request( "http://some.url.1/" );
var p2 = request( "http://some.url.2/" );

Promise.all( [p1,p2] )
.then( function(msgs){
    // both `p1` and `p2` fulfill and pass in
    // their messages here
    return request(
        "http://some.url.3/?v=" + msgs.join(", ")
    );
})
.then( function(msg){
    console.log( msg );
})
;
```

`Promise.all([ .. ])` expects a single argument, an `array`, consisting generally of Promise instances. The promise returned from the `Promise.all([ .. ])` call will receive a fulfillment message (`msgs` in this snippet) that is an `array` of all the fulfillment messages from the passed in promises, in the same order as specified (regardless of fulfillment order).

**Note:** Technically, the `array` of values passed into `Promise.all([ ... ])` can include Promises, thenables, or even immediate values. Each value in the list is essentially passed through `Promise.resolve(...)` to make sure it's a genuine Promise to be waited on, so an immediate value will just be normalized into a Promise for that value. If the `array` is empty, the main Promise is immediately fulfilled.

The main promise returned from `Promise.all([ ... ])` will only be fulfilled if and when all its constituent promises are fulfilled. If any one of those promises instead is rejected, the main `Promise.all([ ... ])` promise is immediately rejected, discarding all results from any other promises.

Remember to always attach a rejection/error handler to every promise, even and especially the one that comes back from `Promise.all([ ... ])`.

## Promise.race([ .. ])

While `Promise.all([ ... ])` coordinates multiple Promises concurrently and assumes all are needed for fulfillment, sometimes you only want to respond to the "first Promise to cross the finish line," letting the other Promises fall away.

This pattern is classically called a "latch," but in Promises it's called a "race."

**Warning:** While the metaphor of "only the first across the finish line wins" fits the behavior well, unfortunately "race" is kind of a loaded term, because "race conditions" are generally taken as bugs in programs (see Chapter 1). Don't confuse `Promise.race([ ... ])` with "race condition."

`Promise.race([ ... ])` also expects a single `array` argument, containing one or more Promises, thenables, or immediate values. It doesn't make much practical sense to have a race with immediate values, because the first one listed will obviously win - like a foot race where one runner starts at the finish line!

Similar to `Promise.all([ ... ])`, `Promise.race([ ... ])` will fulfill if and when any Promise resolution is a fulfillment, and it will reject if and when any Promise resolution is a rejection.

**Warning:** A "race" requires at least one "runner," so if you pass an empty `array`, instead of immediately resolving, the main `race([ ... ])` Promise will never resolve. This is a footgun! ES6 should have specified that it either fulfills, rejects, or just throws some sort of synchronous error. Unfortunately, because of precedence in Promise libraries predating ES6 `Promise`, they had to leave this gotcha in there, so be careful never to send in an empty `array`.

Let's revisit our previous concurrent Ajax example, but in the context of a race between `p1` and `p2`:

```
// `request(..)` is a Promise-aware Ajax utility,
// like we defined earlier in the chapter

var p1 = request( "http://some.url.1/" );
var p2 = request( "http://some.url.2/" );

Promise.race( [p1,p2] )
.then( function(msg){
    // either `p1` or `p2` will win the race
    return request(
        "http://some.url.3/?v=" + msg
    );
})
.then( function(msg){
    console.log( msg );
} );
```

Because only one promise wins, the fulfillment value is a single message, not an `array` as it was for `Promise.all([ ... ])`.

## Timeout Race

We saw this example earlier, illustrating how `Promise.race([ ... ])` can be used to express the "promise timeout" pattern:

```
// `foo()` is a Promise-aware function

// `timeoutPromise(..)`, defined earlier, returns
// a Promise that rejects after a specified delay

// setup a timeout for `foo()`
Promise.race([
  foo(), // attempt `foo()`
  timeoutPromise( 3000 ) // give it 3 seconds
])
.then(
  function(){
    // `foo(..)` fulfilled in time!
  },
  function(err){
    // either `foo()` rejected, or it just
    // didn't finish in time, so inspect
    // `err` to know which
  }
);

```

This timeout pattern works well in most cases. But there are some nuances to consider, and frankly they apply to both `Promise.race([ .. ])` and `Promise.all([ .. ])` equally.

## "Finally"

The key question to ask is, "What happens to the promises that get discarded/ignored?" We're not asking that question from the performance perspective -- they would typically end up garbage collection eligible -- but from the behavioral perspective (side effects, etc.). Promises cannot be canceled -- and shouldn't be as that would destroy the external immutability trust discussed in the "Promise Uncancelable" section later in this chapter -- so they can only be silently ignored.

But what if `foo()` in the previous example is reserving some sort of resource for usage, but the timeout fires first and causes that promise to be ignored? Is there anything in this pattern that proactively frees the reserved resource after the timeout, or otherwise cancels any side effects it may have had? What if all you wanted was to log the fact that `foo()` timed out?

Some developers have proposed that Promises need a `finally(..)` callback registration, which is always called when a Promise resolves, and allows you to specify any cleanup that may be necessary. This doesn't exist in the specification at the moment, but it may come in ES7+. We'll have to wait and see.

It might look like:

```
var p = Promise.resolve( 42 );

p.then( something )
.finally( cleanup )
.then( another )
.finally( cleanup );
```

**Note:** In various Promise libraries, `finally(..)` still creates and returns a new Promise (to keep the chain going). If the `cleanup(..)` function were to return a Promise, it would be linked into the chain, which means you could still have the unhandled rejection issues we discussed earlier.

In the meantime, we could make a static helper utility that lets us observe (without interfering) the resolution of a Promise:

```
// polyfill-safe guard check
if (!Promise.observe) {
  Promise.observe = function(pr,cb) {
    // side-observe `pr`'s resolution
    pr.then(
      function fulfilled(msg){
```

```

        // schedule callback async (as Job)
        Promise.resolve( msg ).then( cb );
    },
    function rejected(err){
        // schedule callback async (as Job)
        Promise.resolve( err ).then( cb );
    }
);

// return original promise
return pr;
};

}

```

Here's how we'd use it in the timeout example from before:

```

Promise.race([
    Promise.observe(
        foo(), // attempt `foo()`
        function cleanup(msg){
            // clean up after `foo()`, even if it
            // didn't finish before the timeout
        }
    ),
    timeoutPromise( 3000 ) // give it 3 seconds
])

```

This `Promise.observe(...)` helper is just an illustration of how you could observe the completions of Promises without interfering with them. Other Promise libraries have their own solutions. Regardless of how you do it, you'll likely have places where you want to make sure your Promises aren't *just* silently ignored by accident.

## Variations on `all([ ... ])` and `race([ ... ])`

While native ES6 Promises come with built-in `Promise.all([ ... ])` and `Promise.race([ ... ])`, there are several other commonly used patterns with variations on those semantics:

- `none([ ... ])` is like `all([ ... ])`, but fulfillments and rejections are transposed. All Promises need to be rejected -- rejections become the fulfillment values and vice versa.
- `any([ ... ])` is like `all([ ... ])`, but it ignores any rejections, so only one needs to fulfill instead of *all* of them.
- `first([ ... ])` is like a race with `any([ ... ])`, which is that it ignores any rejections and fulfills as soon as the first Promise fulfills.
- `last([ ... ])` is like `first([ ... ])`, but only the latest fulfillment wins.

Some Promise abstraction libraries provide these, but you could also define them yourself using the mechanics of Promises, `race([ ... ])` and `all([ ... ])`.

For example, here's how we could define `first([ ... ])`:

```

// polyfill-safe guard check
if (!Promise.first) {
    Promise.first = function(prs) {
        return new Promise( function(resolve,reject){
            // loop through all promises
            prs.forEach( function(pr){
                // normalize the value
                Promise.resolve( pr )
                // whichever one fulfills first wins, and
                // gets to resolve the main promise
                .then( resolve );
            } );
        } );
    }
}

```

```
    };
}
```

**Note:** This implementation of `first(..)` does not reject if all its promises reject; it simply hangs, much like a `Promise.race([])` does. If desired, you could add additional logic to track each promise rejection and if all reject, call `reject()` on the main promise. We'll leave that as an exercise for the reader.

## Concurrent Iterations

Sometimes you want to iterate over a list of Promises and perform some task against all of them, much like you can do with synchronous `array s` (e.g., `forEach(..)`, `map(..)`, `some(..)`, and `every(..)`). If the task to perform against each Promise is fundamentally synchronous, these work fine, just as we used `forEach(..)` in the previous snippet.

But if the tasks are fundamentally asynchronous, or can/should otherwise be performed concurrently, you can use async versions of these utilities as provided by many libraries.

For example, let's consider an asynchronous `map(..)` utility that takes an `array` of values (could be Promises or anything else), plus a function (task) to perform against each. `map(..)` itself returns a promise whose fulfillment value is an `array` that holds (in the same mapping order) the async fulfillment value from each task:

```
if (!Promise.map) {
  Promise.map = function(vals,cb) {
    // new promise that waits for all mapped promises
    return Promise.all(
      // note: regular array `map(..)`, turns
      // the array of values into an array of
      // promises
      vals.map( function(val){
        // replace `val` with a new promise that
        // resolves after `val` is async mapped
        return new Promise( function(resolve){
          cb( val, resolve );
        } );
      } );
    );
  };
}
```

**Note:** In this implementation of `map(..)`, you can't signal async rejection, but if a synchronous exception/error occurs inside of the mapping callback (`cb(..)`), the main `Promise.map(..)` returned promise would reject.

Let's illustrate using `map(..)` with a list of Promises (instead of simple values):

```
var p1 = Promise.resolve( 21 );
var p2 = Promise.resolve( 42 );
var p3 = Promise.reject( "Oops" );

// double values in list even if they're in
// Promises
Promise.map( [p1,p2,p3], function(pr,done){
  // make sure the item itself is a Promise
  Promise.resolve( pr )
  .then(
    // extract value as `v`
    function(v){
      // map fulfillment `v` to new value
      done( v * 2 );
    },
    // or, map to promise rejection message
    done
  );
})
```

```
.then( function(vals){
  console.log( vals ); // [42,84,"Oops"]
} );
```

## Promise API Recap

Let's review the ES6 `Promise` API that we've already seen unfold in bits and pieces throughout this chapter.

**Note:** The following API is native only as of ES6, but there are specification-compliant polyfills (not just extended Promise libraries) which can define `Promise` and all its associated behavior so that you can use native Promises even in pre-ES6 browsers. One such polyfill is "Native Promise Only" (<http://github.com/getify/native-promise-only>), which I wrote!

### `new Promise(..)` Constructor

The *revealing constructor* `Promise(..)` must be used with `new`, and must be provided a function callback that is synchronously/immediately called. This function is passed two function callbacks that act as resolution capabilities for the promise. We commonly label these `resolve(..)` and `reject(..)`:

```
var p = new Promise( function(resolve,reject){
  // `resolve(..)` to resolve/fulfill the promise
  // `reject(..)` to reject the promise
} );
```

`reject(..)` simply rejects the promise, but `resolve(..)` can either fulfill the promise or reject it, depending on what it's passed. If `resolve(..)` is passed an immediate, non-Promise, non-thenable value, then the promise is fulfilled with that value.

But if `resolve(..)` is passed a genuine Promise or thenable value, that value is unwrapped recursively, and whatever its final resolution/state is will be adopted by the promise.

### `Promise.resolve(..)` and `Promise.reject(..)`

A shortcut for creating an already-rejected Promise is `Promise.reject(..)`, so these two promises are equivalent:

```
var p1 = new Promise( function(resolve,reject){
  reject( "Oops" );
} );

var p2 = Promise.reject( "Oops" );
```

`Promise.resolve(..)` is usually used to create an already-fulfilled Promise in a similar way to `Promise.reject(..)`. However, `Promise.resolve(..)` also unwraps thenable values (as discussed several times already). In that case, the Promise returned adopts the final resolution of the thenable you passed in, which could either be fulfillment or rejection:

```
var fulfilledTh = {
  then: function(cb) { cb( 42 ); }
};

var rejectedTh = {
  then: function(cb,errCb) {
    errCb( "Oops" );
  }
};

var p1 = Promise.resolve( fulfilledTh );
var p2 = Promise.resolve( rejectedTh );

// `p1` will be a fulfilled promise
// `p2` will be a rejected promise
```

And remember, `Promise.resolve(..)` doesn't do anything if what you pass is already a genuine Promise; it just returns the value directly. So there's no overhead to calling `Promise.resolve(..)` on values that you don't know the nature of, if one happens to already be a genuine Promise.

## then(..) and catch(..)

Each Promise instance (**not** the `Promise` API namespace) has `then(..)` and `catch(..)` methods, which allow registering of fulfillment and rejection handlers for the Promise. Once the Promise is resolved, one or the other of these handlers will be called, but not both, and it will always be called asynchronously (see "Jobs" in Chapter 1).

`then(..)` takes one or two parameters, the first for the fulfillment callback, and the second for the rejection callback. If either is omitted or is otherwise passed as a non-function value, a default callback is substituted respectively. The default fulfillment callback simply passes the message along, while the default rejection callback simply rethrows (propagates) the error reason it receives.

`catch(..)` takes only the rejection callback as a parameter, and automatically substitutes the default fulfillment callback, as just discussed. In other words, it's equivalent to `then(null,..)`:

```
p.then( fulfilled );
p.then( fulfilled, rejected );
p.catch( rejected ); // or `p.then( null, rejected )`
```

`then(..)` and `catch(..)` also create and return a new promise, which can be used to express Promise chain flow control. If the fulfillment or rejection callbacks have an exception thrown, the returned promise is rejected. If either callback returns an immediate, non-Promise, non-thenable value, that value is set as the fulfillment for the returned promise. If the fulfillment handler specifically returns a promise or thenable value, that value is unwrapped and becomes the resolution of the returned promise.

## Promise.all([ .. ]) and Promise.race([ .. ])

The static helpers `Promise.all([ .. ])` and `Promise.race([ .. ])` on the ES6 `Promise` API both create a Promise as their return value. The resolution of that promise is controlled entirely by the array of promises that you pass in.

For `Promise.all([ .. ])`, all the promises you pass in must fulfill for the returned promise to fulfill. If any promise is rejected, the main returned promise is immediately rejected, too (discarding the results of any of the other promises). For fulfillment, you receive an `array` of all the passed in promises' fulfillment values. For rejection, you receive just the first promise rejection reason value. This pattern is classically called a "gate": all must arrive before the gate opens.

For `Promise.race([ .. ])`, only the first promise to resolve (fulfillment or rejection) "wins," and whatever that resolution is becomes the resolution of the returned promise. This pattern is classically called a "latch": first one to open the latch gets through. Consider:

```
var p1 = Promise.resolve( 42 );
var p2 = Promise.resolve( "Hello World" );
var p3 = Promise.reject( "Oops" );

Promise.race( [p1,p2,p3] )
.then( function(msg){
    console.log( msg );           // 42
} );

Promise.all( [p1,p2,p3] )
.catch( function(err){
    console.error( err );        // "Oops"
} );
```

```
Promise.all( [p1,p2] )
  .then( function(msgs){
    console.log( msgs ); // [42,"Hello World"]
  });

```

**Warning:** Be careful! If an empty `array` is passed to `Promise.all([ ... ])`, it will fulfill immediately, but `Promise.race([ ... ])` will hang forever and never resolve.

The ES6 `Promise` API is pretty simple and straightforward. It's at least good enough to serve the most basic of async cases, and is a good place to start when rearranging your code from callback hell to something better.

But there's a whole lot of async sophistication that apps often demand which Promises themselves will be limited in addressing. In the next section, we'll dive into those limitations as motivations for the benefit of Promise libraries.

## Promise Limitations

Many of the details we'll discuss in this section have already been alluded to in this chapter, but we'll just make sure to review these limitations specifically.

### Sequence Error Handling

We covered Promise-flavored error handling in detail earlier in this chapter. The limitations of how Promises are designed -- how they chain, specifically -- creates a very easy pitfall where an error in a Promise chain can be silently ignored accidentally.

But there's something else to consider with Promise errors. Because a Promise chain is nothing more than its constituent Promises wired together, there's no entity to refer to the entire chain as a single *thing*, which means there's no external way to observe any errors that may occur.

If you construct a Promise chain that has no error handling in it, any error anywhere in the chain will propagate indefinitely down the chain, until observed (by registering a rejection handler at some step). So, in that specific case, having a reference to the *last* promise in the chain is enough (`p` in the following snippet), because you can register a rejection handler there, and it will be notified of any propagated errors:

```
// `foo(..)`, `STEP2(..)` and `STEP3(..)` are
// all promise-aware utilities

var p = foo( 42 )
  .then( STEP2 )
  .then( STEP3 );

```

Although it may seem sneakily confusing, `p` here doesn't point to the first promise in the chain (the one from the `foo(42)` call), but instead from the last promise, the one that comes from the `then(STEP3)` call.

Also, no step in the promise chain is observably doing its own error handling. That means that you could then register a rejection error handler on `p`, and it would be notified if any errors occur anywhere in the chain:

```
p.catch( handleErrors );

```

But if any step of the chain in fact does its own error handling (perhaps hidden/abstracted away from what you can see), your `handleErrors(..)` won't be notified. This may be what you want -- it was, after all, a "handled rejection" -- but it also may *not* be what you want. The complete lack of ability to be notified (of "already handled" rejection errors) is a limitation that restricts capabilities in some use cases.

It's basically the same limitation that exists with a `try..catch` that can catch an exception and simply swallow it. So this isn't a limitation **unique to Promises**, but it is something we might wish to have a workaround for.

Unfortunately, many times there is no reference kept for the intermediate steps in a Promise-chain sequence, so without such references, you cannot attach error handlers to reliably observe the errors.

## Single Value

Promises by definition only have a single fulfillment value or a single rejection reason. In simple examples, this isn't that big of a deal, but in more sophisticated scenarios, you may find this limiting.

The typical advice is to construct a values wrapper (such as an `object` or `array`) to contain these multiple messages. This solution works, but it can be quite awkward and tedious to wrap and unwrap your messages with every single step of your Promise chain.

## Splitting Values

Sometimes you can take this as a signal that you could/should decompose the problem into two or more Promises.

Imagine you have a utility `foo(..)` that produces two values (`x` and `y`) asynchronously:

```
function getY(x) {
  return new Promise( function(resolve,reject){
    setTimeout( function(){
      resolve( (3 * x) - 1 );
    }, 100 );
  } );
}

function foo(bar,baz) {
  var x = bar * baz;

  return getY( x )
    .then( function(y){
      // wrap both values into container
      return [x,y];
    } );
}

foo( 10, 20 )
.then( function(msgs){
  var x = msgs[0];
  var y = msgs[1];

  console.log( x, y );    // 200 599
} );
```

First, let's rearrange what `foo(..)` returns so that we don't have to wrap `x` and `y` into a single `array` value to transport through one Promise. Instead, we can wrap each value into its own promise:

```
function foo(bar,baz) {
  var x = bar * baz;

  // return both promises
  return [
    Promise.resolve( x ),
    getY( x )
  ];
}

Promise.all(
  foo( 10, 20 )
)
.then( function(msgs){
  var x = msgs[0];
```

```

var y = msgs[1];
console.log( x, y );
} );

```

Is an `array` of promises really better than an `array` of values passed through a single promise? Syntactically, it's not much of an improvement.

But this approach more closely embraces the Promise design theory. It's now easier in the future to refactor to split the calculation of `x` and `y` into separate functions. It's cleaner and more flexible to let the calling code decide how to orchestrate the two promises -- using `Promise.all([ ... ])` here, but certainly not the only option -- rather than to abstract such details away inside of `foo(..)`.

## Unwrap/Spread Arguments

The `var x = ..` and `var y = ..` assignments are still awkward overhead. We can employ some functional trickery (hat tip to Reginald Braithwaite, @raganwald on Twitter) in a helper utility:

```

function spread(fn) {
  return Function.prototype.bind( fn, null );
}

Promise.all(
  foo( 10, 20 )
)
.then(
  spread( function(x,y){
    console.log( x, y );      // 200 599
  } )
)

```

That's a bit nicer! Of course, you could inline the functional magic to avoid the extra helper:

```

Promise.all(
  foo( 10, 20 )
)
.then( Function.prototype.bind(
  function(x,y){
    console.log( x, y );      // 200 599
  },
  null
) );

```

These tricks may be neat, but ES6 has an even better answer for us: destructuring. The array destructuring assignment form looks like this:

```

Promise.all(
  foo( 10, 20 )
)
.then( function(msgs){
  var [x,y] = msgs;

  console.log( x, y );      // 200 599
} );

```

But best of all, ES6 offers the array parameter destructuring form:

```

Promise.all(
  foo( 10, 20 )
)

```

```

)
.then( function([x,y]){
    console.log( x, y );    // 200 599
} );

```

We've now embraced the one-value-per-Promise mantra, but kept our supporting boilerplate to a minimum!

**Note:** For more information on ES6 destructuring forms, see the *ES6 & Beyond* title of this series.

## Single Resolution

One of the most intrinsic behaviors of Promises is that a Promise can only be resolved once (fulfillment or rejection). For many async use cases, you're only retrieving a value once, so this works fine.

But there's also a lot of async cases that fit into a different model -- one that's more akin to events and/or streams of data. It's not clear on the surface how well Promises can fit into such use cases, if at all. Without a significant abstraction on top of Promises, they will completely fall short for handling multiple value resolution.

Imagine a scenario where you might want to fire off a sequence of async steps in response to a stimulus (like an event) that can in fact happen multiple times, like a button click.

This probably won't work the way you want:

```

// `click(..)` binds the `click` event to a DOM element
// `request(..)` is the previously defined Promise-aware Ajax

var p = new Promise( function(resolve, reject){
    click( "#mybtn", resolve );
} );

p.then( function(evt){
    var btnID = evt.currentTarget.id;
    return request( "http://some.url.1/?id=" + btnID );
} )
.then( function(text){
    console.log( text );
} );

```

The behavior here only works if your application calls for the button to be clicked just once. If the button is clicked a second time, the `p` promise has already been resolved, so the second `resolve(..)` call would be ignored.

Instead, you'd probably need to invert the paradigm, creating a whole new Promise chain for each event firing:

```

click( "#mybtn", function(evt){
    var btnID = evt.currentTarget.id;

    request( "http://some.url.1/?id=" + btnID )
    .then( function(text){
        console.log( text );
    } );
} );

```

This approach will work in that a whole new Promise sequence will be fired off for each `"click"` event on the button.

But beyond just the ugliness of having to define the entire Promise chain inside the event handler, this design in some respects violates the idea of separation of concerns/capabilities (SoC). You might very well want to define your event handler in a different place in your code from where you define the *response* to the event (the Promise chain). That's pretty awkward to do in this pattern, without helper mechanisms.

**Note:** Another way of articulating this limitation is that it'd be nice if we could construct some sort of "observable" that we can subscribe a Promise chain to. There are libraries that have created these abstractions (such as RxJS -- <http://rxjs.codeplex.com/>), but the abstractions can seem so heavy that you can't even see the nature of Promises anymore. Such heavy abstraction brings important questions to mind such as whether (sans Promises) these mechanisms are as *trustable* as Promises themselves have been designed to be. We'll revisit the "Observable" pattern in Appendix B.

## Inertia

One concrete barrier to starting to use Promises in your own code is all the code that currently exists which is not already Promise-aware. If you have lots of callback-based code, it's far easier to just keep coding in that same style.

"A code base in motion (with callbacks) will remain in motion (with callbacks) unless acted upon by a smart, Promises-aware developer."

Promises offer a different paradigm, and as such, the approach to the code can be anywhere from just a little different to, in some cases, radically different. You have to be intentional about it, because Promises will not just naturally shake out from the same ol' ways of doing code that have served you well thus far.

Consider a callback-based scenario like the following:

```
function foo(x,y,cb) {
  ajax(
    "http://some.url.1/?x=" + x + "&y=" + y,
    cb
  );
}

foo( 11, 31, function(err,text) {
  if (err) {
    console.error( err );
  }
  else {
    console.log( text );
  }
});
```

Is it immediately obvious what the first steps are to convert this callback-based code to Promise-aware code? Depends on your experience. The more practice you have with it, the more natural it will feel. But certainly, Promises don't just advertise on the label exactly how to do it -- there's no one-size-fits-all answer -- so the responsibility is up to you.

As we've covered before, we definitely need an Ajax utility that is Promise-aware instead of callback-based, which we could call `request(..)`. You can make your own, as we have already. But the overhead of having to manually define Promise-aware wrappers for every callback-based utility makes it less likely you'll choose to refactor to Promise-aware coding at all.

Promises offer no direct answer to that limitation. Most Promise libraries do offer a helper, however. But even without a library, imagine a helper like this:

```
// polyfill-safe guard check
if (!Promise.wrap) {
  Promise.wrap = function(fn) {
    return function() {
      var args = [].slice.call( arguments );

      return new Promise( function(resolve,reject){
        fn.apply(
          null,
          args.concat( function(err,v){
            if (err) {
              reject( err );
            }
          })
        );
      });
    };
  };
}
```

```
        else {  
            resolve( v );  
        }  
    } )  
);  
} );  
};  
};  
};  
}
```

OK, that's more than just a tiny trivial utility. However, although it may look a bit intimidating, it's not as bad as you'd think. It takes a function that expects an error-first style callback as its last parameter, and returns a new one that automatically creates a Promise to return, and substitutes the callback for you, wired up to the Promise fulfillment/rejection.

Rather than waste too much time talking about *how* this `Promise.wrap(...)` helper works, let's just look at how we use it:

```
var request = Promise.wrap( ajax )  
  
request( "http://some.url.1/" )  
.then( ... )  
..
```

Wow, that was pretty easy!

`Promise.wrap(..)` does **not** produce a Promise. It produces a function that will produce Promises. In a sense, a Promise-producing function could be seen as a "Promise factory." I propose "promisory" as the name for such a thing ("Promise" + "factory").

The act of wrapping a callback-expecting function to be a Promise-aware function is sometimes referred to as "lifting" or "promisifying". But there doesn't seem to be a standard term for what to call the resultant function other than a "lifted function", so I like "promisory" better as I think it's more descriptive.

**Note:** Promisory isn't a made-up term. It's a real word, and its definition means to contain or convey a promise. That's exactly what these functions are doing, so it turns out to be a pretty perfect terminology match!

So, `Promise.wrap.ajax` produces an `ajax(..)` promisory we call `request(..)`, and that promisory produces Promises for Ajax responses.

If all functions were already promises, we wouldn't need to make them ourselves, so the extra step is a tad bit of a shame. But at least the wrapping pattern is (usually) repeatable so we can put it into a `Promise.wrap(..)` helper as shown to aid our promise coding.

So back to our earlier example, we need a promisory for both `ajax(..)` and `foo(..)`:

```
// make a promisory for `ajax(..)`
var request = Promise.wrap( ajax );

// refactor `foo(..)`, but keep it externally
// callback-based for compatibility with other
// parts of the code for now -- only use
// `request(..)`'s promise internally.

function foo(x,y,cb) {
    request(
        "http://some.url.1/?x=" + x + "&y=" + y
    )
    .then(
        function fulfilled(text){
            cb( null, text );
        },
        cb
    );
}
```

```
// now, for this code's purposes, make a
// promisory for `foo(..)`
var betterFoo = Promise.wrap( foo );

// and use the promisory
betterFoo( 11, 31 )
.then(
  function fulfilled(text){
    console.log( text );
  },
  function rejected(err){
    console.error( err );
  }
);

```

Of course, while we're refactoring `foo(..)` to use our new `request(..)` promisory, we could just make `foo(..)` a promisory itself, instead of remaining callback-based and needing to make and use the subsequent `betterFoo(..)` promisory. This decision just depends on whether `foo(..)` needs to stay callback-based compatible with other parts of the code base or not.

Consider:

```
// `foo(..)` is now also a promisory because it
// delegates to the `request(..)` promisory
function foo(x,y) {
  return request(
    "http://some.url.1/?x=" + x + "&y=" + y
  );
}

foo( 11, 31 )
.then( ... )
...
```

While ES6 Promises don't natively ship with helpers for such promisory wrapping, most libraries provide them, or you can make your own. Either way, this particular limitation of Promises is addressable without too much pain (certainly compared to the pain of callback hell!).

## Promise Uncancelable

Once you create a Promise and register a fulfillment and/or rejection handler for it, there's nothing external you can do to stop that progression if something else happens to make that task moot.

**Note:** Many Promise abstraction libraries provide facilities to cancel Promises, but this is a terrible idea! Many developers wish Promises had natively been designed with external cancelation capability, but the problem is that it would let one consumer/observer of a Promise affect some other consumer's ability to observe that same Promise. This violates the future-value's trustability (external immutability), but moreover is the embodiment of the "action at a distance" anti-pattern ([http://en.wikipedia.org/wiki/Action\\_at\\_a\\_distance\\_%28computer\\_programming%29](http://en.wikipedia.org/wiki/Action_at_a_distance_%28computer_programming%29)). Regardless of how useful it seems, it will actually lead you straight back into the same nightmares as callbacks.

Consider our Promise timeout scenario from earlier:

```
var p = foo( 42 );

Promise.race( [
  p,
  timeoutPromise( 3000 )
] )
.then(
  doSomething,
  handleError
)
```

```

);
p.then( function(){
  // still happens even in the timeout case :(
} );

```

The "timeout" was external to the promise `p`, so `p` itself keeps going, which we probably don't want.

One option is to invasively define your resolution callbacks:

```

var OK = true;

var p = foo( 42 );

Promise.race([
  p,
  timeoutPromise( 3000 )
  .catch( function(err){
    OK = false;
    throw err;
  } )
])
.then(
  doSomething,
  handleError
);

p.then( function(){
  if (OK) {
    // only happens if no timeout! :)
  }
} );

```

This is ugly. It works, but it's far from ideal. Generally, you should try to avoid such scenarios.

But if you can't, the ugliness of this solution should be a clue that *cancellation* is a functionality that belongs at a higher level of abstraction on top of Promises. I'd recommend you look to Promise abstraction libraries for assistance rather than hacking it yourself.

**Note:** My *asynquence* Promise abstraction library provides just such an abstraction and an `abort()` capability for the sequence, all of which will be discussed in Appendix A.

A single Promise is not really a flow-control mechanism (at least not in a very meaningful sense), which is exactly what *cancellation* refers to; that's why Promise cancelation would feel awkward.

By contrast, a chain of Promises taken collectively together -- what I like to call a "sequence" -- *is* a flow control expression, and thus it's appropriate for cancellation to be defined at that level of abstraction.

No individual Promise should be cancelable, but it's sensible for a *sequence* to be cancelable, because you don't pass around a sequence as a single immutable value like you do with a Promise.

## Promise Performance

This particular limitation is both simple and complex.

Comparing how many pieces are moving with a basic callback-based async task chain versus a Promise chain, it's clear Promises have a fair bit more going on, which means they are naturally at least a tiny bit slower. Think back to just the simple list of trust guarantees that Promises offer, as compared to the ad hoc solution code you'd have to layer on top of callbacks to achieve the same protections.

More work to do, more guards to protect, means that Promises *are* slower as compared to naked, untrustable callbacks. That much is obvious, and probably simple to wrap your brain around.

But how much slower? Well... that's actually proving to be an incredibly difficult question to answer absolutely, across the board.

Frankly, it's kind of an apples-to-oranges comparison, so it's probably the wrong question to ask. You should actually compare whether an ad-hoc callback system with all the same protections manually layered in is faster than a Promise implementation.

If Promises have a legitimate performance limitation, it's more that they don't really offer a line-item choice as to which trustability protections you want/need or not -- you get them all, always.

Nevertheless, if we grant that a Promise is generally a *little bit slower* than its non-Promise, non-trustable callback equivalent -- assuming there are places where you feel you can justify the lack of trustability -- does that mean that Promises should be avoided across the board, as if your entire application is driven by nothing but must-be-utterly-the-fastest code possible?

Sanity check: if your code is legitimately like that, **is JavaScript even the right language for such tasks?** JavaScript can be optimized to run applications very performantly (see Chapter 5 and Chapter 6). But is obsessing over tiny performance tradeoffs with Promises, in light of all the benefits they offer, *really* appropriate?

Another subtle issue is that Promises make *everything* async, which means that some immediately (synchronously) complete steps still defer advancement of the next step to a Job (see Chapter 1). That means that it's possible that a sequence of Promise tasks could complete ever-so-slightly slower than the same sequence wired up with callbacks.

Of course, the question here is this: are these potential slips in tiny fractions of performance *worth* all the other articulated benefits of Promises we've laid out across this chapter?

My take is that in virtually all cases where you might think Promise performance is slow enough to be concerned, it's actually an anti-pattern to optimize away the benefits of Promise trustability and composability by avoiding them altogether.

Instead, you should default to using them across the code base, and then profile and analyze your application's hot (critical) paths. Are Promises *really* a bottleneck, or are they just a theoretical slowdown? Only *then*, armed with actual valid benchmarks (see Chapter 6) is it responsible and prudent to factor out the Promises in just those identified critical areas.

Promises are a little slower, but in exchange you're getting a lot of trustability, non-Zalgo predictability, and composability built in. Maybe the limitation is not actually their performance, but your lack of perception of their benefits?

## Review

Promises are awesome. Use them. They solve the *inversion of control* issues that plague us with callbacks-only code.

They don't get rid of callbacks, they just redirect the orchestration of those callbacks to a trustable intermediary mechanism that sits between us and another utility.

Promise chains also begin to address (though certainly not perfectly) a better way of expressing async flow in sequential fashion, which helps our brains plan and maintain async JS code better. We'll see an even better solution to *that* problem in the next chapter!

# You Don't Know JS Yet: Sync & Async - 2nd Edition

## Chapter 4: Generators

### NOTE:

Work in progress

### NOTE:

Everything below here is previous text from 1st edition, and is only here for reference while 2nd edition work is underway.  
**Please ignore this stuff.**

In Chapter 2, we identified two key drawbacks to expressing async flow control with callbacks:

- Callback-based async doesn't fit how our brain plans out steps of a task.
- Callbacks aren't trustable or composable because of *inversion of control*.

In Chapter 3, we detailed how Promises uninvert the *inversion of control* of callbacks, restoring trustability/composability.

Now we turn our attention to expressing async flow control in a sequential, synchronous-looking fashion. The "magic" that makes it possible is ES6 **generators**.

## Breaking Run-to-Completion

In Chapter 1, we explained an expectation that JS developers almost universally rely on in their code: once a function starts executing, it runs until it completes, and no other code can interrupt and run in between.

As bizarre as it may seem, ES6 introduces a new type of function that does not behave with the run-to-completion behavior. This new type of function is called a "generator."

To understand the implications, let's consider this example:

```
var x = 1;

function foo() {
  x++;
  bar();           // <-- what about this line?
  console.log("x:", x);
}
```

```

function bar() {
    x++;
}

foo();           // x: 3

```

In this example, we know for sure that `bar()` runs in between `x++` and `console.log(x)`. But what if `bar()` wasn't there? Obviously, the result would be `2` instead of `3`.

Now let's twist your brain. What if `bar()` wasn't present, but it could still somehow run between the `x++` and `console.log(x)` statements? How would that be possible?

In **preemptive** multithreaded languages, it would essentially be possible for `bar()` to "interrupt" and run at exactly the right moment between those two statements. But JS is not preemptive, nor is it (currently) multithreaded. And yet, a **cooperative** form of this "interruption" (concurrency) is possible, if `foo()` itself could somehow indicate a "pause" at that part in the code.

**Note:** I use the word "cooperative" not only because of the connection to classical concurrency terminology (see Chapter 1), but because as you'll see in the next snippet, the ES6 syntax for indicating a pause point in code is `yield` -- suggesting a politely *cooperative* yielding of control.

Here's the ES6 code to accomplish such cooperative concurrency:

```

var x = 1;

function *foo() {
    x++;
    yield; // pause!
    console.log("x:", x);
}

function bar() {
    x++;
}

```

**Note:** You will likely see most other JS documentation/code that will format a generator declaration as `function* foo() { ... }` instead of as I've done here with `function *foo() { ... }` -- the only difference being the stylistic positioning of the `*`. The two forms are functionally/syntactically identical, as is a third `function*foo() { ... }` (no space) form. There are arguments for both styles, but I basically prefer `function *foo..` because it then matches when I reference a generator in writing with `*foo()`. If I said only `foo()`, you wouldn't know as clearly if I was talking about a generator or a regular function. It's purely a stylistic preference.

Now, how can we run the code in that previous snippet such that `bar()` executes at the point of the `yield` inside of `*foo()`?

```

// construct an iterator `it` to control the generator
var it = foo();

// start `foo()` here!
it.next();
x;           // 2
bar();
x;           // 3
it.next();   // x: 3

```

OK, there's quite a bit of new and potentially confusing stuff in those two code snippets, so we've got plenty to wade through. But before we explain the different mechanics/syntax with ES6 generators, let's walk through the behavior flow:

1. The `it = foo()` operation does *not* execute the `*foo()` generator yet, but it merely constructs an *iterator* that will control its execution. More on *iterators* in a bit.
2. The first `it.next()` starts the `*foo()` generator, and runs the `x++` on the first line of `*foo()`.

3. `*foo()` pauses at the `yield` statement, at which point that first `it.next()` call finishes. At the moment, `*foo()` is still running and active, but it's in a paused state.
4. We inspect the value of `x`, and it's now `2`.
5. We call `bar()`, which increments `x` again with `x++`.
6. We inspect the value of `x` again, and it's now `3`.
7. The final `it.next()` call resumes the `*foo()` generator from where it was paused, and runs the `console.log(..)` statement, which uses the current value of `x` of `3`.

Clearly, `*foo()` started, but did *not* run-to-completion -- it paused at the `yield`. We resumed `*foo()` later, and let it finish, but that wasn't even required.

So, a generator is a special kind of function that can start and stop one or more times, and doesn't necessarily ever have to finish. While it won't be terribly obvious yet why that's so powerful, as we go throughout the rest of this chapter, that will be one of the fundamental building blocks we use to construct generators-as-async-flow-control as a pattern for our code.

## Input and Output

A generator function is a special function with the new processing model we just alluded to. But it's still a function, which means it still has some basic tenets that haven't changed -- namely, that it still accepts arguments (aka "input"), and that it can still return a value (aka "output"):

```
function *foo(x,y) {
  return x * y;
}

var it = foo( 6, 7 );

var res = it.next();

res.value;      // 42
```

We pass in the arguments `6` and `7` to `*foo(..)` as the parameters `x` and `y`, respectively. And `*foo(..)` returns the value `42` back to the calling code.

We now see a difference with how the generator is invoked compared to a normal function. `foo(6,7)` obviously looks familiar. But subtly, the `*foo(..)` generator hasn't actually run yet as it would have with a function.

Instead, we're just creating an *iterator* object, which we assign to the variable `it`, to control the `*foo(..)` generator. Then we call `it.next()`, which instructs the `*foo(..)` generator to advance from its current location, stopping either at the next `yield` or end of the generator.

The result of that `next(..)` call is an object with a `value` property on it holding whatever value (if anything) was returned from `*foo(..)`. In other words, `yield` caused a value to be sent out from the generator during the middle of its execution, kind of like an intermediate `return`.

Again, it won't be obvious yet why we need this whole indirect *iterator* object to control the generator. We'll get there, I promise.

## Iteration Messaging

In addition to generators accepting arguments and having return values, there's even more powerful and compelling input/output messaging capability built into them, via `yield` and `next(..)`.

Consider:

```
function *foo(x) {
  var y = x * (yield);
  return y;
```

```

}

var it = foo( 6 );

// start `foo(..)`
it.next();

var res = it.next( 7 );

res.value;      // 42

```

First, we pass in `6` as the parameter `x`. Then we call `it.next()`, and it starts up `*foo(..)`.

Inside `*foo(..)`, the `var y = x ..` statement starts to be processed, but then it runs across a `yield` expression. At that point, it pauses `*foo(..)` (in the middle of the assignment statement!), and essentially requests the calling code to provide a result value for the `yield` expression. Next, we call `it.next( 7 )`, which is passing the `7` value back in to be that result of the paused `yield` expression.

So, at this point, the assignment statement is essentially `var y = 6 * 7`. Now, `return y` returns that `42` value back as the result of the `it.next( 7 )` call.

Notice something very important but also easily confusing, even to seasoned JS developers: depending on your perspective, there's a mismatch between the `yield` and the `next(..)` call. In general, you're going to have one more `next(..)` call than you have `yield` statements -- the preceding snippet has one `yield` and two `next(..)` calls.

Why the mismatch?

Because the first `next(..)` always starts a generator, and runs to the first `yield`. But it's the second `next(..)` call that fulfills the first paused `yield` expression, and the third `next(..)` would fulfill the second `yield`, and so on.

### Tale of Two Questions

Actually, which code you're thinking about primarily will affect whether there's a perceived mismatch or not.

Consider only the generator code:

```

var y = x * (yield);
return y;

```

This **first** `yield` is basically *asking a question*: "What value should I insert here?"

Who's going to answer that question? Well, the **first** `next()` has already run to get the generator up to this point, so obviously *it* can't answer the question. So, the **second** `next(..)` call must answer the question *posed by the first yield*.

See the mismatch -- second-to-first?

But let's flip our perspective. Let's look at it not from the generator's point of view, but from the iterator's point of view.

To properly illustrate this perspective, we also need to explain that messages can go in both directions -- `yield ..` as an expression can send out messages in response to `next(..)` calls, and `next(..)` can send values to a paused `yield` expression. Consider this slightly adjusted code:

```

function *foo(x) {
    var y = x * (yield "Hello");    // <- yield a value!
    return y;
}

var it = foo( 6 );

var res = it.next();    // first `next()`, don't pass anything
res.value;              // "Hello"

```

```
res = it.next( 7 );           // pass `7` to waiting `yield`
res.value;                  // 42
```

`yield ..` and `next(..)` pair together as a two-way message passing system **during the execution of the generator**.

So, looking only at the *iterator* code:

```
var res = it.next();          // first `next()`, don't pass anything
res.value;                  // "Hello"

res = it.next( 7 );          // pass `7` to waiting `yield`
res.value;                  // 42
```

**Note:** We don't pass a value to the first `next()` call, and that's on purpose. Only a paused `yield` could accept such a value passed by a `next(..)`, and at the beginning of the generator when we call the first `next()`, there is no paused `yield` to accept such a value. The specification and all compliant browsers just silently **discard** anything passed to the first `next()`. It's still a bad idea to pass a value, as you're just creating silently "failing" code that's confusing. So, always start a generator with an argument-free `next()`.

The first `next()` call (with nothing passed to it) is basically *asking a question*: "What `next` value does the `*foo(..)` generator have to give me?" And who answers this question? The first `yield "hello"` expression.

See? No mismatch there.

Depending on *who* you think about asking the question, there is either a mismatch between the `yield` and `next(..)` calls, or not.

But wait! There's still an extra `next()` compared to the number of `yield` statements. So, that final `it.next(7)` call is again asking the question about what `next` value the generator will produce. But there's no more `yield` statements left to answer, is there? So who answers?

The `return` statement answers the question!

And if there is no `return` in your generator -- `return` is certainly not any more required in generators than in regular functions -- there's always an assumed/implicit `return; (aka return undefined; )`, which serves the purpose of default answering the question *posed* by the final `it.next(7)` call.

These questions and answers -- the two-way message passing with `yield` and `next(..)` -- are quite powerful, but it's not obvious at all how these mechanisms are connected to async flow control. We're getting there!

## Multiple Iterators

It may appear from the syntactic usage that when you use an *iterator* to control a generator, you're controlling the declared generator function itself. But there's a subtlety that's easy to miss: each time you construct an *iterator*, you are implicitly constructing an instance of the generator which that *iterator* will control.

You can have multiple instances of the same generator running at the same time, and they can even interact:

```
function *foo() {
  var x = yield 2;
  z++;
  var y = yield (x * z);
  console.log( x, y, z );
}

var z = 1;

var it1 = foo();
var it2 = foo();
```

```

var val1 = it1.next().value;           // 2 <- yield 2
var val2 = it2.next().value;           // 2 <- yield 2

val1 = it1.next( val2 * 10 ).value;    // 40 <- x:20, z:2
val2 = it2.next( val1 * 5 ).value;    // 600 <- x:200, z:3

it1.next( val2 / 2 );                // y:300
                                         // 20 300 3
it2.next( val1 / 4 );                // y:10
                                         // 200 10 3

```

**Warning:** The most common usage of multiple instances of the same generator running concurrently is not such interactions, but when the generator is producing its own values without input, perhaps from some independently connected resource. We'll talk more about value production in the next section.

Let's briefly walk through the processing:

1. Both instances of `*foo()` are started at the same time, and both `next()` calls reveal a `value` of `2` from the `yield 2` statements, respectively.
2. `val2 * 10` is `2 * 10`, which is sent into the first generator instance `it1`, so that `x` gets value `20`. `z` is incremented from `1` to `2`, and then `20 * 2` is yielded out, setting `val1` to `40`.
3. `val1 * 5` is `40 * 5`, which is sent into the second generator instance `it2`, so that `x` gets value `200`. `z` is incremented again, from `2` to `3`, and then `200 * 3` is yielded out, setting `val2` to `600`.
4. `val2 / 2` is `600 / 2`, which is sent into the first generator instance `it1`, so that `y` gets value `300`, then printing out `20 300 3` for its `x y z` values, respectively.
5. `val1 / 4` is `40 / 4`, which is sent into the second generator instance `it2`, so that `y` gets value `10`, then printing out `200 10 3` for its `x y z` values, respectively.

That's a "fun" example to run through in your mind. Did you keep it straight?

## Interleaving

Recall this scenario from the "Run-to-completion" section of Chapter 1:

```

var a = 1;
var b = 2;

function foo() {
  a++;
  b = b * a;
  a = b + 3;
}

function bar() {
  b--;
  a = 8 + b;
  b = a * 2;
}

```

With normal JS functions, of course either `foo()` can run completely first, or `bar()` can run completely first, but `foo()` cannot interleave its individual statements with `bar()`. So, there are only two possible outcomes to the preceding program.

However, with generators, clearly interleaving (even in the middle of statements!) is possible:

```

var a = 1;
var b = 2;

function *foo() {
  a++;

```

```

    yield;
    b = b * a;
    a = (yield b) + 3;
}

function *bar() {
  b--;
  yield;
  a = (yield 8) + b;
  b = a * (yield 2);
}

```

Depending on what respective order the *iterators* controlling `*foo()` and `*bar()` are called, the preceding program could produce several different results. In other words, we can actually illustrate (in a sort of fake-ish way) the theoretical "threaded race conditions" circumstances discussed in Chapter 1, by interleaving the two generator iterations over the same shared variables.

First, let's make a helper called `step(..)` that controls an *iterator*:

```

function step(gen) {
  var it = gen();
  var last;

  return function() {
    // whatever is `yield`ed out, just
    // send it right back in the next time!
    last = it.next( last ).value;
  };
}

```

`step(..)` initializes a generator to create its `it iterator`, then returns a function which, when called, advances the *iterator* by one step. Additionally, the previously `yield` ed out value is sent right back in at the `next` step. So, `yield 8` will just become `8` and `yield b` will just be `b` (whatever it was at the time of `yield`).

Now, just for fun, let's experiment to see the effects of interleaving these different chunks of `*foo()` and `*bar()`. We'll start with the boring base case, making sure `*foo()` totally finishes before `*bar()` (just like we did in Chapter 1):

```

// make sure to reset `a` and `b`
a = 1;
b = 2;

var s1 = step( foo );
var s2 = step( bar );

// run `*foo()` completely first
s1();
s1();
s1();

// now run `*bar()`
s2();
s2();
s2();
s2();

console.log( a, b );    // 11 22

```

The end result is `11` and `22`, just as it was in the Chapter 1 version. Now let's mix up the interleaving ordering and see how it changes the final values of `a` and `b`:

```

// make sure to reset `a` and `b`
a = 1;
b = 2;

```

```

var s1 = step( foo );
var s2 = step( bar );

s2();           // b--;
s2();           // yield 8
s1();           // a++;
s2();           // a = 8 + b;
               // yield 2
s1();           // b = b * a;
               // yield b
s1();           // a = b + 3;
s2();           // b = a * 2;

```

Before I tell you the results, can you figure out what `a` and `b` are after the preceding program? No cheating!

```
console.log( a, b );    // 12 18
```

**Note:** As an exercise for the reader, try to see how many other combinations of results you can get back rearranging the order of the `s1()` and `s2()` calls. Don't forget you'll always need three `s1()` calls and four `s2()` calls. Recall the discussion earlier about matching `next()` with `yield` for the reasons why.

You almost certainly won't want to intentionally create *this* level of interleaving confusion, as it creates incredibly difficult to understand code. But the exercise is interesting and instructive to understand more about how multiple generators can run concurrently in the same shared scope, because there will be places where this capability is quite useful.

We'll discuss generator concurrency in more detail at the end of this chapter.

## Generator'ing Values

In the previous section, we mentioned an interesting use for generators, as a way to produce values. This is **not** the main focus in this chapter, but we'd be remiss if we didn't cover the basics, especially because this use case is essentially the origin of the name: generators.

We're going to take a slight diversion into the topic of *iterators* for a bit, but we'll circle back to how they relate to generators and using a generator to *generate* values.

## Producers and Iterators

Imagine you're producing a series of values where each value has a definable relationship to the previous value. To do this, you're going to need a stateful producer that remembers the last value it gave out.

You can implement something like that straightforwardly using a function closure (see the *Scope & Closures* title of this series):

```

var gimmeSomething = (function(){
  var nextVal;

  return function(){
    if (nextVal === undefined) {
      nextVal = 1;
    } else {
      nextVal = (3 * nextVal) + 6;
    }

    return nextVal;
  };
})();

```

```
gimmeSomething();      // 1
gimmeSomething();      // 9
gimmeSomething();      // 33
gimmeSomething();      // 105
```

**Note:** The `nextVal` computation logic here could have been simplified, but conceptually, we don't want to calculate the *next value* (aka `nextVal`) until the `next` `gimmeSomething()` call happens, because in general that could be a resource-leaky design for producers of more persistent or resource-limited values than simple `number`s.

Generating an arbitrary number series isn't a terribly realistic example. But what if you were generating records from a data source? You could imagine much the same code.

In fact, this task is a very common design pattern, usually solved by iterators. An *iterator* is a well-defined interface for stepping through a series of values from a producer. The JS interface for iterators, as it is in most languages, is to call `next()` each time you want the next value from the producer.

We could implement the standard *iterator* interface for our number series producer:

```
var something = (function(){
  var nextVal;

  return {
    // needed for `for..of` loops
    [Symbol.iterator]: function(){ return this; },

    // standard iterator interface method
    next: function(){
      if (nextVal === undefined) {
        nextVal = 1;
      }
      else {
        nextVal = (3 * nextVal) + 6;
      }

      return { done:false, value:nextVal };
    }
  };
})();

something.next().value;      // 1
something.next().value;      // 9
something.next().value;      // 33
something.next().value;      // 105
```

**Note:** We'll explain why we need the `[Symbol.iterator]: ...` part of this code snippet in the "Iterables" section. Syntactically though, two ES6 features are at play. First, the `[ ... ]` syntax is called a *computed property name* (see the *this & Object Prototypes* title of this series). It's a way in an object literal definition to specify an expression and use the result of that expression as the name for the property. Next, `Symbol.iterator` is one of ES6's predefined special `Symbol` values (see the *ES6 & Beyond* title of this book series).

The `next()` call returns an object with two properties: `done` is a `boolean` value signaling the *iterator*'s complete status; `value` holds the iteration value.

ES6 also adds the `for..of` loop, which means that a standard *iterator* can automatically be consumed with native loop syntax:

```
for (var v of something) {
  console.log( v );

  // don't let the loop run forever!
  if (v > 500) {
    break;
  }
}
```

```

}
// 1 9 33 105 321 969

```

**Note:** Because our `something iterator` always returns `done:false`, this `for..of` loop would run forever, which is why we put the `break` conditional in. It's totally OK for iterators to be never-ending, but there are also cases where the `iterator` will run over a finite set of values and eventually return a `done:true`.

The `for..of` loop automatically calls `next()` for each iteration -- it doesn't pass any values in to the `next()` -- and it will automatically terminate on receiving a `done:true`. It's quite handy for looping over a set of data.

Of course, you could manually loop over iterators, calling `next()` and checking for the `done:true` condition to know when to stop:

```

for (
  var ret;
  (ret = something.next()) && !ret.done;
) {
  console.log( ret.value );

  // don't let the loop run forever!
  if (ret.value > 500) {
    break;
  }
}
// 1 9 33 105 321 969

```

**Note:** This manual `for` approach is certainly uglier than the ES6 `for..of` loop syntax, but its advantage is that it affords you the opportunity to pass in values to the `next(..)` calls if necessary.

In addition to making your own `iterators`, many built-in data structures in JS (as of ES6), like `array`s, also have default `iterators`:

```

var a = [1,3,5,7,9];

for (var v of a) {
  console.log( v );
}
// 1 3 5 7 9

```

The `for..of` loop asks `a` for its `iterator`, and automatically uses it to iterate over `a`'s values.

**Note:** It may seem a strange omission by ES6, but regular `object`s intentionally do not come with a default `iterator` the way `array`s do. The reasons go deeper than we will cover here. If all you want is to iterate over the properties of an object (with no particular guarantee of ordering), `Object.keys(..)` returns an `array`, which can then be used like `for (var k of Object.keys(obj)) { ... }`. Such a `for..of` loop over an object's keys would be similar to a `for..in` loop, except that `Object.keys(..)` does not include properties from the `[[Prototype]]` chain while `for..in` does (see the *this & Object Prototypes* title of this series).

## Iterables

The `something` object in our running example is called an `iterator`, as it has the `next()` method on its interface. But a closely related term is `iterable`, which is an `object` that **contains** an `iterator` that can iterate over its values.

As of ES6, the way to retrieve an `iterator` from an `iterable` is that the `iterable` must have a function on it, with the name being the special ES6 symbol value `Symbol.iterator`. When this function is called, it returns an `iterator`. Though not required, generally each call should return a fresh new `iterator`.

a in the previous snippet is an *iterable*. The `for..of` loop automatically calls its `Symbol.iterator` function to construct an *iterator*. But we could of course call the function manually, and use the *iterator* it returns:

```
var a = [1,3,5,7,9];

var it = a[Symbol.iterator]();

it.next().value;    // 1
it.next().value;    // 3
it.next().value;    // 5
..
```

In the previous code listing that defined `something`, you may have noticed this line:

```
[Symbol.iterator]: function(){ return this; }
```

That little bit of confusing code is making the `something` value -- the interface of the `something iterator` -- also an *iterable*; it's now both an *iterable* and an *iterator*. Then, we pass `something` to the `for..of` loop:

```
for (var v of something) {
  ..
}
```

The `for..of` loop expects `something` to be an *iterable*, so it looks for and calls its `Symbol.iterator` function. We defined that function to simply `return this`, so it just gives itself back, and the `for..of` loop is none the wiser.

## Generator Iterator

Let's turn our attention back to generators, in the context of *iterators*. A generator can be treated as a producer of values that we extract one at a time through an *iterator* interface's `next()` calls.

So, a generator itself is not technically an *iterable*, though it's very similar -- when you execute the generator, you get an *iterator* back:

```
function *foo(){ .. }

var it = foo();
```

We can implement the `something` infinite number series producer from earlier with a generator, like this:

```
function *something() {
  var nextVal;

  while (true) {
    if (nextVal === undefined) {
      nextVal = 1;
    }
    else {
      nextVal = (3 * nextVal) + 6;
    }

    yield nextVal;
  }
}
```

**Note:** A `while..true` loop would normally be a very bad thing to include in a real JS program, at least if it doesn't have a `break` or `return` in it, as it would likely run forever, synchronously, and block/lock-up the browser UI. However, in a generator, such a loop is generally totally OK if it has a `yield` in it, as the generator will pause at each iteration, `yield`ing back to the main program and/or to the event loop queue. To put it glibly, "generators put the `while..true` back in JS programming!"

That's a fair bit cleaner and simpler, right? Because the generator pauses at each `yield`, the state (scope) of the function `*something()` is kept around, meaning there's no need for the closure boilerplate to preserve variable state across calls.

Not only is it simpler code -- we don't have to make our own *iterator* interface -- it actually is more reasonable code, because it more clearly expresses the intent. For example, the `while..true` loop tells us the generator is intended to run forever -- to keep *generating* values as long as we keep asking for them.

And now we can use our shiny new `*something()` generator with a `for..of` loop, and you'll see it works basically identically:

```
for (var v of something()) {
  console.log( v );

  // don't let the loop run forever!
  if (v > 500) {
    break;
  }
}
// 1 9 33 105 321 969
```

But don't skip over `for (var v of something()) ... !` We didn't just reference `something` as a value like in earlier examples, but instead called the `*something()` generator to get its *iterator* for the `for..of` loop to use.

If you're paying close attention, two questions may arise from this interaction between the generator and the loop:

- Why couldn't we say `for (var v of something) ... ?` Because `something` here is a generator, which is not an *iterable*. We have to call `something()` to construct a producer for the `for..of` loop to iterate over.
- The `something()` call produces an *iterator*, but the `for..of` loop wants an *iterable*, right? Yep. The generator's *iterator* also has a `Symbol.iterator` function on it, which basically does a `return this`, just like the `something iterable` we defined earlier. In other words, a generator's *iterator* is also an *iterable*!

## Stopping the Generator

In the previous example, it would appear the *iterator* instance for the `*something()` generator was basically left in a suspended state forever after the `break` in the loop was called.

But there's a hidden behavior that takes care of that for you. "Abnormal completion" (i.e., "early termination") of the `for..of` loop -- generally caused by a `break`, `return`, or an uncaught exception -- sends a signal to the generator's *iterator* for it to terminate.

**Note:** Technically, the `for..of` loop also sends this signal to the *iterator* at the normal completion of the loop. For a generator, that's essentially a moot operation, as the generator's *iterator* had to complete first so the `for..of` loop completed. However, custom *iterators* might desire to receive this additional signal from `for..of` loop consumers.

While a `for..of` loop will automatically send this signal, you may wish to send the signal manually to an *iterator*; you do this by calling `return(..)`.

If you specify a `try..finally` clause inside the generator, it will always be run even when the generator is externally completed. This is useful if you need to clean up resources (database connections, etc.):

```
function *something() {
  try {
    var nextVal;

    while (true) {
```

```

        if (nextVal === undefined) {
            nextVal = 1;
        }
        else {
            nextVal = (3 * nextVal) + 6;
        }

        yield nextVal;
    }
}
// cleanup clause
finally {
    console.log( "cleaning up!" );
}
}

```

The earlier example with `break` in the `for..of` loop will trigger the `finally` clause. But you could instead manually terminate the generator's `iterator` instance from the outside with `return(..)` :

```

var it = something();
for (var v of it) {
    console.log( v );

    // don't let the loop run forever!
    if (v > 500) {
        console.log(
            // complete the generator's iterator
            it.return( "Hello World" ).value
        );
        // no `break` needed here
    }
}
// 1 9 33 105 321 969
// cleaning up!
// Hello World

```

When we call `it.return(..)`, it immediately terminates the generator, which of course runs the `finally` clause. Also, it sets the returned `value` to whatever you passed in to `return(..)`, which is how `"Hello World"` comes right back out. We also don't need to include a `break` now because the generator's `iterator` is set to `done:true`, so the `for..of` loop will terminate on its next iteration.

Generators owe their namesake mostly to this *consuming produced values* use. But again, that's just one of the uses for generators, and frankly not even the main one we're concerned with in the context of this book.

But now that we more fully understand some of the mechanics of how they work, we can *next* turn our attention to how generators apply to async concurrency.

## Iterating Generators Asynchronously

What do generators have to do with async coding patterns, fixing problems with callbacks, and the like? Let's get to answering that important question.

We should revisit one of our scenarios from Chapter 3. Let's recall the callback approach:

```

function foo(x,y,cb) {
    ajax(
        "http://some.url.1/?x=" + x + "&y=" + y,
        cb
    );
}

```

```
foo( 11, 31, function(err, text) {
  if (err) {
    console.error( err );
  }
  else {
    console.log( text );
  }
});
```

If we wanted to express this same task flow control with a generator, we could do:

```
function foo(x,y) {
  ajax(
    "http://some.url.1/?x=" + x + "&y=" + y,
    function(err,data){
      if (err) {
        // throw an error into `*main()`
        it.throw( err );
      }
      else {
        // resume `*main()` with received `data`
        it.next( data );
      }
    }
  );
}

function *main() {
  try {
    var text = yield foo( 11, 31 );
    console.log( text );
  }
  catch (err) {
    console.error( err );
  }
}

var it = main();

// start it all up!
it.next();
```

At first glance, this snippet is longer, and perhaps a little more complex looking, than the callback snippet before it. But don't let that impression get you off track. The generator snippet is actually **much** better! But there's a lot going on for us to explain.

First, let's look at this part of the code, which is the most important:

```
var text = yield foo( 11, 31 );
console.log( text );
```

Think about how that code works for a moment. We're calling a normal function `foo(..)` and we're apparently able to get back the `text` from the Ajax call, even though it's asynchronous.

How is that possible? If you recall the beginning of Chapter 1, we had almost identical code:

```
var data = ajax( "...url 1.." );
console.log( data );
```

And that code didn't work! Can you spot the difference? It's the `yield` used in a generator.

That's the magic! That's what allows us to have what appears to be blocking, synchronous code, but it doesn't actually block the whole program; it only pauses/blocks the code in the generator itself.

In `yield foo(11, 31)`, first the `foo(11, 31)` call is made, which returns nothing (aka `undefined`), so we're making a call to request data, but we're actually then doing `yield undefined`. That's OK, because the code is not currently relying on a `yield` ed value to do anything interesting. We'll revisit this point later in the chapter.

We're not using `yield` in a message passing sense here, only in a flow control sense to pause/block. Actually, it will have message passing, but only in one direction, after the generator is resumed.

So, the generator pauses at the `yield`, essentially asking the question, "what value should I return to assign to the variable `text`?" Who's going to answer that question?

Look at `foo(..)`. If the Ajax request is successful, we call:

```
it.next( data );
```

That's resuming the generator with the response data, which means that our paused `yield` expression receives that value directly, and then as it restarts the generator code, that value gets assigned to the local variable `text`.

Pretty cool, huh?

Take a step back and consider the implications. We have totally synchronous-looking code inside the generator (other than the `yield` keyword itself), but hidden behind the scenes, inside of `foo(..)`, the operations can complete asynchronously.

**That's huge!** That's a nearly perfect solution to our previously stated problem with callbacks not being able to express asynchrony in a sequential, synchronous fashion that our brains can relate to.

In essence, we are abstracting the asynchrony away as an implementation detail, so that we can reason synchronously/sequentially about our flow control: "Make an Ajax request, and when it finishes print out the response." And of course, we just expressed two steps in the flow control, but this same capability extends without bounds, to let us express however many steps we need to.

**Tip:** This is such an important realization, just go back and read the last three paragraphs again to let it sink in!

## Synchronous Error Handling

But the preceding generator code has even more goodness to `yield` to us. Let's turn our attention to the `try..catch` inside the generator:

```
try {
  var text = yield foo( 11, 31 );
  console.log( text );
}
catch (err) {
  console.error( err );
}
```

How does this work? The `foo(..)` call is asynchronously completing, and doesn't `try..catch` fail to catch asynchronous errors, as we looked at in Chapter 3?

We already saw how the `yield` lets the assignment statement pause to wait for `foo(..)` to finish, so that the completed response can be assigned to `text`. The awesome part is that this `yield` pausing *also* allows the generator to `catch` an error. We throw that error into the generator with this part of the earlier code listing:

```
if (err) {
  // throw an error into `*main()`
  it.throw( err );
}
```

The `yield`-pause nature of generators means that not only do we get synchronous-looking `return` values from async function calls, but we can also synchronously `catch` errors from those async function calls!

So we've seen we can throw errors *into* a generator, but what about throwing errors *out of* a generator? Exactly as you'd expect:

```
function *main() {
  var x = yield "Hello World";

  yield x.toLowerCase(); // cause an exception!
}

var it = main();

it.next().value; // Hello World

try {
  it.next( 42 );
}
catch (err) {
  console.error( err ); // TypeError
}
```

Of course, we could have manually thrown an error with `throw ..` instead of causing an exception.

We can even `catch` the same error that we `throw(..)` into the generator, essentially giving the generator a chance to handle it but if it doesn't, the *iterator* code must handle it:

```
function *main() {
  var x = yield "Hello World";

  // never gets here
  console.log( x );
}

var it = main();

it.next();

try {
  // will `*main()` handle this error? we'll see!
  it.throw( "Oops" );
}
catch (err) {
  // nope, didn't handle it!
  console.error( err ); // Oops
}
```

Synchronous-looking error handling (via `try..catch`) with async code is a huge win for readability and reason-ability.

## Generators + Promises

In our previous discussion, we showed how generators can be iterated asynchronously, which is a huge step forward in sequential reason-ability over the spaghetti mess of callbacks. But we lost something very important: the trustability and composable of Promises (see Chapter 3)!

Don't worry -- we can get that back. The best of all worlds in ES6 is to combine generators (synchronous-looking async code) with Promises (trustable and composable).

But how?

Recall from Chapter 3 the Promise-based approach to our running Ajax example:

```

function foo(x,y) {
  return request(
    "http://some.url.1/?x=" + x + "&y=" + y
  );
}

foo( 11, 31 )
.then(
  function(text){
    console.log( text );
  },
  function(err){
    console.error( err );
  }
);

```

In our earlier generator code for the running Ajax example, `foo(..)` returned nothing (`undefined`), and our *iterator* control code didn't care about that `yield`ed value.

But here the Promise-aware `foo(..)` returns a promise after making the Ajax call. That suggests that we could construct a promise with `foo(..)` and then `yield` it from the generator, and then the *iterator* control code would receive that promise.

But what should the *iterator* do with the promise?

It should listen for the promise to resolve (fulfillment or rejection), and then either resume the generator with the fulfillment message or throw an error into the generator with the rejection reason.

Let me repeat that, because it's so important. The natural way to get the most out of Promises and generators is **to yield a Promise**, and wire that Promise to control the generator's *iterator*.

Let's give it a try! First, we'll put the Promise-aware `foo(..)` together with the generator `*main()`:

```

function foo(x,y) {
  return request(
    "http://some.url.1/?x=" + x + "&y=" + y
  );
}

function *main() {
  try {
    var text = yield foo( 11, 31 );
    console.log( text );
  }
  catch (err) {
    console.error( err );
  }
}

```

The most powerful revelation in this refactor is that the code inside `*main()` **did not have to change at all!** Inside the generator, whatever values are `yield`ed out is just an opaque implementation detail, so we're not even aware it's happening, nor do we need to worry about it.

But how are we going to run `*main()` now? We still have some of the implementation plumbing work to do, to receive and wire up the `yield`ed promise so that it resumes the generator upon resolution. We'll start by trying that manually:

```

var it = main();
var p = it.next().value;

// wait for the `p` promise to resolve
p.then(
  function(text){

```

```

        it.next( text );
    },
    function(err){
        it.throw( err );
    }
);

```

Actually, that wasn't so painful at all, was it?

This snippet should look very similar to what we did earlier with the manually wired generator controlled by the error-first callback. Instead of an `if (err) { it.throw...}`, the promise already splits fulfillment (success) and rejection (failure) for us, but otherwise the `iterator` control is identical.

Now, we've glossed over some important details.

Most importantly, we took advantage of the fact that we knew that `*main()` only had one Promise-aware step in it. What if we wanted to be able to Promise-drive a generator no matter how many steps it has? We certainly don't want to manually write out the Promise chain differently for each generator! What would be much nicer is if there was a way to repeat (aka "loop" over) the iteration control, and each time a Promise comes out, wait on its resolution before continuing.

Also, what if the generator throws out an error (intentionally or accidentally) during the `it.next(..)` call? Should we quit, or should we `catch` it and send it right back in? Similarly, what if we `it.throw(..)` a Promise rejection into the generator, but it's not handled, and comes right back out?

## Promise-Aware Generator Runner

The more you start to explore this path, the more you realize, "wow, it'd be great if there was just some utility to do it for me." And you're absolutely correct. This is such an important pattern, and you don't want to get it wrong (or exhaust yourself repeating it over and over), so your best bet is to use a utility that is specifically designed to `run` Promise-`yield`ing generators in the manner we've illustrated.

Several Promise abstraction libraries provide just such a utility, including my `asyncquence` library and its `runner(..)`, which will be discussed in Appendix A of this book.

But for the sake of learning and illustration, let's just define our own standalone utility that we'll call `run(..)`:

```

// thanks to Benjamin Gruenbaum (@benjammingr on GitHub) for
// big improvements here!
function run(gen) {
    var args = [].slice.call( arguments, 1 ), it;

    // initialize the generator in the current context
    it = gen.apply( this, args );

    // return a promise for the generator completing
    return Promise.resolve()
        .then( function handleNext(value){
            // run to the next yielded value
            var next = it.next( value );

            return (function handleResult(next){
                // generator has completed running?
                if (next.done) {
                    return next.value;
                }
                // otherwise keep going
                else {
                    return Promise.resolve( next.value )
                        .then(
                            // resume the async loop on
                            // success, sending the resolved
                            // value back into the generator

```

```

        handleNext,
        // if `value` is a rejected
        // promise, propagate error back
        // into the generator for its own
        // error handling
        function handleErr(err) {
            return Promise.resolve(
                it.throw( err )
            )
            .then( handleResult );
        }
    );
}
})(next);
} );
}

```

As you can see, it's a quite a bit more complex than you'd probably want to author yourself, and you especially wouldn't want to repeat this code for each generator you use. So, a utility/library helper is definitely the way to go. Nevertheless, I encourage you to spend a few minutes studying that code listing to get a better sense of how to manage the generator+Promise negotiation.

How would you use `run(..)` with `*main()` in our *running Ajax* example?

```

function *main() {
    // ...
}

run( main );

```

That's it! The way we wired `run(..)`, it will automatically advance the generator you pass to it, asynchronously until completion.

**Note:** The `run(..)` we defined returns a promise which is wired to resolve once the generator is complete, or receive an uncaught exception if the generator doesn't handle it. We don't show that capability here, but we'll come back to it later in the chapter.

## ES7: `async` and `await` ?

The preceding pattern -- generators yielding Promises that then control the generator's *iterator* to advance it to completion -- is such a powerful and useful approach, it would be nicer if we could do it without the clutter of the library utility helper (aka `run(..)`).

There's probably good news on that front. At the time of this writing, there's early but strong support for a proposal for more syntactic addition in this realm for the post-ES6, ES7-ish timeframe. Obviously, it's too early to guarantee the details, but there's a pretty decent chance it will shake out similar to the following:

```

function foo(x,y) {
    return request(
        "http://some.url.1/?x=" + x + "&y=" + y
    );
}

async function main() {
    try {
        var text = await foo( 11, 31 );
        console.log( text );
    }
    catch (err) {
        console.error( err );
    }
}

```

```

    }
main();

```

As you can see, there's no `run(..)` call (meaning no need for a library utility!) to invoke and drive `main()` -- it's just called as a normal function. Also, `main()` isn't declared as a generator function anymore; it's a new kind of function: `async function`. And finally, instead of `yield` ing a Promise, we `await` for it to resolve.

The `async function` automatically knows what to do if you `await` a Promise -- it will pause the function (just like with generators) until the Promise resolves. We didn't illustrate it in this snippet, but calling an `async` function like `main()` automatically returns a promise that's resolved whenever the function finishes completely.

**Tip:** The `async` / `await` syntax should look very familiar to readers with experience in C#, because it's basically identical.

The proposal essentially codifies support for the pattern we've already derived, into a syntactic mechanism: combining Promises with sync-looking flow control code. That's the best of both worlds combined, to effectively address practically all of the major concerns we outlined with callbacks.

The mere fact that such a ES7-ish proposal already exists and has early support and enthusiasm is a major vote of confidence in the future importance of this `async` pattern.

## Promise Concurrency in Generators

So far, all we've demonstrated is a single-step `async` flow with Promises+generators. But real-world code will often have many `async` steps.

If you're not careful, the sync-looking style of generators may lull you into complacency with how you structure your `async` concurrency, leading to suboptimal performance patterns. So we want to spend a little time exploring the options.

Imagine a scenario where you need to fetch data from two different sources, then combine those responses to make a third request, and finally print out the last response. We explored a similar scenario with Promises in Chapter 3, but let's reconsider it in the context of generators.

Your first instinct might be something like:

```

function *foo() {
  var r1 = yield request( "http://some.url.1" );
  var r2 = yield request( "http://some.url.2" );

  var r3 = yield request(
    "http://some.url.3/?v=" + r1 + "," + r2
  );

  console.log( r3 );
}

// use previously defined `run(..)` utility
run( foo );

```

This code will work, but in the specifics of our scenario, it's not optimal. Can you spot why?

Because the `r1` and `r2` requests can -- and for performance reasons, *should* -- run concurrently, but in this code they will run sequentially; the `"http://some.url.2"` URL isn't Ajax fetched until after the `"http://some.url.1"` request is finished. These two requests are independent, so the better performance approach would likely be to have them run at the same time.

But how exactly would you do that with a generator and `yield`? We know that `yield` is only a single pause point in the code, so you can't really do two pauses at the same time.

The most natural and effective answer is to base the async flow on Promises, specifically on their capability to manage state in a time-independent fashion (see "Future Value" in Chapter 3).

The simplest approach:

```
function *foo() {
  // make both requests "in parallel"
  var p1 = request( "http://some.url.1" );
  var p2 = request( "http://some.url.2" );

  // wait until both promises resolve
  var r1 = yield p1;
  var r2 = yield p2;

  var r3 = yield request(
    "http://some.url.3/?v=" + r1 + "," + r2
  );
  console.log( r3 );
}

// use previously defined `run(..)` utility
run( foo );
```

Why is this different from the previous snippet? Look at where the `yield` is and is not. `p1` and `p2` are promises for Ajax requests made concurrently (aka "in parallel"). It doesn't matter which one finishes first, because promises will hold onto their resolved state for as long as necessary.

Then we use two subsequent `yield` statements to wait for and retrieve the resolutions from the promises (into `r1` and `r2`, respectively). If `p1` resolves first, the `yield p1` resumes first then waits on the `yield p2` to resume. If `p2` resolves first, it will just patiently hold onto that resolution value until asked, but the `yield p1` will hold on first, until `p1` resolves.

Either way, both `p1` and `p2` will run concurrently, and both have to finish, in either order, before the `r3 = yield request..` Ajax request will be made.

If that flow control processing model sounds familiar, it's basically the same as what we identified in Chapter 3 as the "gate" pattern, enabled by the `Promise.all([ ... ])` utility. So, we could also express the flow control like this:

```
function *foo() {
  // make both requests "in parallel," and
  // wait until both promises resolve
  var results = yield Promise.all( [
    request( "http://some.url.1" ),
    request( "http://some.url.2" )
  ]);

  var r1 = results[0];
  var r2 = results[1];

  var r3 = yield request(
    "http://some.url.3/?v=" + r1 + "," + r2
  );
  console.log( r3 );
}

// use previously defined `run(..)` utility
run( foo );
```

**Note:** As we discussed in Chapter 3, we can even use ES6 destructuring assignment to simplify the `var r1 = ... var r2 = ...` assignments, with `var [r1,r2] = results`.

In other words, all of the concurrency capabilities of Promises are available to us in the generator+Promise approach. So in any place where you need more than sequential this-then-that async flow control steps, Promises are likely your best bet.

## Promises, Hidden

As a word of stylistic caution, be careful about how much Promise logic you include **inside your generators**. The whole point of using generators for asynchrony in the way we've described is to create simple, sequential, sync-looking code, and to hide as much of the details of asynchrony away from that code as possible.

For example, this might be a cleaner approach:

```
// note: normal function, not generator
function bar(url1,url2) {
  return Promise.all( [
    request( url1 ),
    request( url2 )
  ]);
}

function *foo() {
  // hide the Promise-based concurrency details
  // inside `bar(..)`
  var results = yield bar(
    "http://some.url.1",
    "http://some.url.2"
  );

  var r1 = results[0];
  var r2 = results[1];

  var r3 = yield request(
    "http://some.url.3/?v=" + r1 + "," + r2
  );

  console.log( r3 );
}

// use previously defined `run(..)` utility
run( foo );
```

Inside `*foo()`, it's cleaner and clearer that all we're doing is just asking `bar(..)` to get us some `results`, and we'll `yield` - wait on that to happen. We don't have to care that under the covers a `Promise.all([ .. ])` Promise composition will be used to make that happen.

**We treat asynchrony, and indeed Promises, as an implementation detail.**

Hiding your Promise logic inside a function that you merely call from your generator is especially useful if you're going to do a sophisticated series flow-control. For example:

```
function bar() {
  return Promise.all( [
    baz( ... )
    .then( ... ),
    Promise.race( [ ... ] )
  ])
  .then( ... )
}
```

That kind of logic is sometimes required, and if you dump it directly inside your generator(s), you've defeated most of the reason why you would want to use generators in the first place. We *should* intentionally abstract such details away from our generator code so that they don't clutter up the higher level task expression.

Beyond creating code that is both functional and performant, you should also strive to make code that is as reasonable and maintainable as possible.

**Note:** Abstraction is not *always* a healthy thing for programming -- many times it can increase complexity in exchange for terseness. But in this case, I believe it's much healthier for your generator+Promise async code than the alternatives. As with all such advice, though, pay attention to your specific situations and make proper decisions for you and your team.

## Generator Delegation

In the previous section, we showed calling regular functions from inside a generator, and how that remains a useful technique for abstracting away implementation details (like async Promise flow). But the main drawback of using a normal function for this task is that it has to behave by the normal function rules, which means it cannot pause itself with `yield` like a generator can.

It may then occur to you that you might try to call one generator from another generator, using our `run(..)` helper, such as:

```
function *foo() {
  var r2 = yield request( "http://some.url.2" );
  var r3 = yield request( "http://some.url.3/?v=" + r2 );

  return r3;
}

function *bar() {
  var r1 = yield request( "http://some.url.1" );

  // "delegating" to `*foo()` via `run(..)`
  var r3 = yield run( foo );

  console.log( r3 );
}

run( bar );
```

We run `*foo()` inside of `*bar()` by using our `run(..)` utility again. We take advantage here of the fact that the `run(..)` we defined earlier returns a promise which is resolved when its generator is run to completion (or errors out), so if we `yield` out to a `run(..)` instance the promise from another `run(..)` call, it automatically pauses `*bar()` until `*foo()` finishes.

But there's an even better way to integrate calling `*foo()` into `*bar()`, and it's called `yield`-delegation. The special syntax for `yield`-delegation is: `yield * __` (notice the extra `*`). Before we see it work in our previous example, let's look at a simpler scenario:

```
function *foo() {
  console.log( `'*foo()' starting` );
  yield 3;
  yield 4;
  console.log( `'*foo()' finished` );
}

function *bar() {
  yield 1;
  yield 2;
  yield *foo();    // `yield`-delegation!
  yield 5;
}

var it = bar();

it.next().value;    // 1
it.next().value;    // 2
it.next().value;    // `'*foo()' starting
                  // 3
```

```
it.next().value;    // 4
it.next().value;    // `*foo()` finished
                  // 5
```

**Note:** Similar to a note earlier in the chapter where I explained why I prefer `function *foo() ...` instead of `function* foo()` ..., I also prefer -- differing from most other documentation on the topic -- to say `yield *foo()` instead of `yield* foo()`. The placement of the `*` is purely stylistic and up to your best judgment. But I find the consistency of styling attractive.

How does the `yield *foo()` delegation work?

First, calling `foo()` creates an *iterator* exactly as we've already seen. Then, `yield *` delegates/transfers the *iterator* instance control (of the present `*bar()` generator) over to this other `*foo()` *iterator*.

So, the first two `it.next()` calls are controlling `*bar()`, but when we make the third `it.next()` call, now `*foo()` starts up, and now we're controlling `*foo()` instead of `*bar()`. That's why it's called delegation -- `*bar()` delegated its iteration control to `*foo()`.

As soon as the `it iterator` control exhausts the entire `*foo()` *iterator*, it automatically returns to controlling `*bar()`.

So now back to the previous example with the three sequential Ajax requests:

```
function *foo() {
  var r2 = yield request( "http://some.url.2" );
  var r3 = yield request( "http://some.url.3/?v=" + r2 );

  return r3;
}

function *bar() {
  var r1 = yield request( "http://some.url.1" );

  // "delegating" to `*foo()` via `yield*`
  var r3 = yield *foo();

  console.log( r3 );
}

run( bar );
```

The only difference between this snippet and the version used earlier is the use of `yield *foo()` instead of the previous `yield run(foo)`.

**Note:** `yield *` yields iteration control, not generator control; when you invoke the `*foo()` generator, you're now `yield`-delegating to its *iterator*. But you can actually `yield`-delegate to any *iterable*; `yield *[1,2,3]` would consume the default *iterator* for the `[1,2,3]` array value.

## Why Delegation?

The purpose of `yield`-delegation is mostly code organization, and in that way is symmetrical with normal function calling.

Imagine two modules that respectively provide methods `foo()` and `bar()`, where `bar()` calls `foo()`. The reason the two are separate is generally because the proper organization of code for the program calls for them to be in separate functions. For example, there may be cases where `foo()` is called standalone, and other places where `bar()` calls `foo()`.

For all these exact same reasons, keeping generators separate aids in program readability, maintenance, and debuggability. In that respect, `yield *` is a syntactic shortcut for manually iterating over the steps of `*foo()` while inside of `*bar()`.

Such manual approach would be especially complex if the steps in `*foo()` were asynchronous, which is why you'd probably need to use that `run(..)` utility to do it. And as we've shown, `yield *foo()` eliminates the need for a sub-instance of the `run(..)` utility (like `run(foo)`).

## Delegating Messages

You may wonder how this `yield`-delegation works not just with `iterator` control but with the two-way message passing. Carefully follow the flow of messages in and out, through the `yield`-delegation:

```
function *foo() {
  console.log( "inside `*foo()`:", yield "B" );

  console.log( "inside `*foo()`:", yield "C" );

  return "D";
}

function *bar() {
  console.log( "inside `*bar()`:", yield "A" );

  // `yield`-delegation!
  console.log( "inside `*bar()`:", yield *foo() );

  console.log( "inside `*bar()`:", yield "E" );

  return "F";
}

var it = bar();

console.log( "outside:", it.next().value );
// outside: A

console.log( "outside:", it.next( 1 ).value );
// inside `*bar()`: 1
// outside: B

console.log( "outside:", it.next( 2 ).value );
// inside `*foo()`: 2
// outside: C

console.log( "outside:", it.next( 3 ).value );
// inside `*foo()`: 3
// inside `*bar()`: D
// outside: E

console.log( "outside:", it.next( 4 ).value );
// inside `*bar()`: 4
// outside: F
```

Pay particular attention to the processing steps after the `it.next(3)` call:

1. The `3` value is passed (through the `yield`-delegation in `*bar()`) into the waiting `yield "C"` expression inside of `*foo()`.
2. `*foo()` then calls `return "D"`, but this value doesn't get returned all the way back to the outside `it.next(3)` call.
3. Instead, the `"D"` value is sent as the result of the waiting `yield *foo()` expression inside of `*bar()` -- this `yield`-delegation expression has essentially been paused while all of `*foo()` was exhausted. So `"D"` ends up inside of `*bar()` for it to print out.
4. `yield "E"` is called inside of `*bar()`, and the `"E"` value is yielded to the outside as the result of the `it.next(3)` call.

From the perspective of the external `iterator` (`it`), it doesn't appear any differently between controlling the initial generator or a delegated one.

In fact, `yield`-delegation doesn't even have to be directed to another generator; it can just be directed to a non-generator, general `iterable`. For example:

```
function *bar() {
```

```

console.log( "inside `*bar()`:", yield "A" );

// `yield`-delegation to a non-generator!
console.log( "inside `*bar()`:", yield *[ "B", "C", "D" ] );

console.log( "inside `*bar()`:", yield "E" );

return "F";
}

var it = bar();

console.log( "outside:", it.next().value );
// outside: A

console.log( "outside:", it.next( 1 ).value );
// inside `*bar()`: 1
// outside: B

console.log( "outside:", it.next( 2 ).value );
// outside: C

console.log( "outside:", it.next( 3 ).value );
// outside: D

console.log( "outside:", it.next( 4 ).value );
// inside `*bar()`: undefined
// outside: E

console.log( "outside:", it.next( 5 ).value );
// inside `*bar()`: 5
// outside: F

```

Notice the differences in where the messages were received/reported between this example and the one previous.

Most strikingly, the default `array iterator` doesn't care about any messages sent in via `next(..)` calls, so the values `2`, `3`, and `4` are essentially ignored. Also, because that `iterator` has no explicit `return` value (unlike the previously used `*foo()`), the `yield *` expression gets an `undefined` when it finishes.

## Exceptions Delegated, Too!

In the same way that `yield`-delegation transparently passes messages through in both directions, errors/exceptions also pass in both directions:

```

function *foo() {
  try {
    yield "B";
  }
  catch (err) {
    console.log( "error caught inside `*foo()`:", err );
  }

  yield "C";

  throw "D";
}

function *bar() {
  yield "A";

  try {
    yield *foo();
  }
  catch (err) {
    console.log( "error caught inside `*bar()`:", err );
  }
}

```

```

        }

        yield "E";

        yield *baz();

        // note: can't get here!
        yield "G";
    }

    function *baz() {
        throw "F";
    }

    var it = bar();

    console.log( "outside:", it.next().value );
    // outside: A

    console.log( "outside:", it.next( 1 ).value );
    // outside: B

    console.log( "outside:", it.throw( 2 ).value );
    // error caught inside `*foo()`: 2
    // outside: C

    console.log( "outside:", it.next( 3 ).value );
    // error caught inside `*bar()`: D
    // outside: E

    try {
        console.log( "outside:", it.next( 4 ).value );
    }
    catch (err) {
        console.log( "error caught outside:", err );
    }
    // error caught outside: F

```

Some things to note from this snippet:

1. When we call `it.throw(2)`, it sends the error message `2` into `*bar()`, which delegates that to `*foo()`, which then catches it and handles it gracefully. Then, the `yield "C"` sends `"C"` back out as the return `value` from the `it.throw(2)` call.
2. The `"D"` value that's next `throw n` from inside `*foo()` propagates out to `*bar()`, which catches it and handles it gracefully. Then the `yield "E"` sends `"E"` back out as the return `value` from the `it.next(3)` call.
3. Next, the exception `throw n` from `*baz()` isn't caught in `*bar()` -- though we did catch it outside -- so both `*baz()` and `*bar()` are set to a completed state. After this snippet, you would not be able to get the `"G"` value out with any subsequent `next(..)` call(s) -- they will just return `undefined` for `value`.

## Delegating Asynchrony

Let's finally get back to our earlier `yield`-delegation example with the multiple sequential Ajax requests:

```

function *foo() {
    var r2 = yield request( "http://some.url.2" );
    var r3 = yield request( "http://some.url.3/?v=" + r2 );

    return r3;
}

function *bar() {
    var r1 = yield request( "http://some.url.1" );

    var r3 = yield *foo();
}

```

```

    console.log( r3 );
}

run( bar );

```

Instead of calling `yield run(foo)` inside of `*bar()`, we just call `yield *foo()`.

In the previous version of this example, the Promise mechanism (controlled by `run(..)`) was used to transport the value from `return r3` in `*foo()` to the local variable `r3` inside `*bar()`. Now, that value is just returned back directly via the `yield *` mechanics.

Otherwise, the behavior is pretty much identical.

## Delegating "Recursion"

Of course, `yield`-delegation can keep following as many delegation steps as you wire up. You could even use `yield`-delegation for async-capable generator "recursion" -- a generator `yield`-delegating to itself:

```

function *foo(val) {
  if (val > 1) {
    // generator recursion
    val = yield *foo( val - 1 );
  }

  return yield request( "http://some.url/?v=" + val );
}

function *bar() {
  var r1 = yield *foo( 3 );
  console.log( r1 );
}

run( bar );

```

**Note:** Our `run(..)` utility could have been called with `run( foo, 3 )`, because it supports additional parameters being passed along to the initialization of the generator. However, we used a parameter-free `*bar()` here to highlight the flexibility of `yield *`.

What processing steps follow from that code? Hang on, this is going to be quite intricate to describe in detail:

1. `run(bar)` starts up the `*bar()` generator.
2. `foo(3)` creates an *iterator* for `*foo(..)` and passes `3` as its `val` parameter.
3. Because `3 > 1`, `foo(2)` creates another *iterator* and passes in `2` as its `val` parameter.
4. Because `2 > 1`, `foo(1)` creates yet another *iterator* and passes in `1` as its `val` parameter.
5. `1 > 1` is `false`, so we next call `request(..)` with the `1` value, and get a promise back for that first Ajax call.
6. That promise is `yield`ed out, which comes back to the `*foo(2)` generator instance.
7. The `yield *` passes that promise back out to the `*foo(3)` generator instance. Another `yield *` passes the promise out to the `*bar()` generator instance. And yet again another `yield *` passes the promise out to the `run(..)` utility, which will wait on that promise (for the first Ajax request) to proceed.
8. When the promise resolves, its fulfillment message is sent to resume `*bar()`, which passes through the `yield *` into the `*foo(3)` instance, which then passes through the `yield *` to the `*foo(2)` generator instance, which then passes through the `yield *` to the normal `yield` that's waiting in the `*foo(3)` generator instance.
9. That first call's Ajax response is now immediately `return`ed from the `*foo(3)` generator instance, which sends that value back as the result of the `yield *` expression in the `*foo(2)` instance, and assigned to its local `val` variable.
10. Inside `*foo(2)`, a second Ajax request is made with `request(..)`, whose promise is `yield`ed back to the `*foo(1)` instance, and then `yield *` propagates all the way out to `run(..)` (step 7 again). When the promise resolves, the second

Ajax response propagates all the way back into the `*foo(2)` generator instance, and is assigned to its local `val` variable.

11. Finally, the third Ajax request is made with `request(..)`, its promise goes out to `run(..)`, and then its resolution value comes all the way back, which is then `return`ed so that it comes back to the waiting `yield *` expression in `*bar()`.

Phew! A lot of crazy mental juggling, huh? You might want to read through that a few more times, and then go grab a snack to clear your head!

## Generator Concurrency

As we discussed in both Chapter 1 and earlier in this chapter, two simultaneously running "processes" can cooperatively interleave their operations, and many times this can *yield* (pun intended) very powerful asynchrony expressions.

Frankly, our earlier examples of concurrency interleaving of multiple generators showed how to make it really confusing. But we hinted that there's places where this capability is quite useful.

Recall a scenario we looked at in Chapter 1, where two different simultaneous Ajax response handlers needed to coordinate with each other to make sure that the data communication was not a race condition. We slotted the responses into the `res` array like this:

```
function response(data) {
  if (data.url == "http://some.url.1") {
    res[0] = data;
  }
  else if (data.url == "http://some.url.2") {
    res[1] = data;
  }
}
```

But how can we use multiple generators concurrently for this scenario?

```
// `request(..)` is a Promise-aware Ajax utility

var res = [];

function *reqData(url) {
  res.push(
    yield request(url)
  );
}
```

**Note:** We're going to use two instances of the `*reqData(..)` generator here, but there's no difference to running a single instance of two different generators; both approaches are reasoned about identically. We'll see two different generators coordinating in just a bit.

Instead of having to manually sort out `res[0]` and `res[1]` assignments, we'll use coordinated ordering so that `res.push(..)` properly slots the values in the expected and predictable order. The expressed logic thus should feel a bit cleaner.

But how will we actually orchestrate this interaction? First, let's just do it manually, with Promises:

```
var it1 = reqData("http://some.url.1");
var it2 = reqData("http://some.url.2");

var p1 = it1.next().value;
var p2 = it2.next().value;

p1
.then(function(data){
  it1.next(data);
  return p2;
})
```

```

} )
.then( function(data){
    it2.next( data );
} );

```

`*reqData(..)`'s two instances are both started to make their Ajax requests, then paused with `yield`. Then we choose to resume the first instance when `p1` resolves, and then `p2`'s resolution will restart the second instance. In this way, we use Promise orchestration to ensure that `res[0]` will have the first response and `res[1]` will have the second response.

But frankly, this is awfully manual, and it doesn't really let the generators orchestrate themselves, which is where the true power can lie. Let's try it a different way:

```

// `request(..)` is a Promise-aware Ajax utility

var res = [];

function *reqData(url) {
    var data = yield request( url );

    // transfer control
    yield;

    res.push( data );
}

var it1 = reqData( "http://some.url.1" );
var it2 = reqData( "http://some.url.2" );

var p1 = it1.next().value;
var p2 = it2.next().value;

p1.then( function(data){
    it1.next( data );
} );

p2.then( function(data){
    it2.next( data );
} );

Promise.all( [p1,p2] )
.then( function(){
    it1.next();
    it2.next();
} );

```

OK, this is a bit better (though still manual!), because now the two instances of `*reqData(..)` run truly concurrently, and (at least for the first part) independently.

In the previous snippet, the second instance was not given its data until after the first instance was totally finished. But here, both instances receive their data as soon as their respective responses come back, and then each instance does another `yield` for control transfer purposes. We then choose what order to resume them in the `Promise.all([ .. ])` handler.

What may not be as obvious is that this approach hints at an easier form for a reusable utility, because of the symmetry. We can do even better. Let's imagine using a utility called `runAll(..)`:

```

// `request(..)` is a Promise-aware Ajax utility

var res = [];

runAll(
    function*(){
        var p1 = request( "http://some.url.1" );

```

```

    // transfer control
    yield;

    res.push( yield p1 );
},
function*(){
    var p2 = request( "http://some.url.2" );

    // transfer control
    yield;

    res.push( yield p2 );
}
);

```

**Note:** We're not including a code listing for `runAll(..)` as it is not only long enough to bog down the text, but is an extension of the logic we've already implemented in `run(..)` earlier. So, as a good supplementary exercise for the reader, try your hand at evolving the code from `run(..)` to work like the imagined `runAll(..)`. Also, my *asynquence* library provides a previously mentioned `runner(..)` utility with this kind of capability already built in, and will be discussed in Appendix A of this book.

Here's how the processing inside `runAll(..)` would operate:

1. The first generator gets a promise for the first Ajax response from `"http://some.url.1"`, then `yield`s control back to the `runAll(..)` utility.
2. The second generator runs and does the same for `"http://some.url.2"`, `yield`ing control back to the `runAll(..)` utility.
3. The first generator resumes, and then `yield`s out its promise `p1`. The `runAll(..)` utility does the same in this case as our previous `run(..)`, in that it waits on that promise to resolve, then resumes the same generator (no control transfer!). When `p1` resolves, `runAll(..)` resumes the first generator again with that resolution value, and then `res[0]` is given its value. When the first generator then finishes, that's an implicit transfer of control.
4. The second generator resumes, `yield`s out its promise `p2`, and waits for it to resolve. Once it does, `runAll(..)` resumes the second generator with that value, and `res[1]` is set.

In this running example, we use an outer variable called `res` to store the results of the two different Ajax responses -- that's our concurrency coordination making that possible.

But it might be quite helpful to further extend `runAll(..)` to provide an inner variable space for the multiple generator instances to *share*, such as an empty object we'll call `data` below. Also, it could take non-Promise values that are `yield`ed and hand them off to the next generator.

Consider:

```

// `request(..)` is a Promise-aware Ajax utility

runAll(
    function*(data){
        data.res = [];

        // transfer control (and message pass)
        var url1 = yield "http://some.url.2";

        var p1 = request( url1 ); // "http://some.url.1"

        // transfer control
        yield;

        data.res.push( yield p1 );
},
function*(data){
    // transfer control (and message pass)
    var url2 = yield "http://some.url.1";
}
);

```

```

        var p2 = request( url2 ); // "http://some.url.2"

        // transfer control
        yield;

        data.res.push( yield p2 );
    }
);

```

In this formulation, the two generators are not just coordinating control transfer, but actually communicating with each other, both through `data.res` and the `yield` ed messages that trade `url1` and `url2` values. That's incredibly powerful!

Such realization also serves as a conceptual base for a more sophisticated asynchrony technique called CSP (Communicating Sequential Processes), which we will cover in Appendix B of this book.

## Thunks

So far, we've made the assumption that `yield` ing a Promise from a generator -- and having that Promise resume the generator via a helper utility like `run(...)` -- was the best possible way to manage asynchrony with generators. To be clear, it is.

But we skipped over another pattern that has some mildly widespread adoption, so in the interest of completeness we'll take a brief look at it.

In general computer science, there's an old pre-JS concept called a "thunk." Without getting bogged down in the historical nature, a narrow expression of a thunk in JS is a function that -- without any parameters -- is wired to call another function.

In other words, you wrap a function definition around function call -- with any parameters it needs -- to *defer* the execution of that call, and that wrapping function is a thunk. When you later execute the thunk, you end up calling the original function.

For example:

```

function foo(x,y) {
  return x + y;
}

function fooThunk() {
  return foo( 3, 4 );
}

// later

console.log( fooThunk() ); // 7

```

So, a synchronous thunk is pretty straightforward. But what about an async thunk? We can essentially extend the narrow thunk definition to include it receiving a callback.

Consider:

```

function foo(x,y,cb) {
  setTimeout( function(){
    cb( x + y );
  }, 1000 );
}

function fooThunk(cb) {
  foo( 3, 4, cb );
}

// later

fooThunk( function(sum){

```

```
    console.log( sum );           // 7
} );
```

As you can see, `fooThunk(..)` only expects a `cb(..)` parameter, as it already has values `3` and `4` (for `x` and `y`, respectively) pre-specified and ready to pass to `foo(..)`. A thunk is just waiting around patiently for the last piece it needs to do its job: the callback.

You don't want to make thunks manually, though. So, let's invent a utility that does this wrapping for us.

Consider:

```
function thunkify(fn) {
  var args = [].slice.call( arguments, 1 );
  return function(cb) {
    args.push( cb );
    return fn.apply( null, args );
  };
}

var fooThunk = thunkify( foo, 3, 4 );

// later

fooThunk( function(sum) {
  console.log( sum );           // 7
} );
```

**Tip:** Here we assume that the original (`foo(..)`) function signature expects its callback in the last position, with any other parameters coming before it. This is a pretty ubiquitous "standard" for async JS function standards. You might call it "callback-last style." If for some reason you had a need to handle "callback-first style" signatures, you would just make a utility that used `args.unshift(..)` instead of `args.push(..)`.

The preceding formulation of `thunkify(..)` takes both the `foo(..)` function reference, and any parameters it needs, and returns back the thunk itself (`fooThunk(..)`). However, that's not the typical approach you'll find to thunks in JS.

Instead of `thunkify(..)` making the thunk itself, typically -- if not perplexingly -- the `thunkify(..)` utility would produce a function that produces thunks.

Uhhhh... yeah.

Consider:

```
function thunkify(fn) {
  return function() {
    var args = [].slice.call( arguments );
    return function(cb) {
      args.push( cb );
      return fn.apply( null, args );
    };
  };
}
```

The main difference here is the extra `return function() { .. }` layer. Here's how its usage differs:

```
var whatIsThis = thunkify( foo );

var fooThunk = whatIsThis( 3, 4 );

// later

fooThunk( function(sum) {
  console.log( sum );           // 7
```

```
 } );
```

Obviously, the big question this snippet implies is what is `whatIsThis` properly called? It's not the thunk, it's the thing that will produce thunks from `foo(..)` calls. It's kind of like a "factory" for "thunks." There doesn't seem to be any kind of standard agreement for naming such a thing.

So, my proposal is "thunkory" ("thunk" + "factory"). So, `thunkify(..)` produces a thunkory, and a thunkory produces thunks. That reasoning is symmetric to my proposal for "promisory" in Chapter 3:

```
var fooThunkory = thunkify( foo );

var fooThunk1 = fooThunkory( 3, 4 );
var fooThunk2 = fooThunkory( 5, 6 );

// later

fooThunk1( function(sum) {
  console.log( sum );           // 7
} );

fooThunk2( function(sum) {
  console.log( sum );           // 11
} );
```

**Note:** The running `foo(..)` example expects a style of callback that's not "error-first style." Of course, "error-first style" is much more common. If `foo(..)` had some sort of legitimate error-producing expectation, we could change it to expect and use an error-first callback. None of the subsequent `thunkify(..)` machinery cares what style of callback is assumed. The only difference in usage would be `fooThunk1(function(err,sum){...})`.

Exposing the thunkory method -- instead of how the earlier `thunkify(..)` hides this intermediary step -- may seem like unnecessary complication. But in general, it's quite useful to make thunkories at the beginning of your program to wrap existing API methods, and then be able to pass around and call those thunkories when you need thunks. The two distinct steps preserve a cleaner separation of capability.

To illustrate:

```
// cleaner:
var fooThunkory = thunkify( foo );

var fooThunk1 = fooThunkory( 3, 4 );
var fooThunk2 = fooThunkory( 5, 6 );

// instead of:
var fooThunk1 = thunkify( foo, 3, 4 );
var fooThunk2 = thunkify( foo, 5, 6 );
```

Regardless of whether you like to deal with the thunkories explicitly or not, the usage of thunks `fooThunk1(..)` and `fooThunk2(..)` remains the same.

## s/promise/thunk/

So what's all this thunk stuff have to do with generators?

Comparing thunks to promises generally: they're not directly interchangeable as they're not equivalent in behavior. Promises are vastly more capable and trustable than bare thunks.

But in another sense, they both can be seen as a request for a value, which may be `async` in its answering.

Recall from Chapter 3 we defined a utility for promisifying a function, which we called `Promise.wrap(..)` -- we could have called it `promisify(..)`, too! This Promise-wrapping utility doesn't produce Promises; it produces promisories that in turn produce Promises. This is completely symmetric to the thunkories and thunks presently being discussed.

To illustrate the symmetry, let's first alter the running `foo(..)` example from earlier to assume an "error-first style" callback:

```
function foo(x,y,cb) {
  setTimeout( function(){
    // assume `cb(..)` as "error-first style"
    cb( null, x + y );
  }, 1000 );
}
```

Now, we'll compare using `thunkify(..)` and `promisify(..)` (aka `Promise.wrap(..)` from Chapter 3):

```
// symmetrical: constructing the question asker
var fooThunkory = thunkify( foo );
var fooPromisory = promisify( foo );

// symmetrical: asking the question
var fooThunk = fooThunkory( 3, 4 );
var fooPromise = fooPromisory( 3, 4 );

// get the thunk answer
fooThunk( function(err,sum){
  if (err) {
    console.error( err );
  }
  else {
    console.log( sum );           // 7
  }
} );

// get the promise answer
fooPromise
.then(
  function(sum){
    console.log( sum );           // 7
  },
  function(err){
    console.error( err );
  }
);
```

Both the thunkory and the promisory are essentially asking a question (for a value), and respectively the thunk `fooThunk` and promise `fooPromise` represent the future answers to that question. Presented in that light, the symmetry is clear.

With that perspective in mind, we can see that generators which `yield` Promises for asynchrony could instead `yield` thunks for asynchrony. All we'd need is a smarter `run(..)` utility (like from before) that can not only look for and wire up to a `yield` ed Promise but also to provide a callback to a `yield` ed thunk.

Consider:

```
function *foo() {
  var val = yield request( "http://some.url.1" );
  console.log( val );
}

run( foo );
```

In this example, `request(..)` could either be a promisory that returns a promise, or a thunkory that returns a thunk. From the perspective of what's going on inside the generator code logic, we don't care about that implementation detail, which is quite powerful!

So, `request(..)` could be either:

```
// promisory `request(..)` (see Chapter 3)
var request = Promise.wrap( ajax );

// vs.

// thunkory `request(..)`
var request = thunkify( ajax );
```

Finally, as a thunk-aware patch to our earlier `run(..)` utility, we would need logic like this:

```
// ..
// did we receive a thunk back?
else if (typeof next.value == "function") {
    return new Promise( function(resolve,reject){
        // call the thunk with an error-first callback
        next.value( function(err,msg) {
            if (err) {
                reject( err );
            }
            else {
                resolve( msg );
            }
        } );
    } )
    .then(
        handleNext,
        function handleErr(err) {
            return Promise.resolve(
                it.throw( err )
            )
            .then( handleResult );
        }
    );
}
```

Now, our generators can either call promisories to `yield` Promises, or call thunkories to `yield` thunks, and in either case, `run(..)` would handle that value and use it to wait for the completion to resume the generator.

Symmetry wise, these two approaches look identical. However, we should point out that's true only from the perspective of Promises or thunks representing the future value continuation of a generator.

From the larger perspective, thunks do not in and of themselves have hardly any of the trustability or composability guarantees that Promises are designed with. Using a thunk as a stand-in for a Promise in this particular generator asynchrony pattern is workable but should be seen as less than ideal when compared to all the benefits that Promises offer (see Chapter 3).

If you have the option, prefer `yield pr` rather than `yield th`. But there's nothing wrong with having a `run(..)` utility which can handle both value types.

**Note:** The `runner(..)` utility in my *asynquence* library, which will be discussed in Appendix A, handles `yield`s of Promises, thunks and *asynquence* sequences.

## Pre-ES6 Generators

You're hopefully convinced now that generators are a very important addition to the async programming toolbox. But it's a new syntax in ES6, which means you can't just polyfill generators like you can Promises (which are just a new API). So what can we do to bring generators to our browser JS if we don't have the luxury of ignoring pre-ES6 browsers?

For all new syntax extensions in ES6, there are tools -- the most common term for them is transpilers, for trans-compilers -- which can take your ES6 syntax and transform it into equivalent (but obviously uglier!) pre-ES6 code. So, generators can be transpiled into code that will have the same behavior but work in ES5 and below.

But how? The "magic" of `yield` doesn't obviously sound like code that's easy to transpile. We actually hinted at a solution in our earlier discussion of closure-based *iterators*.

## Manual Transformation

Before we discuss the transpilers, let's derive how manual transpilation would work in the case of generators. This isn't just an academic exercise, because doing so will actually help further reinforce how they work.

Consider:

```
// `request(..)` is a Promise-aware Ajax utility

function *foo(url) {
  try {
    console.log("requesting:", url);
    var val = yield request(url);
    console.log(val);
  }
  catch (err) {
    console.log("Oops:", err);
    return false;
  }
}

var it = foo("http://some.url.1");
```

The first thing to observe is that we'll still need a normal `foo()` function that can be called, and it will still need to return an *iterator*. So, let's sketch out the non-generator transformation:

```
function foo(url) {
  // ...

  // make and return an iterator
  return {
    next: function(v) {
      // ...
    },
    throw: function(e) {
      // ...
    }
  };
}

var it = foo("http://some.url.1");
```

The next thing to observe is that a generator does its "magic" by suspending its scope/state, but we can emulate that with function closure (see the *Scope & Closures* title of this series). To understand how to write such code, we'll first annotate different parts of our generator with state values:

```
// `request(..)` is a Promise-aware Ajax utility
```

```

function *foo(url) {
  // STATE *1*

  try {
    console.log("requesting:", url);
    var TMP1 = request(url);

    // STATE *2*
    var val = yield TMP1;
    console.log(val);
  }
  catch (err) {
    // STATE *3*
    console.log("Oops:", err);
    return false;
  }
}

```

**Note:** For more accurate illustration, we split up the `val = yield request..` statement into two parts, using the temporary `TMP1` variable. `request(..)` happens in state `*1*`, and the assignment of its completion value to `val` happens in state `*2*`. We'll get rid of that intermediate `TMP1` when we convert the code to its non-generator equivalent.

In other words, `*1*` is the beginning state, `*2*` is the state if the `request(..)` succeeds, and `*3*` is the state if the `request(..)` fails. You can probably imagine how any extra `yield` steps would just be encoded as extra states.

Back to our transpiled generator, let's define a variable `state` in the closure we can use to keep track of the state:

```

function foo(url) {
  // manage generator state
  var state;

  //
}

```

Now, let's define an inner function called `process(..)` inside the closure which handles each state, using a `switch` statement:

```

// `request(..)` is a Promise-aware Ajax utility

function foo(url) {
  // manage generator state
  var state;

  // generator-wide variable declarations
  var val;

  function process(v) {
    switch (state) {
      case 1:
        console.log("requesting:", url);
        return request(url);
      case 2:
        val = v;
        console.log(val);
        return;
      case 3:
        var err = v;
        console.log("Oops:", err);
        return false;
    }
  }

  //
}

```

Each state in our generator is represented by its own `case` in the `switch` statement. `process(..)` will be called each time we need to process a new state. We'll come back to how that works in just a moment.

For any generator-wide variable declarations (`val`), we move those to a `var` declaration outside of `process(..)` so they can survive multiple calls to `process(..)`. But the "block scoped" `err` variable is only needed for the `*3*` state, so we leave it in place.

In state `*1*`, instead of `yield request(..)`, we did `return request(..)`. In terminal state `*2*`, there was no explicit `return`, so we just do a `return;` which is the same as `return undefined`. In terminal state `*3*`, there was a `return false`, so we preserve that.

Now we need to define the code in the *iterator* functions so they call `process(..)` appropriately:

```
function foo(url) {
  // manage generator state
  var state;

  // generator-wide variable declarations
  var val;

  function process(v) {
    switch (state) {
      case 1:
        console.log("requesting:", url);
        return request(url);
      case 2:
        val = v;
        console.log(val);
        return;
      case 3:
        var err = v;
        console.log("Oops:", err);
        return false;
    }
  }

  // make and return an iterator
  return {
    next: function(v) {
      // initial state
      if (!state) {
        state = 1;
        return {
          done: false,
          value: process()
        };
      }
      // yield resumed successfully
      else if (state == 1) {
        state = 2;
        return {
          done: true,
          value: process(v)
        };
      }
      // generator already completed
      else {
        return {
          done: true,
          value: undefined
        };
      }
    },
    "throw": function(e) {
      // the only explicit error handling is in
      // state *1*
    }
  };
}
```

```

        if (state == 1) {
            state = 3;
            return {
                done: true,
                value: process( e )
            };
        }
        // otherwise, an error won't be handled,
        // so just throw it right back out
        else {
            throw e;
        }
    }
};

}

```

How does this code work?

1. The first call to the `iterator's next()` call would move the generator from the uninitialized state to state `1`, and then call `process()` to handle that state. The return value from `request(..)`, which is the promise for the Ajax response, is returned back as the `value` property from the `next()` call.
2. If the Ajax request succeeds, the second call to `next(..)` should send in the Ajax response value, which moves our state to `2`. `process(..)` is again called (this time with the passed in Ajax response value), and the `value` property returned from `next(..)` will be `undefined`.
3. However, if the Ajax request fails, `throw(..)` should be called with the error, which would move the state from `1` to `3` (instead of `2`). Again `process(..)` is called, this time with the error value. That `case` returns `false`, which is set as the `value` property returned from the `throw(..)` call.

From the outside -- that is, interacting only with the `iterator` -- this `foo(..)` normal function works pretty much the same as the `*foo(..)` generator would have worked. So we've effectively "transpiled" our ES6 generator to pre-ES6 compatibility!

We could then manually instantiate our generator and control its iterator -- calling `var it = foo(..)` and `it.next(..)` and such -- or better, we could pass it to our previously defined `run(..)` utility as `run(foo,..)`.

## Automatic Transpilation

The preceding exercise of manually deriving a transformation of our ES6 generator to pre-ES6 equivalent teaches us how generators work conceptually. But that transformation was really intricate and very non-portable to other generators in our code. It would be quite impractical to do this work by hand, and would completely obviate all the benefit of generators.

But luckily, several tools already exist that can automatically convert ES6 generators to things like what we derived in the previous section. Not only do they do the heavy lifting work for us, but they also handle several complications that we glossed over.

One such tool is regenerator (<https://facebook.github.io/regenerator/>), from the smart folks at Facebook.

If we use regenerator to transpile our previous generator, here's the code produced (at the time of this writing):

```

// `request(..)` is a Promise-aware Ajax utility

var foo = regeneratorRuntime.mark(function foo(url) {
    var val;

    return regeneratorRuntime.wrap(function foo$($context$1$0) {
        while (1) switch ($context$1$0.prev = $context$1$0.next) {
            case 0:
                $context$1$0.prev = 0;
                console.log( "requesting:", url );
                $context$1$0.next = 4;
                return request( url );
            case 4:

```

```

    val = context$1$0.sent;
    console.log( val );
    context$1$0.next = 12;
    break;
  case 8:
    context$1$0.prev = 8;
    context$1$0.t0 = context$1$0.catch(0);
    console.log("Oops:", context$1$0.t0);
    return context$1$0.abrupt("return", false);
  case 12:
  case "end":
    return context$1$0.stop();
}
}, foo, this, [[0, 8]]);
});

```

There's some obvious similarities here to our manual derivation, such as the `switch` / `case` statements, and we even see `val` pulled out of the closure just as we did.

Of course, one trade-off is that regenerator's transpilation requires a helper library `regeneratorRuntime` that holds all the reusable logic for managing a general generator / *iterator*. A lot of that boilerplate looks different than our version, but even then, the concepts can be seen, like with `context$1$0.next = 4` keeping track of the next state for the generator.

The main takeaway is that generators are not restricted to only being useful in ES6+ environments. Once you understand the concepts, you can employ them throughout your code, and use tools to transform the code to be compatible with older environments.

This is more work than just using a `Promise` API polyfill for pre-ES6 Promises, but the effort is totally worth it, because generators are so much better at expressing async flow control in a reasonable, sensible, synchronous-looking, sequential fashion.

Once you get hooked on generators, you'll never want to go back to the hell of async spaghetti callbacks!

## Review

Generators are a new ES6 function type that does not run-to-completion like normal functions. Instead, the generator can be paused in mid-completion (entirely preserving its state), and it can later be resumed from where it left off.

This pause/resume interchange is cooperative rather than preemptive, which means that the generator has the sole capability to pause itself, using the `yield` keyword, and yet the *iterator* that controls the generator has the sole capability (via `next(..)`) to resume the generator.

The `yield` / `next(..)` duality is not just a control mechanism, it's actually a two-way message passing mechanism. A `yield ..` expression essentially pauses waiting for a value, and the next `next(..)` call passes a value (or implicit `undefined`) back to that paused `yield` expression.

The key benefit of generators related to async flow control is that the code inside a generator expresses a sequence of steps for the task in a naturally sync/sequential fashion. The trick is that we essentially hide potential asynchrony behind the `yield` keyword -- moving the asynchrony to the code where the generator's *iterator* is controlled.

In other words, generators preserve a sequential, synchronous, blocking code pattern for async code, which lets our brains reason about the code much more naturally, addressing one of the two key drawbacks of callback-based async.

# You Don't Know JS Yet: Sync & Async - 2nd Edition

## Chapter 5: Program Performance

**NOTE:**

Work in progress

**NOTE:**

Everything below here is previous text from 1st edition, and is only here for reference while 2nd edition work is underway.  
**Please ignore this stuff.**

This book so far has been all about how to leverage asynchrony patterns more effectively. But we haven't directly addressed why asynchrony really matters to JS. The most obvious explicit reason is **performance**.

For example, if you have two Ajax requests to make, and they're independent, but you need to wait on them both to finish before doing the next task, you have two options for modeling that interaction: serial and concurrent.

You could make the first request and wait to start the second request until the first finishes. Or, as we've seen both with promises and generators, you could make both requests "in parallel," and express the "gate" to wait on both of them before moving on.

Clearly, the latter is usually going to be more performant than the former. And better performance generally leads to better user experience.

It's even possible that asynchrony (interleaved concurrency) can improve just the perception of performance, even if the overall program still takes the same amount of time to complete. User perception of performance is every bit -- if not more! -- as important as actual measurable performance.

We want to now move beyond localized asynchrony patterns to talk about some bigger picture performance details at the program level.

**Note:** You may be wondering about micro-performance issues like if `a++` or `++a` is faster. We'll look at those sorts of performance details in the next chapter on "Benchmarking & Tuning."

## Web Workers

If you have processing-intensive tasks but you don't want them to run on the main thread (which may slow down the browser/UI), you might have wished that JavaScript could operate in a multithreaded manner.

In Chapter 1, we talked in detail about how JavaScript is single threaded. And that's still true. But a single thread isn't the only way to organize the execution of your program.

Imagine splitting your program into two pieces, and running one of those pieces on the main UI thread, and running the other piece on an entirely separate thread.

What kinds of concerns would such an architecture bring up?

For one, you'd want to know if running on a separate thread meant that it ran in parallel (on systems with multiple CPUs/cores) such that a long-running process on that second thread would **not** block the main program thread. Otherwise, "virtual threading" wouldn't be of much benefit over what we already have in JS with async concurrency.

And you'd want to know if these two pieces of the program have access to the same shared scope/resources. If they do, then you have all the questions that multithreaded languages (Java, C++, etc.) deal with, such as needing cooperative or preemptive locking (mutexes, etc.). That's a lot of extra work, and shouldn't be undertaken lightly.

Alternatively, you'd want to know how these two pieces could "communicate" if they couldn't share scope/resources.

All these are great questions to consider as we explore a feature added to the web platform circa HTML5 called "Web Workers." This is a feature of the browser (aka host environment) and actually has almost nothing to do with the JS language itself. That is, JavaScript does not *currently* have any features that support threaded execution.

But an environment like your browser can easily provide multiple instances of the JavaScript engine, each on its own thread, and let you run a different program in each thread. Each of those separate threaded pieces of your program is called a "(Web) Worker." This type of parallelism is called "task parallelism," as the emphasis is on splitting up chunks of your program to run in parallel.

From your main JS program (or another Worker), you instantiate a Worker like so:

```
var w1 = new Worker( "http://some.url.1/mycoolworker.js" );
```

The URL should point to the location of a JS file (not an HTML page!) which is intended to be loaded into a Worker. The browser will then spin up a separate thread and let that file run as an independent program in that thread.

**Note:** The kind of Worker created with such a URL is called a "Dedicated Worker." But instead of providing a URL to an external file, you can also create an "Inline Worker" by providing a Blob URL (another HTML5 feature); essentially it's an inline file stored in a single (binary) value. However, Blobs are beyond the scope of what we'll discuss here.

Workers do not share any scope or resources with each other or the main program -- that would bring all the nightmares of threaded programming to the forefront -- but instead have a basic event messaging mechanism connecting them.

The `w1` Worker object is an event listener and trigger, which lets you subscribe to events sent by the Worker as well as send events to the Worker.

Here's how to listen for events (actually, the fixed `"message"` event):

```
w1.addEventListener( "message", function(evt){
    // evt.data
} );
```

And you can send the `"message"` event to the Worker:

```
w1.postMessage( "something cool to say" );
```

Inside the Worker, the messaging is totally symmetrical:

```
// "mycoolworker.js"
```

```

addEventListener( "message", function(evt){
    // evt.data
} );

postMessage( "a really cool reply" );

```

Notice that a dedicated Worker is in a one-to-one relationship with the program that created it. That is, the `"message"` event doesn't need any disambiguation here, because we're sure that it could only have come from this one-to-one relationship -- either it came from the Worker or the main page.

Usually the main page application creates the Workers, but a Worker can instantiate its own child Worker(s) -- known as subworkers -- as necessary. Sometimes this is useful to delegate such details to a sort of "master" Worker that spawns other Workers to process parts of a task. Unfortunately, at the time of this writing, Chrome still does not support subworkers, while Firefox does.

To kill a Worker immediately from the program that created it, call `terminate()` on the Worker object (like `w1` in the previous snippets). Abruptly terminating a Worker thread does not give it any chance to finish up its work or clean up any resources. It's akin to you closing a browser tab to kill a page.

If you have two or more pages (or multiple tabs with the same page!) in the browser that try to create a Worker from the same file URL, those will actually end up as completely separate Workers. Shortly, we'll discuss a way to "share" a Worker.

**Note:** It may seem like a malicious or ignorant JS program could easily perform a denial-of-service attack on a system by spawning hundreds of Workers, seemingly each with their own thread. While it's true that it's somewhat of a guarantee that a Worker will end up on a separate thread, this guarantee is not unlimited. The system is free to decide how many actual threads/CPUs/cores it really wants to create. There's no way to predict or guarantee how many you'll have access to, though many people assume it's at least as many as the number of CPUs/cores available. I think the safest assumption is that there's at least one other thread besides the main UI thread, but that's about it.

## Worker Environment

Inside the Worker, you do not have access to any of the main program's resources. That means you cannot access any of its global variables, nor can you access the page's DOM or other resources. Remember: it's a totally separate thread.

You can, however, perform network operations (Ajax, WebSockets) and set timers. Also, the Worker has access to its own copy of several important global variables/features, including `navigator`, `location`, `JSON`, and `applicationCache`.

You can also load extra JS scripts into your Worker, using `importScripts(..)`:

```

// inside the Worker
importScripts( "foo.js", "bar.js" );

```

These scripts are loaded synchronously, which means the `importScripts(..)` call will block the rest of the Worker's execution until the file(s) are finished loading and executing.

**Note:** There have also been some discussions about exposing the `<canvas>` API to Workers, which combined with having canvases be Transferables (see the "Data Transfer" section), would allow Workers to perform more sophisticated off-thread graphics processing, which can be useful for high-performance gaming (WebGL) and other similar applications. Although this doesn't exist yet in any browsers, it's likely to happen in the near future.

What are some common uses for Web Workers?

- Processing intensive math calculations
- Sorting large data sets
- Data operations (compression, audio analysis, image pixel manipulations, etc.)
- High-traffic network communications

## Data Transfer

You may notice a common characteristic of most of those uses, which is that they require a large amount of information to be transferred across the barrier between threads using the event mechanism, perhaps in both directions.

In the early days of Workers, serializing all data to a string value was the only option. In addition to the speed penalty of the two-way serializations, the other major negative was that the data was being copied, which meant a doubling of memory usage (and the subsequent churn of garbage collection).

Thankfully, we now have a few better options.

If you pass an object, a so-called "Structured Cloning Algorithm" ([https://developer.mozilla.org/en-US/docs/Web/Guide/API/DOM/The\\_structured\\_clone\\_algorithm](https://developer.mozilla.org/en-US/docs/Web/Guide/API/DOM/The_structured_clone_algorithm)) is used to copy/duplicate the object on the other side. This algorithm is fairly sophisticated and can even handle duplicating objects with circular references. The to-string/from-string performance penalty is not paid, but we still have duplication of memory using this approach. There is support for this in IE10 and above, as well as all the other major browsers.

An even better option, especially for larger data sets, is "Transferable Objects" (<http://updates.html5rocks.com/2011/12/Transferable-Objects-Lightning-Fast>). What happens is that the object's "ownership" is transferred, but the data itself is not moved. Once you transfer away an object to a Worker, it's empty or inaccessible in the originating location -- that eliminates the hazards of threaded programming over a shared scope. Of course, transfer of ownership can go in both directions.

There really isn't much you need to do to opt into a Transferable Object; any data structure that implements the Transferable interface (<https://developer.mozilla.org/en-US/docs/Web/API/Transferable>) will automatically be transferred this way (support Firefox & Chrome).

For example, typed arrays like `Uint8Array` (see the *ES6 & Beyond* title of this series) are "Transferables." This is how you'd send a Transferable Object using `postMessage(...)`:

```
// `foo` is a `Uint8Array` for instance
postMessage( foo.buffer, [ foo.buffer ] );
```

The first parameter is the raw buffer and the second parameter is a list of what to transfer.

Browsers that don't support Transferable Objects simply degrade to structured cloning, which means performance reduction rather than outright feature breakage.

## Shared Workers

If your site or app allows for loading multiple tabs of the same page (a common feature), you may very well want to reduce the resource usage of their system by preventing duplicate dedicated Workers; the most common limited resource in this respect is a socket network connection, as browsers limit the number of simultaneous connections to a single host. Of course, limiting multiple connections from a client also eases your server resource requirements.

In this case, creating a single centralized Worker that all the page instances of your site or app can *share* is quite useful.

That's called a `SharedWorker`, which you create like so (support for this is limited to Firefox and Chrome):

```
var w1 = new SharedWorker( "http://some.url.1/mycoolworker.js" );
```

Because a shared Worker can be connected to or from more than one program instance or page on your site, the Worker needs a way to know which program a message comes from. This unique identification is called a "port" -- think network socket ports. So the calling program must use the `port` object of the Worker for communication:

```
w1.port.addEventListener( "message", handleMessages );

// ...

w1.port.postMessage( "something cool" );
```

Also, the port connection must be initialized, as:

```
w1.port.start();
```

Inside the shared Worker, an extra event must be handled: `"connect"`. This event provides the `port` object for that particular connection. The most convenient way to keep multiple connections separate is to use closure (see *Scope & Closures* title of this series) over the `port`, as shown next, with the event listening and transmitting for that connection defined inside the handler for the `"connect"` event:

```
// inside the shared Worker
addEventListener( "connect", function(evt){
    // the assigned port for this connection
    var port = evt.ports[0];

    port.addEventListener( "message", function(evt){
        // ...
        port.postMessage( ... );
        // ...
    } );
    // initialize the port connection
    port.start();
});
```

Other than that difference, shared and dedicated Workers have the same capabilities and semantics.

**Note:** Shared Workers survive the termination of a port connection if other port connections are still alive, whereas dedicated Workers are terminated whenever the connection to their initiating program is terminated.

## Polyfilling Web Workers

Web Workers are very attractive performance-wise for running JS programs in parallel. However, you may be in a position where your code needs to run in older browsers that lack support. Because Workers are an API and not a syntax, they can be polyfilled, to an extent.

If a browser doesn't support Workers, there's simply no way to fake multithreading from the performance perspective. Iframes are commonly thought of to provide a parallel environment, but in all modern browsers they actually run on the same thread as the main page, so they're not sufficient for faking parallelism.

As we detailed in Chapter 1, JS's asynchronicity (not parallelism) comes from the event loop queue, so you can force faked Workers to be asynchronous using timers (`setTimeout(..)`, etc.). Then you just need to provide a polyfill for the Worker API. There are some listed here (<https://github.com/Modernizr/Modernizr/wiki/HTML5-Cross-Browser-Polyfills#web-workers>), but frankly none of them look great.

I've written a sketch of a polyfill for `worker` here (<https://gist.github.com/getify/1b26accc1a09aa53ad25>). It's basic, but it should get the job done for simple `worker` support, given that the two-way messaging works correctly as well as `"onerror"` handling. You could probably also extend it with more features, such as `terminate()` or faked Shared Workers, as you see fit.

**Note:** You can't fake synchronous blocking, so this polyfill just disallows use of `importScripts(...)`. Another option might have been to parse and transform the Worker's code (once Ajax loaded) to handle rewriting to some asynchronous form of an `importScripts(...)` polyfill, perhaps with a promise-aware interface.

## SIMD

Single instruction, multiple data (SIMD) is a form of "data parallelism," as contrasted to "task parallelism" with Web Workers, because the emphasis is not really on program logic chunks being parallelized, but rather multiple bits of data being processed in parallel.

With SIMD, threads don't provide the parallelism. Instead, modern CPUs provide SIMD capability with "vectors" of numbers -- think: type specialized arrays -- as well as instructions that can operate in parallel across all the numbers; these are low-level operations leveraging instruction-level parallelism.

The effort to expose SIMD capability to JavaScript is primarily spearheaded by Intel (<https://01.org/node/1495>), namely by Mohammad Haghigat (at the time of this writing), in cooperation with Firefox and Chrome teams. SIMD is on an early standards track with a good chance of making it into a future revision of JavaScript, likely in the ES7 timeframe.

SIMD JavaScript proposes to expose short vector types and APIs to JS code, which on those SIMD-enabled systems would map the operations directly through to the CPU equivalents, with fallback to non-parallelized operation "shims" on non-SIMD systems.

The performance benefits for data-intensive applications (signal analysis, matrix operations on graphics, etc.) with such parallel math processing are quite obvious!

Early proposal forms of the SIMD API at the time of this writing look like this:

```
var v1 = SIMD.float32x4( 3.14159, 21.0, 32.3, 55.55 );
var v2 = SIMD.float32x4( 2.1, 3.2, 4.3, 5.4 );

var v3 = SIMD.int32x4( 10, 101, 1001, 10001 );
var v4 = SIMD.int32x4( 10, 20, 30, 40 );

SIMD.float32x4.mul( v1, v2 );      // [ 6.597339, 67.2, 138.89, 299.97 ]
SIMD.int32x4.add( v3, v4 );       // [ 20, 121, 1031, 10041 ]
```

Shown here are two different vector data types, 32-bit floating-point numbers and 32-bit integer numbers. You can see that these vectors are sized exactly to four 32-bit elements, as this matches the SIMD vector sizes (128-bit) available in most modern CPUs. It's also possible we may see an `x8` (or larger!) version of these APIs in the future.

Besides `mul()` and `add()`, many other operations are likely to be included, such as `sub()`, `div()`, `abs()`, `neg()`, `sqrt()`, `reciprocal()`, `reciprocalSqrt()` (arithmetic), `shuffle()` (rearrange vector elements), `and()`, `or()`, `xor()`, `not()` (logical), `equal()`, `greaterThan()`, `lessThan()` (comparison), `shiftLeft()`, `shiftRightLogical()`, `shiftRightArithmetic()` (shifts), `fromFloat32x4()`, and `fromInt32x4()` (conversions).

**Note:** There's an official "prolyfill" (hopeful, expectant, future-leaning polyfill) for the SIMD functionality available ([https://github.com/johnmcutchan/ecmascript\\_simd](https://github.com/johnmcutchan/ecmascript_simd)), which illustrates a lot more of the planned SIMD capability than we've illustrated in this section.

## asm.js

"asm.js" (<http://asmjs.org/>) is a label for a highly optimizable subset of the JavaScript language. By carefully avoiding certain mechanisms and patterns that are *hard* to optimize (garbage collection, coercion, etc.), asm.js-styled code can be recognized by the JS engine and given special attention with aggressive low-level optimizations.

Distinct from other program performance mechanisms discussed in this chapter, asm.js isn't necessarily something that needs to be adopted into the JS language specification. There is an asm.js specification (<http://asmjs.org/spec/latest/>), but it's mostly for tracking an agreed upon set of candidate inferences for optimization rather than a set of requirements of JS engines.

There's not currently any new syntax being proposed. Instead, asm.js suggests ways to recognize existing standard JS syntax that conforms to the rules of asm.js and let engines implement their own optimizations accordingly.

There's been some disagreement between browser vendors over exactly how asm.js should be activated in a program. Early versions of the asm.js experiment required a `"use asm";` pragma (similar to strict mode's `"use strict";`) to help clue the JS engine to be looking for asm.js optimization opportunities and hints. Others have asserted that asm.js should just be a set of heuristics that engines automatically recognize without the author having to do anything extra, meaning that existing programs could theoretically benefit from asm.js-style optimizations without doing anything special.

## How to Optimize with asm.js

The first thing to understand about asm.js optimizations is around types and coercion (see the *Types & Grammar* title of this series). If the JS engine has to track multiple different types of values in a variable through various operations, so that it can handle coercions between types as necessary, that's a lot of extra work that keeps the program optimization suboptimal.

**Note:** We're going to use asm.js-style code here for illustration purposes, but be aware that it's not commonly expected that you'll author such code by hand. asm.js is more intended to a compilation target from other tools, such as Emscripten (<https://github.com/kripken/emscripten/wiki>). It's of course possible to write your own asm.js code, but that's usually a bad idea because the code is very low level and managing it can be very time consuming and error prone. Nevertheless, there may be cases where you'd want to hand tweak your code for asm.js optimization purposes.

There are some "tricks" you can use to hint to an asm.js-aware JS engine what the intended type is for variables/operations, so that it can skip these coercion tracking steps.

For example:

```
var a = 42;
// ...
var b = a;
```

In that program, the `b = a` assignment leaves the door open for type divergence in variables. However, it could instead be written as:

```
var a = 42;
// ...
var b = a | 0;
```

Here, we've used the `|` ("binary OR") with value `0`, which has no effect on the value other than to make sure it's a 32-bit integer. That code run in a normal JS engine works just fine, but when run in an asm.js-aware JS engine it *can* signal that `b` should always be treated as a 32-bit integer, so the coercion tracking can be skipped.

Similarly, the addition operation between two variables can be restricted to a more performant integer addition (instead of floating point):

```
(a + b) | 0
```

Again, the asm.js-aware JS engine can see that hint and infer that the `+` operation should be 32-bit integer addition because the end result of the whole expression would automatically be 32-bit integer conformed anyway.

## asm.js Modules

One of the biggest detractors to performance in JS is around memory allocation, garbage collection, and scope access. asm.js suggests one of the ways around these issues is to declare a more formalized asm.js "module" -- do not confuse these with ES6 modules; see the *ES6 & Beyond* title of this series.

For an asm.js module, you need to explicitly pass in a tightly conformed namespace -- this is referred to in the spec as `stdlib` , as it should represent standard libraries needed -- to import necessary symbols, rather than just using globals via lexical scope. In the base case, the `window` object is an acceptable `stdlib` object for asm.js module purposes, but you could and perhaps should construct an even more restricted one.

You also must declare a "heap" -- which is just a fancy term for a reserved spot in memory where variables can already be used without asking for more memory or releasing previously used memory -- and pass that in, so that the asm.js module won't need to do anything that would cause memory churn; it can just use the pre-reserved space.

A "heap" is likely a typed `ArrayBuffer` , such as:

```
var heap = new ArrayBuffer( 0x10000 ); // 64k heap
```

Using that pre-reserved 64k of binary space, an asm.js module can store and retrieve values in that buffer without any memory allocation or garbage collection penalties. For example, the `heap` buffer could be used inside the module to back an array of 64-bit float values like this:

```
var arr = new Float64Array( heap );
```

OK, so let's make a quick, silly example of an asm.js-styled module to illustrate how these pieces fit together. We'll define a `foo(..)` that takes a start (`x`) and end (`y`) integer for a range, and calculates all the inner adjacent multiplications of the values in the range, and then finally averages those values together:

```
function fooASM(stdlib,foreign,heap) {
  "use asm";

  var arr = new stdlib.Int32Array( heap );

  function foo(x,y) {
    x = x | 0;
    y = y | 0;

    var i = 0;
    var p = 0;
    var sum = 0;
    var count = ((y|0) - (x|0)) | 0;

    // calculate all the inner adjacent multiplications
    for (i = x | 0;
         (i | 0) < (y | 0);
         p = (p + 8) | 0, i = (i + 1) | 0
    ) {
      // store result
      arr[ p >> 3 ] = (i * (i + 1)) | 0;
    }

    // calculate average of all intermediate values
    for (i = 0, p = 0;
         (i | 0) < (count | 0);
         p = (p + 8) | 0, i = (i + 1) | 0
    ) {
      sum = (sum + arr[ p >> 3 ]) | 0;
    }

    return +(sum / count);
  }
}
```

```

    }

    return {
      foo: foo
    };
}

var heap = new ArrayBuffer( 0x1000 );
var foo = fooASM( window, null, heap ).foo;

foo( 10, 20 );           // 233

```

**Note:** This asm.js example is hand authored for illustration purposes, so it doesn't represent the same code that would be produced from a compilation tool targeting asm.js. But it does show the typical nature of asm.js code, especially the type hinting and use of the `heap` buffer for temporary variable storage.

The first call to `fooASM(..)` is what sets up our asm.js module with its `heap` allocation. The result is a `foo(..)` function we can call as many times as necessary. Those `foo(..)` calls should be specially optimized by an asm.js-aware JS engine.

Importantly, the preceding code is completely standard JS and would run just fine (without special optimization) in a non-asm.js engine.

Obviously, the nature of restrictions that make asm.js code so optimizable reduces the possible uses for such code significantly. asm.js won't necessarily be a general optimization set for any given JS program. Instead, it's intended to provide an optimized way of handling specialized tasks such as intensive math operations (e.g., those used in graphics processing for games).

## Review

The first four chapters of this book are based on the premise that async coding patterns give you the ability to write more performant code, which is generally a very important improvement. But async behavior only gets you so far, because it's still fundamentally bound to a single event loop thread.

So in this chapter we've covered several program-level mechanisms for improving performance even further.

Web Workers let you run a JS file (aka program) in a separate thread using async events to message between the threads. They're wonderful for offloading long-running or resource-intensive tasks to a different thread, leaving the main UI thread more responsive.

SIMD proposes to map CPU-level parallel math operations to JavaScript APIs for high-performance data-parallel operations, like number processing on large data sets.

Finally, asm.js describes a small subset of JavaScript that avoids the hard-to-optimize parts of JS (like garbage collection and coercion) and lets the JS engine recognize and run such code through aggressive optimizations. asm.js could be hand authored, but that's extremely tedious and error prone, akin to hand authoring assembly language (hence the name). Instead, the main intent is that asm.js would be a good target for cross-compilation from other highly optimized program languages -- for example, Emscripten (<https://github.com/kripken/emscripten/wiki>) transpiling C/C++ to JavaScript.

While not covered explicitly in this chapter, there are even more radical ideas under very early discussion for JavaScript, including approximations of direct threaded functionality (not just hidden behind data structure APIs). Whether that happens explicitly, or we just see more parallelism creep into JS behind the scenes, the future of more optimized program-level performance in JS looks really *promising*.

# You Don't Know JS Yet: Sync & Async - 2nd Edition

## Chapter 6: Benchmarking & Tuning

**NOTE:**

Work in progress

**NOTE:**

Everything below here is previous text from 1st edition, and is only here for reference while 2nd edition work is underway.  
**Please ignore this stuff.**

As the first four chapters of this book were all about performance as a coding pattern (asynchrony and concurrency), and Chapter 5 was about performance at the macro program architecture level, this chapter goes after the topic of performance at the micro level, focusing on single expressions/statements.

One of the most common areas of curiosity -- indeed, some developers can get quite obsessed about it -- is in analyzing and testing various options for how to write a line or chunk of code, and which one is faster.

We're going to look at some of these issues, but it's important to understand from the outset that this chapter is **not** about feeding the obsession of micro-performance tuning, like whether some given JS engine can run `++a` faster than `a++`. The more important goal of this chapter is to figure out what kinds of JS performance matter and which ones don't, *and how to tell the difference*.

But even before we get there, we need to explore how to most accurately and reliably test JS performance, because there's tons of misconceptions and myths that have flooded our collective cult knowledge base. We've got to sift through all that junk to find some clarity.

## Benchmarking

OK, time to start dispelling some misconceptions. I'd wager the vast majority of JS developers, if asked to benchmark the speed (execution time) of a certain operation, would initially go about it something like this:

```
var start = (new Date()).getTime();    // or `Date.now()`

// do some operation

var end = (new Date()).getTime();
```

```
console.log( "Duration:", (end - start) );
```

Raise your hand if that's roughly what came to your mind. Yep, I thought so. There's a lot wrong with this approach, but don't feel bad; **we've all been there**.

What did that measurement tell you, exactly? Understanding what it does and doesn't say about the execution time of the operation in question is key to learning how to appropriately benchmark performance in JavaScript.

If the duration reported is `0`, you may be tempted to believe that it took less than a millisecond. But that's not very accurate. Some platforms don't have single millisecond precision, but instead only update the timer in larger increments. For example, older versions of windows (and thus IE) had only 15ms precision, which means the operation has to take at least that long for anything other than `0` to be reported!

Moreover, whatever duration is reported, the only thing you really know is that the operation took approximately that long on that exact single run. You have near-zero confidence that it will always run at that speed. You have no idea if the engine or system had some sort of interference at that exact moment, and that at other times the operation could run faster.

What if the duration reported is `4`? Are you more sure it took about four milliseconds? Nope. It might have taken less time, and there may have been some other delay in getting either `start` or `end` timestamps.

More troublingly, you also don't know that the circumstances of this operation test aren't overly optimistic. It's possible that the JS engine figured out a way to optimize your isolated test case, but in a more real program such optimization would be diluted or impossible, such that the operation would run slower than your test.

So... what do we know? Unfortunately, with those realizations stated, **we know very little**. Something of such low confidence isn't even remotely good enough to build your determinations on. Your "benchmark" is basically useless. And worse, it's dangerous in that it implies false confidence, not just to you but also to others who don't think critically about the conditions that led to those results.

## Repetition

"OK," you now say, "Just put a loop around it so the whole test takes longer." If you repeat an operation 100 times, and that whole loop reportedly takes a total of 137ms, then you can just divide by 100 and get an average duration of 1.37ms for each operation, right?

Well, not exactly.

A straight mathematical average by itself is definitely not sufficient for making judgments about performance which you plan to extrapolate to the breadth of your entire application. With a hundred iterations, even a couple of outliers (high or low) can skew the average, and then when you apply that conclusion repeatedly, you even further inflate the skew beyond credulity.

Instead of just running for a fixed number of iterations, you can instead choose to run the loop of tests until a certain amount of time has passed. That might be more reliable, but how do you decide how long to run? You might guess that it should be some multiple of how long your operation should take to run once. Wrong.

Actually, the length of time to repeat across should be based on the accuracy of the timer you're using, specifically to minimize the chances of inaccuracy. The less precise your timer, the longer you need to run to make sure you've minimized the error percentage. A 15ms timer is pretty bad for accurate benchmarking; to minimize its uncertainty (aka "error rate") to less than 1%, you need to run your each cycle of test iterations for 750ms. A 1ms timer only needs a cycle to run for 50ms to get the same confidence.

But then, that's just a single sample. To be sure you're factoring out the skew, you'll want lots of samples to average across. You'll also want to understand something about just how slow the worst sample is, how fast the best sample is, how far apart those best and worse cases were, and so on. You'll want to know not just a number that tells you how fast something ran, but also to have some quantifiable measure of how trustable that number is.

Also, you probably want to combine these different techniques (as well as others), so that you get the best balance of all the possible approaches.

That's all bare minimum just to get started. If you've been approaching performance benchmarking with anything less serious than what I just glossed over, well... "you don't know: proper benchmarking."

## Benchmark.js

Any relevant and reliable benchmark should be based on statistically sound practices. I am not going to write a chapter on statistics here, so I'll hand wave around some terms: standard deviation, variance, margin of error. If you don't know what those terms really mean -- I took a stats class back in college and I'm still a little fuzzy on them -- you are not actually qualified to write your own benchmarking logic.

Luckily, smart folks like John-David Dalton and Mathias Bynens do understand these concepts, and wrote a statistically sound benchmarking tool called Benchmark.js (<http://benchmarkjs.com/>). So I can end the suspense by simply saying: "just use that tool."

I won't repeat their whole documentation for how Benchmark.js works; they have fantastic API Docs (<http://benchmarkjs.com/docs>) you should read. Also there are some great (<http://calendar.perfplanet.com/2010/bulletproof-javascript-benchmarks/>) writeups (<http://monsur.hossa.in/2012/12/11/benchmarkjs.html>) on more of the details and methodology.

But just for quick illustration purposes, here's how you could use Benchmark.js to run a quick performance test:

```
function foo() {
  // operation(s) to test
}

var bench = new Benchmark(
  "foo test",           // test name
  foo,                  // function to test (just contents)
  {                     // ...
    // optional extra options (see docs)
  }
);

bench.hz;              // number of operations per second
bench.stats.moe;       // margin of error
bench.stats.variance;  // variance across samples
// ...
```

There's *lots* more to learn about using Benchmark.js besides this glance I'm including here. But the point is that it's handling all of the complexities of setting up a fair, reliable, and valid performance benchmark for a given piece of JavaScript code. If you're going to try to test and benchmark your code, this library is the first place you should turn.

We're showing here the usage to test a single operation like X, but it's fairly common that you want to compare X to Y. This is easy to do by simply setting up two different tests in a "Suite" (a Benchmark.js organizational feature). Then, you run them head-to-head, and compare the statistics to conclude whether X or Y was faster.

Benchmark.js can of course be used to test JavaScript in a browser (see the "jsPerf.com" section later in this chapter), but it can also run in non-browser environments (Node.js, etc.).

One largely untapped potential use-case for Benchmark.js is to use it in your Dev or QA environments to run automated performance regression tests against critical path parts of your application's JavaScript. Similar to how you might run unit test suites before deployment, you can also compare the performance against previous benchmarks to monitor if you are improving or degrading application performance.

## Setup/Teardown

In the previous code snippet, we glossed over the "extra options" `{ ... }` object. But there are two options we should discuss: `setup` and `teardown`.

These two options let you define functions to be called before and after your test case runs.

It's incredibly important to understand that your `setup` and `teardown` code **does not run for each test iteration**. The best way to think about it is that there's an outer loop (repeating cycles), and an inner loop (repeating test iterations). `setup` and `teardown` are run at the beginning and end of each *outer* loop (aka cycle) iteration, but not inside the inner loop.

Why does this matter? Let's imagine you have a test case that looks like this:

```
a = a + "w";
b = a.charAt( 1 );
```

Then, you set up your test `setup` as follows:

```
var a = "x";
```

Your temptation is probably to believe that `a` is starting out as `"x"` for each test iteration.

But it's not! It's starting `a` at `"x"` for each test cycle, and then your repeated `+ "w"` concatenations will be making a larger and larger `a` value, even though you're only ever accessing the character `"w"` at the `1` position.

Where this most commonly bites you is when you make side effect changes to something like the DOM, like appending a child element. You may think your parent element is set as empty each time, but it's actually getting lots of elements added, and that can significantly sway the results of your tests.

## Context Is King

Don't forget to check the context of a particular performance benchmark, especially a comparison between X and Y tasks. Just because your test reveals that X is faster than Y doesn't mean that the conclusion "X is faster than Y" is actually relevant.

For example, let's say a performance test reveals that X runs 10,000,000 operations per second, and Y runs at 8,000,000 operations per second. You could claim that Y is 20% slower than X, and you'd be mathematically correct, but your assertion doesn't hold as much water as you'd think.

Let's think about the results more critically: 10,000,000 operations per second is 10,000 operations per millisecond, and 10 operations per microsecond. In other words, a single operation takes 0.1 microseconds, or 100 nanoseconds. It's hard to fathom just how small 100ns is, but for comparison, it's often cited that the human eye isn't generally capable of distinguishing anything less than 100ms, which is one million times slower than the 100ns speed of the X operation.

Even recent scientific studies showing that maybe the brain can process as quick as 13ms (about 8x faster than previously asserted) would mean that X is still running 125,000 times faster than the human brain can perceive a distinct thing happening. **X is going really, really fast.**

But more importantly, let's talk about the difference between X and Y, the 2,000,000 operations per second difference. If X takes 100ns, and Y takes 80ns, the difference is 20ns, which in the best case is still one 650-thousandth of the interval the human brain can perceive.

What's my point? **None of this performance difference matters, at all!**

But wait, what if this operation is going to happen a whole bunch of times in a row? Then the difference could add up, right?

OK, so what we're asking then is, how likely is it that operation X is going to be run over and over again, one right after the other, and that this has to happen 650,000 times just to get a sliver of a hope the human brain could perceive it. More likely, it'd have to happen 5,000,000 to 10,000,000 times together in a tight loop to even approach relevance.

While the computer scientist in you might protest that this is possible, the louder voice of realism in you should sanity check just how likely or unlikely that really is. Even if it is relevant in rare occasions, it's irrelevant in most situations.

The vast majority of your benchmark results on tiny operations -- like the `++x` vs `x++` myth -- **are just totally bogus** for supporting the conclusion that X should be favored over Y on a performance basis.

## Engine Optimizations

You simply cannot reliably extrapolate that if X was 10 microseconds faster than Y in your isolated test, that means X is always faster than Y and should always be used. That's not how performance works. It's vastly more complicated.

For example, let's imagine (purely hypothetical) that you test some microperformance behavior such as comparing:

```
var twelve = "12";
var foo = "foo";

// test 1
var X1 = parseInt( twelve );
var X2 = parseInt( foo );

// test 2
var Y1 = Number( twelve );
var Y2 = Number( foo );
```

If you understand what `parseInt(..)` does compared to `Number(..)`, you might intuit that `parseInt(..)` potentially has "more work" to do, especially in the `foo` case. Or you might intuit that they should have the same amount of work to do in the `foo` case, as both should be able to stop at the first character `"f"`.

Which intuition is correct? I honestly don't know. But I'll make the case it doesn't matter what your intuition is. What might the results be when you test it? Again, I'm making up a pure hypothetical here, I haven't actually tried, nor do I care.

Let's pretend the test comes back that `x` and `y` are statistically identical. Have you then confirmed your intuition about the `"f"` character thing? Nope.

It's possible in our hypothetical that the engine might recognize that the variables `twelve` and `foo` are only being used in one place in each test, and so it might decide to inline those values. Then it may realize that `Number("12")` can just be replaced by `12`. And maybe it comes to the same conclusion with `parseInt(..)`, or maybe not.

Or an engine's dead-code removal heuristic could kick in, and it could realize that variables `x` and `y` aren't being used, so declaring them is irrelevant, so it doesn't end up doing anything at all in either test.

And all that's just made with the mindset of assumptions about a single test run. Modern engines are fantastically more complicated than what we're intuiting here. They do all sorts of tricks, like tracing and tracking how a piece of code behaves over a short period of time, or with a particularly constrained set of inputs.

What if the engine optimizes a certain way because of the fixed input, but in your real program you give more varied input and the optimization decisions shake out differently (or not at all!)? Or what if the engine kicks in optimizations because it sees the code being run tens of thousands of times by the benchmarking utility, but in your real program it will only run a hundred times in near proximity, and under those conditions the optimizations are not worth it?

And all those optimizations we just hypothesized about might happen in our constrained test but maybe the engine wouldn't do them in a more complex program (for various reasons). Or it could be reversed -- the engine might not optimize such trivial code but may be more inclined to optimize it more aggressively when the system is already more taxed by a more sophisticated program.

The point I'm trying to make is that you really don't know for sure exactly what's going on under the covers. All the guesses and hypothesis you can muster don't amount to hardly anything concrete for really making such decisions.

Does that mean you can't really do any useful testing? **Definitely not!**

What this boils down to is that testing *not real* code gives you *not real* results. In so much as is possible and practical, you should test actual real, non-trivial snippets of your code, and under as best of real conditions as you can actually hope to. Only then will the results you get have a chance to approximate reality.

Microbenchmarks like `++x` vs `x++` are so incredibly likely to be bogus, we might as well just flatly assume them as such.

## jsPerf.com

While Benchmark.js is useful for testing the performance of your code in whatever JS environment you're running, it cannot be stressed enough that you need to compile test results from lots of different environments (desktop browsers, mobile devices, etc.) if you want to have any hope of reliable test conclusions.

For example, Chrome on a high-end desktop machine is not likely to perform anywhere near the same as Chrome mobile on a smartphone. And a smartphone with a full battery charge is not likely to perform anywhere near the same as a smartphone with 2% battery life left, when the device is starting to power down the radio and processor.

If you want to make assertions like "X is faster than Y" in any reasonable sense across more than just a single environment, you're going to need to actually test as many of those real world environments as possible. Just because Chrome executes some X operation faster than Y doesn't mean that all browsers do. And of course you also probably will want to cross-reference the results of multiple browser test runs with the demographics of your users.

There's an awesome website for this purpose called jsPerf (<http://jsperf.com>). It uses the Benchmark.js library we talked about earlier to run statistically accurate and reliable tests, and makes the test on an openly available URL that you can pass around to others.

Each time a test is run, the results are collected and persisted with the test, and the cumulative test results are graphed on the page for anyone to see.

When creating a test on the site, you start out with two test cases to fill in, but you can add as many as you need. You also have the ability to set up `setup` code that is run at the beginning of each test cycle and `teardown` code run at the end of each cycle.

**Note:** A trick for doing just one test case (if you're benchmarking a single approach instead of a head-to-head) is to fill in the second test input boxes with placeholder text on first creation, then edit the test and leave the second test blank, which will delete it. You can always add more test cases later.

You can define the initial page setup (importing libraries, defining utility helper functions, declaring variables, etc.). There are also options for defining setup and teardown behavior if needed -- consult the "Setup/Teardown" section in the Benchmark.js discussion earlier.

## Sanity Check

jsPerf is a fantastic resource, but there's an awful lot of tests published that when you analyze them are quite flawed or bogus, for any of a variety of reasons as outlined so far in this chapter.

Consider:

```
// Case 1
var x = [];
for (var i=0; i<10; i++) {
  x[i] = "x";
}

// Case 2
var x = [];
for (var i=0; i<10; i++) {
  x[x.length] = "x";
}
```

```
// Case 3
var x = [];
for (var i=0; i<10; i++) {
  x.push( "x" );
}
```

Some observations to ponder about this test scenario:

- It's extremely common for devs to put their own loops into test cases, and they forget that Benchmark.js already does all the repetition you need. There's a really strong chance that the `for` loops in these cases are totally unnecessary noise.
- The declaring and initializing of `x` is included in each test case, possibly unnecessarily. Recall from earlier that if `x = []` were in the `setup` code, it wouldn't actually be run before each test iteration, but instead once at the beginning of each cycle. That means `x` would continue growing quite large, not just the size `10` implied by the `for` loops.

So is the intent to make sure the tests are constrained only to how the JS engine behaves with very small arrays (size `10`)? That *could* be the intent, but if it is, you have to consider if that's not focusing far too much on nuanced internal implementation details.

On the other hand, does the intent of the test embrace the context that the arrays will actually be growing quite large? Is the JS engines' behavior with larger arrays relevant and accurate when compared with the intended real world usage?

- Is the intent to find out how much `x.length` or `x.push(..)` add to the performance of the operation to append to the `x` array? OK, that might be a valid thing to test. But then again, `push(..)` is a function call, so of course it's going to be slower than `[..]` access. Arguably, cases 1 and 2 are fairer than case 3.

Here's another example that illustrates a common apples-to-oranges flaw:

```
// Case 1
var x = ["John", "Albert", "Sue", "Frank", "Bob"];
x.sort();

// Case 2
var x = ["John", "Albert", "Sue", "Frank", "Bob"];
x.sort( function mySort(a,b){
  if (a < b) return -1;
  if (a > b) return 1;
  return 0;
});
```

Here, the obvious intent is to find out how much slower the custom `mySort(..)` comparator is than the built-in default comparator. But by specifying the function `mySort(..)` as inline function expression, you've created an unfair/bogus test. Here, the second case is not only testing a custom user JS function, **but it's also testing creating a new function expression for each iteration.**

Would it surprise you to find out that if you run a similar test but update it to isolate only for creating an inline function expression versus using a pre-declared function, the inline function expression creation can be from 2% to 20% slower!?

Unless your intent with this test *is* to consider the inline function expression creation "cost," a better/fairer test would put `mySort(..)`'s declaration in the page setup -- don't put it in the test `setup` as that's unnecessary redeclaration for each cycle -- and simply reference it by name in the test case: `x.sort(mySort)`.

Building on the previous example, another pitfall is in opaquely avoiding or adding "extra work" to one test case that creates an apples-to-oranges scenario:

```
// Case 1
var x = [12, -14, 0, 3, 18, 0, 2.9];
x.sort();

// Case 2
var x = [12, -14, 0, 3, 18, 0, 2.9];
```

```
x.sort( function mySort(a,b){  
    return a - b;  
} );
```

Setting aside the previously mentioned inline function expression pitfall, the second case's `mySort(...)` works in this case because you have provided it numbers, but would have of course failed with strings. The first case doesn't throw an error, but it actually behaves differently and has a different outcome! It should be obvious, but: **a different outcome between two test cases almost certainly invalidates the entire test!**

But beyond the different outcomes, in this case, the built in `sort(...)`'s comparator is actually doing "extra work" that `mySort()` does not, in that the built-in one coerces the compared values to strings and does lexicographic comparison. The first snippet results in `[-14, 0, 0, 12, 18, 2.9, 3]` while the second snippet results (likely more accurately based on intent) in `[-14, 0, 0, 2.9, 3, 12, 18]`.

So that test is unfair because it's not actually doing the same task between the cases. Any results you get are bogus.

These same pitfalls can even be much more subtle:

```
// Case 1  
var x = false;  
var y = x ? 1 : 2;  
  
// Case 2  
var x;  
var y = x ? 1 : 2;
```

Here, the intent might be to test the performance impact of the coercion to a Boolean that the `? :` operator will do if the `x` expression is not already a Boolean (see the *Types & Grammar* title of this book series). So, you're apparently OK with the fact that there is extra work to do the coercion in the second case.

The subtle problem? You're setting `x`'s value in the first case and not setting it in the other, so you're actually doing work in the first case that you're not doing in the second. To eliminate any potential (albeit minor) skew, try:

```
// Case 1  
var x = false;  
var y = x ? 1 : 2;  
  
// Case 2  
var x = undefined;  
var y = x ? 1 : 2;
```

Now there's an assignment in both cases, so the thing you want to test -- the coercion of `x` or not -- has likely been more accurately isolated and tested.

## Writing Good Tests

Let me see if I can articulate the bigger point I'm trying to make here.

Good test authoring requires careful analytical thinking about what differences exist between two test cases and whether the differences between them are *intentional* or *unintentional*.

Intentional differences are of course normal and OK, but it's too easy to create unintentional differences that skew your results. You have to be really, really careful to avoid that skew. Moreover, you may intend a difference but it may not be obvious to other readers of your test what your intent was, so they may doubt (or trust!) your test incorrectly. How do you fix that?

**Write better, clearer tests.** But also, take the time to document (using the jsPerf.com "Description" field and/or code comments) exactly what the intent of your test is, even to the nuanced detail. Call out the intentional differences, which will help others and your future self to better identify unintentional differences that could be skewing the test results.

Isolate things which aren't relevant to your test by pre-declaring them in the page or test setup settings so they're outside the timed parts of the test.

Instead of trying to narrow in on a tiny snippet of your real code and benchmarking just that piece out of context, tests and benchmarks are better when they include a larger (while still relevant) context. Those tests also tend to run slower, which means any differences you spot are more relevant in context.

## Microperformance

OK, until now we've been dancing around various microperformance issues and generally looking disfavorably upon obsessing about them. I want to take just a moment to address them directly.

The first thing you need to get more comfortable with when thinking about performance benchmarking your code is that the code you write is not always the code the engine actually runs. We briefly looked at that topic back in Chapter 1 when we discussed statement reordering by the compiler, but here we're going to suggest the compiler can sometimes decide to run different code than you wrote, not just in different orders but different in substance.

Let's consider this piece of code:

```
var foo = 41;

(function(){
  (function(){
    (function(baz){
      var bar = foo + baz;
      // ...
    })(1);
  })();
})();
```

You may think about the `foo` reference in the innermost function as needing to do a three-level scope lookup. We covered in the *Scope & Closures* title of this book series how lexical scope works, and the fact that the compiler generally caches such lookups so that referencing `foo` from different scopes doesn't really practically "cost" anything extra.

But there's something deeper to consider. What if the compiler realizes that `foo` isn't referenced anywhere else but that one location, and it further notices that the value never is anything except the `41` as shown?

Isn't it quite possible and acceptable that the JS compiler could decide to just remove the `foo` variable entirely, and *inline* the value, such as this:

```
(function(){
  (function(){
    (function(baz){
      var bar = 41 + baz;
      // ...
    })(1);
  })();
})();
```

**Note:** Of course, the compiler could probably also do a similar analysis and rewrite with the `baz` variable here, too.

When you begin to think about your JS code as being a hint or suggestion to the engine of what to do, rather than a literal requirement, you realize that a lot of the obsession over discrete syntactic minutia is most likely unfounded.

Another example:

```
function factorial(n) {
  if (n < 2) return 1;
  return n * factorial( n - 1 );
}

factorial( 5 );           // 120
```

Ah, the good ol' fashioned "factorial" algorithm! You might assume that the JS engine will run that code mostly as is. And to be honest, it might -- I'm not really sure.

But as an anecdote, the same code expressed in C and compiled with advanced optimizations would result in the compiler realizing that the call `factorial(5)` can just be replaced with the constant value `120`, eliminating the function and call entirely!

Moreover, some engines have a practice called "unrolling recursion," where it can realize that the recursion you've expressed can actually be done "easier" (i.e., more optimally) with a loop. It's possible the preceding code could be *rewritten* by a JS engine to run as:

```
function factorial(n) {
  if (n < 2) return 1;

  var res = 1;
  for (var i=n; i>1; i--) {
    res *= i;
  }
  return res;
}

factorial( 5 );           // 120
```

Now, let's imagine that in the earlier snippet you had been worried about whether `n * factorial(n-1)` or `n *= factorial(--n)` runs faster. Maybe you even did a performance benchmark to try to figure out which was better. But you miss the fact that in the bigger context, the engine may not run either line of code because it may unroll the recursion!

Speaking of `--`, `--n` versus `n--` is often cited as one of those places where you can optimize by choosing the `--n` version, because theoretically it requires less effort down at the assembly level of processing.

That sort of obsession is basically nonsense in modern JavaScript. That's the kind of thing you should be letting the engine take care of. You should write the code that makes the most sense. Compare these three `for` loops:

```
// Option 1
for (var i=0; i<10; i++) {
  console.log( i );
}

// Option 2
for (var i=0; i<10; ++i) {
  console.log( i );
}

// Option 3
for (var i=-1; ++i<10; ) {
  console.log( i );
}
```

Even if you have some theory where the second or third option is more performant than the first option by a tiny bit, which is dubious at best, the third loop is more confusing because you have to start with `-1` for `i` to account for the fact that `++i` pre-increment is used. And the difference between the first and second options is really quite irrelevant.

It's entirely possible that a JS engine may see a place where `i++` is used and realize that it can safely replace it with the `+i` equivalent, which means your time spent deciding which one to pick was completely wasted and the outcome moot.

Here's another common example of silly microperformance obsession:

```
var x = [ .. ];

// Option 1
for (var i=0; i < x.length; i++) {
    // ..
}

// Option 2
for (var i=0, len = x.length; i < len; i++) {
    // ..
}
```

The theory here goes that you should cache the length of the `x` array in the variable `len`, because ostensibly it doesn't change, to avoid paying the price of `x.length` being consulted for each iteration of the loop.

If you run performance benchmarks around `x.length` usage compared to caching it in a `len` variable, you'll find that while the theory sounds nice, in practice any measured differences are statistically completely irrelevant.

In fact, in some engines like v8, it can be shown (<http://mrale.ph/blog/2014/12/24/array-length-caching.html>) that you could make things slightly worse by pre-caching the length instead of letting the engine figure it out for you. Don't try to outsmart your JavaScript engine, you'll probably lose when it comes to performance optimizations.

## Not All Engines Are Alike

The different JS engines in various browsers can all be "spec compliant" while having radically different ways of handling code. The JS specification doesn't require anything performance related -- well, except ES6's "Tail Call Optimization" covered later in this chapter.

The engines are free to decide that one operation will receive its attention to optimize, perhaps trading off for lesser performance on another operation. It can be very tenuous to find an approach for an operation that always runs faster in all browsers.

There's a movement among some in the JS dev community, especially those who work with Node.js, to analyze the specific internal implementation details of the v8 JavaScript engine and make decisions about writing JS code that is tailored to take best advantage of how v8 works. You can actually achieve a surprisingly high degree of performance optimization with such endeavors, so the payoff for the effort can be quite high.

Some commonly cited examples (<https://github.com/petkaantonov/bluebird/wiki/Optimization-killers>) for v8:

- Don't pass the `arguments` variable from one function to any other function, as such "leakage" slows down the function implementation.
- Isolate a `try..catch` in its own function. Browsers struggle with optimizing any function with a `try..catch` in it, so moving that construct to its own function means you contain the de-optimization harm while letting the surrounding code be optimizable.

But rather than focus on those tips specifically, let's sanity check the v8-only optimization approach in a general sense.

Are you genuinely writing code that only needs to run in one JS engine? Even if your code is entirely intended for Node.js *right now*, is the assumption that v8 will *always* be the used JS engine reliable? Is it possible that someday a few years from now, there's another server-side JS platform besides Node.js that you choose to run your code on? What if what you optimized for before is now a much slower way of doing that operation on the new engine?

Or what if your code always stays running on v8 from here on out, but v8 decides at some point to change the way some set of operations works such that what used to be fast is now slow, and vice versa?

These scenarios aren't just theoretical, either. It used to be that it was faster to put multiple string values into an array and then call `join("")` on the array to concatenate the values than to just use `+` concatenation directly with the values. The historical reason for this is nuanced, but it has to do with internal implementation details about how string values were stored and managed in memory.

As a result, "best practice" advice at the time disseminated across the industry suggesting developers always use the array `join(..)` approach. And many followed.

Except, somewhere along the way, the JS engines changed approaches for internally managing strings, and specifically put in optimizations for `+` concatenation. They didn't slow down `join(..)` per se, but they put more effort into helping `+` usage, as it was still quite a bit more widespread.

**Note:** The practice of standardizing or optimizing some particular approach based mostly on its existing widespread usage is often called (metaphorically) "paving the cowpath."

Once that new approach to handling strings and concatenation took hold, unfortunately all the code out in the wild that was using array `join(..)` to concatenate strings was then sub-optimal.

Another example: at one time, the Opera browser differed from other browsers in how it handled the boxing/unboxing of primitive wrapper objects (see the *Types & Grammar* title of this book series). As such, their advice to developers was to use a `string` object instead of the primitive `string` value if properties like `length` or methods like `charAt(..)` needed to be accessed. This advice may have been correct for Opera at the time, but it was literally completely opposite for other major contemporary browsers, as they had optimizations specifically for the `string` primitives and not their object wrapper counterparts.

I think these various gotchas are at least possible, if not likely, for code even today. So I'm very cautious about making wide ranging performance optimizations in my JS code based purely on engine implementation details, **especially if those details are only true of a single engine.**

The reverse is also something to be wary of: you shouldn't necessarily change a piece of code to work around one engine's difficulty with running a piece of code in an acceptably performant way.

Historically, IE has been the brunt of many such frustrations, given that there have been plenty of scenarios in older IE versions where it struggled with some performance aspect that other major browsers of the time seemed not to have much trouble with. The string concatenation discussion we just had was actually a real concern back in the IE6 and IE7 days, where it was possible to get better performance out of `join(..)` than `+`.

But it's troublesome to suggest that just one browser's trouble with performance is justification for using a code approach that quite possibly could be sub-optimal in all other browsers. Even if the browser in question has a large market share for your site's audience, it may be more practical to write the proper code and rely on the browser to update itself with better optimizations eventually.

"There is nothing more permanent than a temporary hack." Chances are, the code you write now to work around some performance bug will probably outlive the performance bug in the browser itself.

In the days when a browser only updated once every five years, that was a tougher call to make. But as it stands now, browsers across the board are updating at a much more rapid interval (though obviously the mobile world still lags), and they're all competing to optimize web features better and better.

If you run across a case where a browser *does* have a performance wart that others don't suffer from, make sure to report it to them through whatever means you have available. Most browsers have open public bug trackers suitable for this purpose.

**Tip:** I'd only suggest working around a performance issue in a browser if it was a really drastic show-stopper, not just an annoyance or frustration. And I'd be very careful to check that the performance hack didn't have noticeable negative side effects in another browser.

## Big Picture

---

Instead of worrying about all these microperformance nuances, we should instead be looking at big-picture types of optimizations.

How do you know what's big picture or not? You have to first understand if your code is running on a critical path or not. If it's not on the critical path, chances are your optimizations are not worth much.

Ever heard the admonition, "that's premature optimization!"? It comes from a famous quote from Donald Knuth: "premature optimization is the root of all evil.". Many developers cite this quote to suggest that most optimizations are "premature" and are thus a waste of effort. The truth is, as usual, more nuanced.

Here is Knuth's quote, in context:

Programmers waste enormous amounts of time thinking about, or worrying about, the speed of **noncritical** parts of their programs, and these attempts at efficiency actually have a strong negative impact when debugging and maintenance are considered. We should forget about small efficiencies, say about 97% of the time: premature optimization is the root of all evil. Yet we should not pass up our opportunities in that **critical** 3%. [emphasis added]

([http://web.archive.org/web/20130731202547/http://pplab.snu.ac.kr/courses/adv\\_pl05/papers/p261-knuth.pdf](http://web.archive.org/web/20130731202547/http://pplab.snu.ac.kr/courses/adv_pl05/papers/p261-knuth.pdf), Computing Surveys, Vol 6, No 4, December 1974)

I believe it's a fair paraphrasing to say that Knuth *meant*: "non-critical path optimization is the root of all evil." So the key is to figure out if your code is on the critical path -- you should optimize it! -- or not.

I'd even go so far as to say this: no amount of time spent optimizing critical paths is wasted, no matter how little is saved; but no amount of optimization on noncritical paths is justified, no matter how much is saved.

If your code is on the critical path, such as a "hot" piece of code that's going to be run over and over again, or in UX critical places where users will notice, like an animation loop or CSS style updates, then you should spare no effort in trying to employ relevant, measurably significant optimizations.

For example, consider a critical path animation loop that needs to coerce a string value to a number. There are of course multiple ways to do that (see the *Types & Grammar* title of this book series), but which one if any is the fastest?

```
var x = "42";      // need number `42`  
  
// Option 1: let implicit coercion automatically happen  
var y = x / 2;  
  
// Option 2: use `parseInt(..)`  
var y = parseInt( x, 0 ) / 2;  
  
// Option 3: use `Number(..)`  
var y = Number( x ) / 2;  
  
// Option 4: use `+` unary operator  
var y = +x / 2;  
  
// Option 5: use `|` unary operator  
var y = (x | 0) / 2;
```

**Note:** I will leave it as an exercise to the reader to set up a test if you're interested in examining the minute differences in performance among these options.

When considering these different options, as they say, "One of these things is not like the others." `parseInt(..)` does the job, but it also does a lot more -- it parses the string rather than just coercing. You can probably guess, correctly, that `parseInt(..)` is a slower option, and you should probably avoid it.

Of course, if `x` can ever be a value that **needs parsing**, such as `"42px"` (like from a CSS style lookup), then `parseInt(..)` really is the only suitable option!

`Number(..)` is also a function call. From a behavioral perspective, it's identical to the `+ unary operator` option, but it may in fact be a little slower, requiring more machinery to execute the function. Of course, it's also possible that the JS engine recognizes this behavioral symmetry and just handles the inlining of `Number(..)`'s behavior (aka `+x`) for you!

But remember, obsessing about `+x` versus `x | 0` is in most cases likely a waste of effort. This is a microperformance issue, and one that you shouldn't let dictate/upgrade the readability of your program.

While performance is very important in critical paths of your program, it's not the only factor. Among several options that are roughly similar in performance, readability should be another important concern.

## Tail Call Optimization (TCO)

As we briefly mentioned earlier, ES6 includes a specific requirement that ventures into the world of performance. It's related to a specific form of optimization that can occur with function calls: *tail call optimization*.

Briefly, a "tail call" is a function call that appears at the "tail" of another function, such that after the call finishes, there's nothing left to do (except perhaps return its result value).

For example, here's a non-recursive setup with tail calls:

```
function foo(x) {
  return x;
}

function bar(y) {
  return foo(y + 1);    // tail call
}

function baz() {
  return 1 + bar(40);   // not tail call
}

baz();                // 42
```

`foo(y+1)` is a tail call in `bar(..)` because after `foo(..)` finishes, `bar(..)` is also finished except in this case returning the result of the `foo(..)` call. However, `bar(40)` is *not* a tail call because after it completes, its result value must be added to `1` before `baz()` can return it.

Without getting into too much nitty-gritty detail, calling a new function requires an extra amount of reserved memory to manage the call stack, called a "stack frame." So the preceding snippet would generally require a stack frame for each of `baz()`, `bar(..)`, and `foo(..)` all at the same time.

However, if a TCO-capable engine can realize that the `foo(y+1)` call is in *tail position* meaning `bar(..)` is basically complete, then when calling `foo(..)`, it doesn't need to create a new stack frame, but can instead reuse the existing stack frame from `bar(..)`. That's not only faster, but it also uses less memory.

That sort of optimization isn't a big deal in a simple snippet, but it becomes a *much bigger deal* when dealing with recursion, especially if the recursion could have resulted in hundreds or thousands of stack frames. With TCO the engine can perform all those calls with a single stack frame!

Recursion is a hairy topic in JS because without TCO, engines have had to implement arbitrary (and different!) limits to how deep they will let the recursion stack get before they stop it, to prevent running out of memory. With TCO, recursive functions with *tail position* calls can essentially run unbounded, because there's never any extra usage of memory!

Consider that recursive `factorial(..)` from before, but rewritten to make it TCO friendly:

```
function factorial(n) {
  function fact(n, res) {
```

```

        if (n < 2) return res;

        return fact( n - 1, n * res );
    }

    return fact( n, 1 );
}

factorial( 5 );           // 120

```

This version of `factorial(...)` is still recursive, but it's also optimizable with TCO, because both inner `fact(...)` calls are in *tail position*.

**Note:** It's important to note that TCO only applies if there's actually a tail call. If you write recursive functions without tail calls, the performance will still fall back to normal stack frame allocation, and the engines' limits on such recursive call stacks will still apply. Many recursive functions can be rewritten as we just showed with `factorial(...)`, but it takes careful attention to detail.

One reason that ES6 requires engines to implement TCO rather than leaving it up to their discretion is because the *lack of TCO* actually tends to reduce the chances that certain algorithms will be implemented in JS using recursion, for fear of the call stack limits.

If the lack of TCO in the engine would just gracefully degrade to slower performance in all cases, it wouldn't probably have been something that ES6 needed to *require*. But because the lack of TCO can actually make certain programs impractical, it's more an important feature of the language than just a hidden implementation detail.

ES6 guarantees that from now on, JS developers will be able to rely on this optimization across all ES6+ compliant browsers. That's a win for JS performance!

## Review

Effectively benchmarking performance of a piece of code, especially to compare it to another option for that same code to see which approach is faster, requires careful attention to detail.

Rather than rolling your own statistically valid benchmarking logic, just use the Benchmark.js library, which does that for you. But be careful about how you author tests, because it's far too easy to construct a test that seems valid but that's actually flawed -- even tiny differences can skew the results to be completely unreliable.

It's important to get as many test results from as many different environments as possible to eliminate hardware/device bias. jsPerf.com is a fantastic website for crowdsourcing performance benchmark test runs.

Many common performance tests unfortunately obsess about irrelevant microperformance details like `x++` versus `++x`. Writing good tests means understanding how to focus on big picture concerns, like optimizing on the critical path, and avoiding falling into traps like different JS engines' implementation details.

Tail call optimization (TCO) is a required optimization as of ES6 that will make some recursive patterns practical in JS where they would have been impossible otherwise. TCO allows a function call in the *tail position* of another function to execute without needing any extra resources, which means the engine no longer needs to place arbitrary restrictions on call stack depth for recursive algorithms.

# You Don't Know JS Yet: Sync & Async - 2nd Edition

## Appendix A: *asynquence* Library

<b>NOTE:</b>
Work in progress

<b>NOTE:</b>
Everything below here is previous text from 1st edition, and is only here for reference while 2nd edition work is underway. <b>Please ignore this stuff.</b>

Everything below here is previous text from 1st edition, and is only here for reference while 2nd edition work is underway.  
**Please ignore this stuff.**

Chapters 1 and 2 went into quite a bit of detail about typical asynchronous programming patterns and how they're commonly solved with callbacks. But we also saw why callbacks are fatally limited in capability, which led us to Chapters 3 and 4, with Promises and generators offering a much more solid, trustable, and reasonable base to build your asynchrony on.

I referenced my own asynchronous library *asynquence* (<http://github.com/getify/asynquence>) -- "async" + "sequence" = "asynquence" -- several times in this book, and I want to now briefly explain how it works and why its unique design is important and helpful.

In the next appendix, we'll explore some advanced async patterns, but you'll probably want a library to make those palatable enough to be useful. We'll use *asynquence* to express those patterns, so you'll want to spend a little time here getting to know the library first.

*asynquence* is obviously not the only option for good async coding; certainly there are many great libraries in this space. But *asynquence* provides a unique perspective by combining the best of all these patterns into a single library, and moreover is built on a single basic abstraction: the (async) sequence.

My premise is that sophisticated JS programs often need bits and pieces of various different asynchronous patterns woven together, and this is usually left entirely up to each developer to figure out. Instead of having to bring in two or more different async libraries that focus on different aspects of asynchrony, *asynquence* unifies them into variated sequence steps, with just one core library to learn and deploy.

I believe the value is strong enough with *asynquence* to make async flow control programming with Promise-style semantics super easy to accomplish, so that's why we'll exclusively focus on that library here.

To begin, I'll explain the design principles behind *asynquence*, and then we'll illustrate how its API works with code examples.

## Sequences, Abstraction Design

Understanding *asynquence* begins with understanding a fundamental abstraction: any series of steps for a task, whether they separately are synchronous or asynchronous, can be collectively thought of as a "sequence". In other words, a sequence is a container that represents a task, and is comprised of individual (potentially *async*) steps to complete that task.

Each step in the sequence is controlled under the covers by a Promise (see Chapter 3). That is, every step you add to a sequence implicitly creates a Promise that is wired to the previous end of the sequence. Because of the semantics of Promises, every single step advancement in a sequence is asynchronous, even if you synchronously complete the step.

Moreover, a sequence will always proceed linearly from step to step, meaning that step 2 always comes after step 1 finishes, and so on.

Of course, a new sequence can be forked off an existing sequence, meaning the fork only occurs once the main sequence reaches that point in the flow. Sequences can also be combined in various ways, including having one sequence subsumed by another sequence at a particular point in the flow.

A sequence is kind of like a Promise chain. However, with Promise chains, there is no "handle" to grab that references the entire chain. Whichever Promise you have a reference to only represents the current step in the chain plus any other steps hanging off it. Essentially, you cannot hold a reference to a Promise chain unless you hold a reference to the first Promise in the chain.

There are many cases where it turns out to be quite useful to have a handle that references the entire sequence collectively. The most important of those cases is with sequence `abort/cancel`. As we covered extensively in Chapter 3, Promises themselves should never be able to be canceled, as this violates a fundamental design imperative: external immutability.

But sequences have no such immutability design principle, mostly because sequences are not passed around as future-value containers that need immutable value semantics. So sequences are the proper level of abstraction to handle `abort/cancel` behavior. *asynquence* sequences can be `abort()`ed at any time, and the sequence will stop at that point and not go for any reason.

There's plenty more reasons to prefer a sequence abstraction on top of Promise chains, for flow control purposes.

First, Promise chaining is a rather manual process -- one that can get pretty tedious once you start creating and chaining Promises across a wide swath of your programs -- and this tedium can act counterproductively to dissuade the developer from using Promises in places where they are quite appropriate.

Abstractions are meant to reduce boilerplate and tedium, so the sequence abstraction is a good solution to this problem. With Promises, your focus is on the individual step, and there's little assumption that you will keep the chain going. With sequences, the opposite approach is taken, assuming the sequence will keep having more steps added indefinitely.

This abstraction complexity reduction is especially powerful when you start thinking about higher-order Promise patterns (beyond `race([...])` and `all([...])`).

For example, in the middle of a sequence, you may want to express a step that is conceptually like a `try..catch` in that the step will always result in success, either the intended main success resolution or a positive nonerror signal for the caught error. Or, you might want to express a step that is like a retry/until loop, where it keeps trying the same step over and over until success occurs.

These sorts of abstractions are quite nontrivial to express using only Promise primitives, and doing so in the middle of an existing Promise chain is not pretty. But if you abstract your thinking to a sequence, and consider a step as a wrapper around a Promise, that step wrapper can hide such details, freeing you to think about the flow control in the most sensible way without being bothered by the details.

Second, and perhaps more importantly, thinking of *async* flow control in terms of steps in a sequence allows you to abstract out the details of what types of asynchronicity are involved with each individual step. Under the covers, a Promise will always control the step, but above the covers, that step can look either like a continuation callback (the simple default), or like a real Promise, or as a run-to-completion generator, or ... Hopefully, you get the picture.

Third, sequences can more easily be twisted to adapt to different modes of thinking, such as event-, stream-, or reactive-based coding. *asynquence* provides a pattern I call "reactive sequences" (which we'll cover later) as a variation on the "reactive observable" ideas in RxJS ("Reactive Extensions"), that lets a repeatable event fire off a new sequence instance each time. Promises are one-shot-only, so it's quite awkward to express repetitious asynchrony with Promises alone.

Another alternate mode of thinking inverts the resolution/control capability in a pattern I call "iterable sequences". Instead of each individual step internally controlling its own completion (and thus advancement of the sequence), the sequence is inverted so the advancement control is through an external iterator, and each step in the *iterable sequence* just responds to the `next(..)` iterator control.

We'll explore all of these different variations as we go throughout the rest of this appendix, so don't worry if we ran over those bits far too quickly just now.

The takeaway is that sequences are a more powerful and sensible abstraction for complex asynchrony than just Promises (Promise chains) or just generators, and *asynquence* is designed to express that abstraction with just the right level of sugar to make async programming more understandable and more enjoyable.

## asynquence API

To start off, the way you create a sequence (an *asynquence* instance) is with the `ASQ(..)` function. An `ASQ()` call with no parameters creates an empty initial sequence, whereas passing one or more values or functions to `ASQ(..)` sets up the sequence with each argument representing the initial steps of the sequence.

**Note:** For the purposes of all code examples here, I will use the *asynquence* top-level identifier in global browser usage: `ASQ`. If you include and use *asynquence* through a module system (browser or server), you of course can define whichever symbol you prefer, and *asynquence* won't care!

Many of the API methods discussed here are built into the core of *asynquence*, but others are provided through including the optional "contrib" plug-ins package. See the documentation for *asynquence* for whether a method is built in or defined via plug-in: <http://github.com/getify/asynquence>

## Steps

If a function represents a normal step in the sequence, that function is invoked with the first parameter being the continuation callback, and any subsequent parameters being any messages passed on from the previous step. The step will not complete until the continuation callback is called. Once it's called, any arguments you pass to it will be sent along as messages to the next step in the sequence.

To add an additional normal step to the sequence, call `then(..)` (which has essentially the exact same semantics as the `ASQ(..)` call):

```
ASQ(
  // step 1
  function(done){
    setTimeout( function(){
      done( "Hello" );
    }, 100 );
  },
  // step 2
  function(done,greeting) {
    setTimeout( function(){
      done( greeting + " World" );
    }, 100 );
  }
)
// step 3
.then( function(done,msg){
```

```

    setTimeout( function(){
        done( msg.toUpperCase() );
    }, 100 );
}
// step 4
.then( function(done,msg){
    console.log( msg );           // HELLO WORLD
} );

```

**Note:** Though the name `then(..)` is identical to the native Promises API, this `then(..)` is different. You can pass as few or as many functions or values to `then(..)` as you'd like, and each is taken as a separate step. There's no two-callback fulfilled/rejected semantics involved.

Unlike with Promises, where to chain one Promise to the next you have to create and `return` that Promise from a `then(..)` fulfillment handler, with *asynquence*, all you need to do is call the continuation callback -- I always call it `done()` but you can name it whatever suits you -- and optionally pass it completion messages as arguments.

Each step defined by `then(..)` is assumed to be asynchronous. If you have a step that's synchronous, you can either just call `done(..)` right away, or you can use the simpler `val(..)` step helper:

```

// step 1 (sync)
ASQ( function(done){
    done( "Hello" );      // manually synchronous
} )
// step 2 (sync)
.val( function(greeting){
    return greeting + " World";
} )
// step 3 (async)
.then( function(done,msg){
    setTimeout( function(){
        done( msg.toUpperCase() );
    }, 100 );
} )
// step 4 (sync)
.val( function(msg){
    console.log( msg );
} );

```

As you can see, `val(..)`-invoked steps don't receive a continuation callback, as that part is assumed for you -- and the parameter list is less cluttered as a result! To send a message along to the next step, you simply use `return`.

Think of `val(..)` as representing a synchronous "value-only" step, which is useful for synchronous value operations, logging, and the like.

## Errors

One important difference with *asynquence* compared to Promises is with error handling.

With Promises, each individual Promise (step) in a chain can have its own independent error, and each subsequent step has the ability to handle the error or not. The main reason for this semantic comes (again) from the focus on individual Promises rather than on the chain (sequence) as a whole.

I believe that most of the time, an error in one part of a sequence is generally not recoverable, so the subsequent steps in the sequence are moot and should be skipped. So, by default, an error at any step of a sequence throws the entire sequence into error mode, and the rest of the normal steps are ignored.

If you *do* need to have a step where its error is recoverable, there are several different API methods that can accommodate, such as `try(..)` -- previously mentioned as a kind of `try..catch` step -- or `until(..)` -- a retry loop that keeps attempting the step until it succeeds or you manually `break()` the loop. *asynquence* even has `pThen(..)` and `pCatch(..)` methods, which work

identically to how normal Promise `then(..)` and `catch(..)` work (see Chapter 3), so you can do localized mid-sequence error handling if you so choose.

The point is, you have both options, but the more common one in my experience is the default. With Promises, to get a chain of steps to ignore all steps once an error occurs, you have to take care not to register a rejection handler at any step; otherwise, that error gets swallowed as handled, and the sequence may continue (perhaps unexpectedly). This kind of desired behavior is a bit awkward to properly and reliably handle.

To register a sequence error notification handler, *asynquence* provides an `or(..)` sequence method, which also has an alias of `onerror(..)`. You can call this method anywhere in the sequence, and you can register as many handlers as you'd like. That makes it easy for multiple different consumers to listen in on a sequence to know if it failed or not; it's kind of like an error event handler in that respect.

Just like with Promises, all JS exceptions become sequence errors, or you can programmatically signal a sequence error:

```
var sq = ASQ( function(done){
  setTimeout( function(){
    // signal an error for the sequence
    done.fail( "Oops" );
  }, 100 );
} )
.then( function(done){
  // will never get here
} )
.or( function(err){
  console.log( err );           // Oops
} )
.then( function(done){
  // won't get here either
} );

// later

sq.or( function(err){
  console.log( err );           // Oops
} );
```

Another really important difference with error handling in *asynquence* compared to native Promises is the default behavior of "unhandled exceptions". As we discussed at length in Chapter 3, a rejected Promise without a registered rejection handler will just silently hold (aka swallow) the error; you have to remember to always end a chain with a final `catch(..)`.

In *asynquence*, the assumption is reversed.

If an error occurs on a sequence, and it **at that moment** has no error handlers registered, the error is reported to the `console`. In other words, unhandled rejections are by default always reported so as not to be swallowed and missed.

As soon as you register an error handler against a sequence, it opts that sequence out of such reporting, to prevent duplicate noise.

There may, in fact, be cases where you want to create a sequence that may go into the error state before you have a chance to register the handler. This isn't common, but it can happen from time to time.

In those cases, you can also **opt a sequence instance out** of error reporting by calling `defer()` on the sequence. You should only opt out of error reporting if you are sure that you're going to eventually handle such errors:

```
var sq1 = ASQ( function(done){
  doesntExist();           // will throw exception to console
} );

var sq2 = ASQ( function(done){
  doesntExist();           // will throw only a sequence error
} );
// opt-out of error reporting
```

```
.defer();

setTimeout( function(){
    sq1.or( function(err){
        console.log( err ); // ReferenceError
    } );
}

sq2.or( function(err){
    console.log( err ); // ReferenceError
} );
}, 100 );

// ReferenceError (from sq1)
```

This is better error handling behavior than Promises themselves have, because it's the Pit of Success, not the Pit of Failure (see Chapter 3).

**Note:** If a sequence is piped into (aka subsumed by) another sequence -- see "Combining Sequences" for a complete description -- then the source sequence is opted out of error reporting, but now the target sequence's error reporting or lack thereof must be considered.

## Parallel Steps

Not all steps in your sequences will have just a single (async) task to perform; some will need to perform multiple steps "in parallel" (concurrently). A step in a sequence in which multiple substeps are processing concurrently is called a `gate(..)` -- there's an `all(..)` alias if you prefer -- and is directly symmetric to native `Promise.all([..])`.

If all the steps in the `gate(..)` complete successfully, all success messages will be passed to the next sequence step. If any of them generate errors, the whole sequence immediately goes into an error state.

Consider:

```
ASQ( function(done){
    setTimeout( done, 100 );
} )
.gate(
    function(done){
        setTimeout( function(){
            done( "Hello" );
        }, 100 );
    },
    function(done){
        setTimeout( function(){
            done( "World", "!" );
        }, 100 );
    }
)
.val( function(msg1,msg2){
    console.log( msg1 ); // Hello
    console.log( msg2 ); // [ "World", "!" ]
} );
```

For illustration, let's compare that example to native Promises:

```
new Promise( function(resolve,reject){
    setTimeout( resolve, 100 );
} )
.then( function(){
    return Promise.all( [
        new Promise( function(resolve,reject){
            setTimeout( function(){
                resolve( "Hello" );
            }, 100 );
        })
    ]);
}
)
.then( function(result){
    console.log( result ); // [ "Hello" ]
} );
```

```

        }, 100 );
    } ),
    new Promise( function(resolve,reject){
        setTimeout( function(){
            // note: we need a [ ] array here
            resolve( [ "World", "!" ] );
        }, 100 );
    } )
],
);
.then( function(msgs){
    console.log( msgs[0] ); // Hello
    console.log( msgs[1] ); // [ "World", "!" ]
} );

```

Yuck. Promises require a lot more boilerplate overhead to express the same asynchronous flow control. That's a great illustration of why the *asynquence* API and abstraction make dealing with Promise steps a lot nicer. The improvement only goes higher the more complex your asynchrony is.

## Step Variations

There are several variations in the contrib plug-ins on *asynquence*'s `gate(..)` step type that can be quite helpful:

- `any(..)` is like `gate(..)`, except just one segment has to eventually succeed to proceed on the main sequence.
- `first(..)` is like `any(..)`, except as soon as any segment succeeds, the main sequence proceeds (ignoring subsequent results from other segments).
- `race(..)` (symmetric with `Promise.race(..)`) is like `first(..)`, except the main sequence proceeds as soon as any segment completes (either success or failure).
- `last(..)` is like `any(..)`, except only the latest segment to complete successfully sends its message(s) along to the main sequence.
- `none(..)` is the inverse of `gate(..)`: the main sequence proceeds only if all the segments fail (with all segment error message(s) transposed as success message(s) and vice versa).

Let's first define some helpers to make illustration cleaner:

```

function success1(done) {
    setTimeout( function(){
        done( 1 );
    }, 100 );
}

function success2(done) {
    setTimeout( function(){
        done( 2 );
    }, 100 );
}

function failure3(done) {
    setTimeout( function(){
        done.fail( 3 );
    }, 100 );
}

function output(msg) {
    console.log( msg );
}

```

Now, let's demonstrate these `gate(..)` step variations:

```

ASQ().race(
    failure3,

```

```

        success1
    )
.or( output );           // 3

ASQ().any(
    success1,
    failure3,
    success2
)
.val( function(){
    var args = [].slice.call( arguments );
    console.log(
        args          // [ 1, undefined, 2 ]
    );
} );

ASQ().first(
    failure3,
    success1,
    success2
)
.val( output );           // 1

ASQ().last(
    failure3,
    success1,
    success2
)
.val( output );           // 2

ASQ().none(
    failure3
)
.val( output )           // 3
.none(
    failure3
    success1
)
.or( output );           // 1

```

Another step variation is `map(..)`, which lets you asynchronously map elements of an array to different values, and the step doesn't proceed until all the mappings are complete. `map(..)` is very similar to `gate(..)`, except it gets the initial values from an array instead of from separately specified functions, and also because you define a single function callback to operate on each value:

```

function double(x,done) {
    setTimeout( function(){
        done( x * 2 );
    }, 100 );
}

ASQ().map( [1,2,3], double )
.val( output );           // [2,4,6]

```

Also, `map(..)` can receive either of its parameters (the array or the callback) from messages passed from the previous step:

```

function plusOne(x,done) {
    setTimeout( function(){
        done( x + 1 );
    }, 100 );
}

```

```
ASQ( [1,2,3] )
  .map( double )           // message `[1,2,3]` comes in
  .map( plusOne )          // message `[2,4,6]` comes in
  .val( output );          // [3,5,7]
```

Another variation is `waterfall(..)`, which is kind of like a mixture between `gate(..)`'s message collection behavior but `then(..)`'s sequential processing.

Step 1 is first executed, then the success message from step 1 is given to step 2, and then both success messages go to step 3, and then all three success messages go to step 4, and so on, such that the messages sort of collect and cascade down the "waterfall".

Consider:

```
function double(done) {
  var args = [].slice.call( arguments, 1 );
  console.log( args );

  setTimeout( function(){
    done( args[args.length - 1] * 2 );
  }, 100 );
}

ASQ( 3 )
  .waterfall(
    double,                  // [ 3 ]
    double,                  // [ 6 ]
    double,                  // [ 6, 12 ]
    double                   // [ 6, 12, 24 ]
  )
  .val( function(){
    var args = [].slice.call( arguments );
    console.log( args );    // [ 6, 12, 24, 48 ]
  } );
```

If at any point in the "waterfall" an error occurs, the whole sequence immediately goes into an error state.

## Error Tolerance

Sometimes you want to manage errors at the step level and not let them necessarily send the whole sequence into the error state. `asynquence` offers two step variations for that purpose.

`try(..)` attempts a step, and if it succeeds, the sequence proceeds as normal, but if the step fails, the failure is turned into a success message formated as `{ catch: .. }` with the error message(s) filled in:

```
ASQ()
  .try( success1 )           // 1
  .val( output )             // { catch: 1 }
  .try( failure3 )           // { catch: 3 }
  .val( output )             // { catch: 3 }
  .or( function(err){
    // never gets here
  } );
```

You could instead set up a retry loop using `until(..)`, which tries the step and if it fails, retries the step again on the next event loop tick, and so on.

This retry loop can continue indefinitely, but if you want to break out of the loop, you can call the `break()` flag on the completion trigger, which sends the main sequence into an error state:

```
var count = 0;
```

```

ASQ( 3 )
.until( double )
.val( output )           // 6
.until( function(done){
    count++;

    setTimeout( function(){
        if (count < 5) {
            done.fail();
        }
        else {
            // break out of the `until(..)` retry loop
            done.break( "Oops" );
        }
    }, 100 );
} )
.or( output );          // Oops

```

## Promise-Style Steps

If you would prefer to have, inline in your sequence, Promise-style semantics like Promises' `then(..)` and `catch(..)` (see Chapter 3), you can use the `pThen` and `pCatch` plug-ins:

```

ASQ( 21 )
.pThen( function(msg){
    return msg * 2;
} )
.pThen( output )           // 42
.pThen( function(){
    // throw an exception
    doesntExist();
} )
.pCatch( function(err){
    // caught the exception (rejection)
    console.log( err );      // ReferenceError
} )
.val( function(){
    // main sequence is back in a
    // success state because previous
    // exception was caught by
    // `pCatch(..)`
} );

```

`pThen(..)` and `pCatch(..)` are designed to run in the sequence, but behave as if it was a normal Promise chain. As such, you can either resolve genuine Promises or *asynquence* sequences from the "fulfillment" handler passed to `pThen(..)` (see Chapter 3).

## Forking Sequences

One feature that can be quite useful about Promises is that you can attach multiple `then(..)` handler registrations to the same promise, effectively "forking" the flow-control at that promise:

```

var p = Promise.resolve( 21 );

// fork 1 (from `p`)
p.then( function(msg){
    return msg * 2;
} )
.then( function(msg){
    console.log( msg );      // 42
} )

```

```
// fork 2 (from `p`)
p.then( function(msg){
    console.log( msg );           // 21
} );
```

The same "forking" is easy in *asynquence* with `fork()` :

```
var sq = ASQ(...).then(...).then(...);

var sq2 = sq.fork();

// fork 1
sq.then(...);

// fork 2
sq2.then(...);
```

## Combining Sequences

The reverse of `fork()` ing, you can combine two sequences by subsuming one into another, using the `seq(..)` instance method:

```
var sq = ASQ( function(done){
    setTimeout( function(){
        done( "Hello World" );
    }, 200 );
} );

ASQ( function(done){
    setTimeout( done, 100 );
} )
// subsume `sq` sequence into this sequence
.seq( sq )
.val( function(msg){
    console.log( msg );           // Hello World
} )
```

`seq(..)` can either accept a sequence itself, as shown here, or a function. If a function, it's expected that the function when called will return a sequence, so the preceding code could have been done with:

```
// ...
.seq( function(){
    return sq;
} )
// ...
```

Also, that step could instead have been accomplished with a `pipe(..)` :

```
// ...
.then( function(done){
    // pipe `sq` into the `done` continuation callback
    sq.pipe( done );
} )
// ...
```

When a sequence is subsumed, both its success message stream and its error stream are piped in.

**Note:** As mentioned in an earlier note, piping (manually with `pipe(..)` or automatically with `seq(..)`) opts the source sequence out of error-reporting, but doesn't affect the error reporting status of the target sequence.

## Value and Error Sequences

If any step of a sequence is just a normal value, that value is just mapped to that step's completion message:

```
var sq = ASQ( 42 );

sq.val( function(msg){
  console.log( msg );           // 42
} );
```

If you want to make a sequence that's automatically errored:

```
var sq = ASQ.failed( "Oops" );

ASQ()
.seq( sq )
.val( function(msg){
  // won't get here
} )
.or( function(err){
  console.log( err );          // Oops
} );
```

You also may want to automatically create a delayed-value or a delayed-error sequence. Using the `after` and `failAfter` contrib plug-ins, this is easy:

```
var sq1 = ASQ.after( 100, "Hello", "World" );
var sq2 = ASQ.failAfter( 100, "Oops" );

sq1.val( function(msg1,msg2){
  console.log( msg1, msg2 );      // Hello World
} );

sq2.or( function(err){
  console.log( err );            // Oops
} );
```

You can also insert a delay in the middle of a sequence using `after(..)`:

```
ASQ( 42 )
// insert a delay into the sequence
.after( 100 )
.val( function(msg){
  console.log( msg );           // 42
} );
```

## Promises and Callbacks

I think *asynquence* sequences provide a lot of value on top of native Promises, and for the most part you'll find it more pleasant and more powerful to work at that level of abstraction. However, integrating *asynquence* with other non-*asynquence* code will be a reality.

You can easily subsume a promise (e.g., `thenable` -- see Chapter 3) into a sequence using the `promise(..)` instance method:

```
var p = Promise.resolve( 42 );

ASQ()
.promise( p )           // could also: `function(){ return p; }`
```

```
.val( function(msg){
  console.log( msg );    // 42
} );
```

And to go the opposite direction and fork/vend a promise from a sequence at a certain step, use the `toPromise` contrib plug-in:

```
var sq = ASQ.after( 100, "Hello World" );

sq.toPromise()
// this is a standard promise chain now
.then( function(msg){
  return msg.toUpperCase();
} )
.then( function(msg){
  console.log( msg );      // HELLO WORLD
} );
```

To adapt *asynquence* to systems using callbacks, there are several helper facilities. To automatically generate an "error-first style" callback from your sequence to wire into a callback-oriented utility, use `errfcb`:

```
var sq = ASQ( function(done){
  // note: expecting "error-first style" callback
  someAsyncFuncWithCB( 1, 2, done.errfcb )
} )
.val( function(msg){
  // ..
} )
.or( function(err){
  // ..
} );

// note: expecting "error-first style" callback
anotherAsyncFuncWithCB( 1, 2, sq.errfcb() );
```

You also may want to create a sequence-wrapped version of a utility -- compare to "promisory" in Chapter 3 and "thunkory" in Chapter 4 -- and *asynquence* provides `ASQ.wrap(...)` for that purpose:

```
var coolUtility = ASQ.wrap( someAsyncFuncWithCB );

coolUtility( 1, 2 )
.val( function(msg){
  // ..
} )
.or( function(err){
  // ..
} );
```

**Note:** For the sake of clarity (and for fun!), let's coin yet another term, for a sequence-producing function that comes from `ASQ.wrap(...)`, like `coolutility` here. I propose "sequory" ("sequence" + "factory").

## Iterable Sequences

The normal paradigm for a sequence is that each step is responsible for completing itself, which is what advances the sequence. Promises work the same way.

The unfortunate part is that sometimes you need external control over a Promise/step, which leads to awkward "capability extraction".

Consider this Promises example:

```

var domready = new Promise( function(resolve,reject){
    // don't want to put this here, because
    // it belongs logically in another part
    // of the code
    document.addEventListener( "DOMContentLoaded", resolve );
} );

// ...

domready.then( function(){
    // DOM is ready!
} );

```

The "capability extraction" anti-pattern with Promises looks like this:

```

var ready;

var domready = new Promise( function(resolve,reject){
    // extract the `resolve()` capability
    ready = resolve;
} );

// ...

domready.then( function(){
    // DOM is ready!
} );

// ...

document.addEventListener( "DOMContentLoaded", ready );

```

**Note:** This anti-pattern is an awkward code smell, in my opinion, but some developers like it, for reasons I can't grasp.

*asynquence* offers an inverted sequence type I call "iterable sequences", which externalizes the control capability (it's quite useful in use cases like the `domready`):

```

// note: `domready` here is an *iterator* that
// controls the sequence
var domready = ASQ iterable();

// ...

domready.val( function(){
    // DOM is ready
} );

// ...

document.addEventListener( "DOMContentLoaded", domready.next );

```

There's more to iterable sequences than what we see in this scenario. We'll come back to them in Appendix B.

## Running Generators

In Chapter 4, we derived a utility called `run(..)` which can run generators to completion, listening for `yield`ed Promises and using them to `async` resume the generator. *asynquence* has just such a utility built in, called `runner(..)`.

Let's first set up some helpers for illustration:

```

function doublePr(x) {
  return new Promise( function(resolve,reject){
    setTimeout( function(){
      resolve( x * 2 );
    }, 100 );
  } );
}

function doubleSeq(x) {
  return ASQ( function(done){
    setTimeout( function(){
      done( x * 2 )
    }, 100 );
  } );
}

```

Now, we can use `runner(..)` as a step in the middle of a sequence:

```

ASQ( 10, 11 )
  .runner( function*(token){
    var x = token.messages[0] + token.messages[1];

    // yield a real promise
    x = yield doublePr( x );

    // yield a sequence
    x = yield doubleSeq( x );

    return x;
  } )
  .val( function(msg){
    console.log( msg );           // 84
  } );

```

## Wrapped Generators

You can also create a self-packaged generator -- that is, a normal function that runs your specified generator and returns a sequence for its completion -- by `ASQ.wrap(..)` ing it:

```

var foo = ASQ.wrap( function*(token){
  var x = token.messages[0] + token.messages[1];

  // yield a real promise
  x = yield doublePr( x );

  // yield a sequence
  x = yield doubleSeq( x );

  return x;
}, { gen: true } );
// ..

foo( 8, 9 )
  .val( function(msg){
    console.log( msg );           // 68
  } );

```

There's a lot more awesome that `runner(..)` is capable of, but we'll come back to that in Appendix B.

## Review

*asynquence* is a simple abstraction -- a sequence is a series of (async) steps -- on top of Promises, aimed at making working with various asynchronous patterns much easier, without any compromise in capability.

There are other goodies in the *asynquence* core API and its contrib plug-ins beyond what we saw in this appendix, but we'll leave that as an exercise for the reader to go check the rest of the capabilities out.

You've now seen the essence and spirit of *asynquence*. The key take away is that a sequence is comprised of steps, and those steps can be any of dozens of different variations on Promises, or they can be a generator-run, or... The choice is up to you, you have all the freedom to weave together whatever async flow control logic is appropriate for your tasks. No more library switching to catch different async patterns.

If these *asynquence* snippets have made sense to you, you're now pretty well up to speed on the library; it doesn't take that much to learn, actually!

If you're still a little fuzzy on how it works (or why!), you'll want to spend a little more time examining the previous examples and playing around with *asynquence* yourself, before going on to the next appendix. Appendix B will push *asynquence* into several more advanced and powerful async patterns.

# You Don't Know JS Yet: Sync & Async - 2nd Edition

## Appendix B: Advanced Async Patterns

**NOTE:**

Work in progress

**NOTE:**

Everything below here is previous text from 1st edition, and is only here for reference while 2nd edition work is underway.  
**Please ignore this stuff.**

Appendix A introduced the *asynquence* library for sequence-oriented async flow control, primarily based on Promises and generators.

Now we'll explore other advanced asynchronous patterns built on top of that existing understanding and functionality, and see how *asynquence* makes those sophisticated async techniques easy to mix and match in our programs without needing lots of separate libraries.

## Iterable Sequences

We introduced *asynquence*'s iterable sequences in the previous appendix, but we want to revisit them in more detail.

To refresh, recall:

```
var domready = ASQ.iterable();

// ...

domready.val( function(){
    // DOM is ready
} );

// ...

document.addEventListener( "DOMContentLoaded", domready.next );
```

Now, let's define a sequence of multiple steps as an iterable sequence:

```

var steps = ASQ iterable();

steps
  .then( function STEP1(x){
    return x * 2;
  })
  .then( function STEP2(x){
    return x + 3;
  })
  .then( function STEP3(x){
    return x * 4;
  });

steps.next( 8 ).value;      // 16
steps.next( 16 ).value;    // 19
steps.next( 19 ).value;    // 76
steps.next().done;         // true

```

As you can see, an iterable sequence is a standard-compliant *iterator* (see Chapter 4). So, it can be iterated with an ES6 `for..of` loop, just like a generator (or any other *iterable*) can:

```

var steps = ASQ iterable();

steps
  .then( function STEP1(){ return 2; } )
  .then( function STEP2(){ return 4; } )
  .then( function STEP3(){ return 6; } )
  .then( function STEP4(){ return 8; } )
  .then( function STEP5(){ return 10; } );

for (var v of steps) {
  console.log( v );
}
// 2 4 6 8 10

```

Beyond the event triggering example shown in the previous appendix, iterable sequences are interesting because in essence they can be seen as a stand-in for generators or Promise chains, but with even more flexibility.

Consider a multiple Ajax request example -- we've seen the same scenario in Chapters 3 and 4, both as a Promise chain and as a generator, respectively -- expressed as an iterable sequence:

```

// sequence-aware ajax
var request = ASQ.wrap( ajax );

ASQ( "http://some.url.1" )
  .runner(
    ASQ iterable()

    .then( function STEP1(token){
      var url = token.messages[0];
      return request( url );
    })

    .then( function STEP2(resp){
      return ASQ().gate(
        request( "http://some.url.2/?v=" + resp ),
        request( "http://some.url.3/?v=" + resp )
      );
    })

    .then( function STEP3(r1,r2){ return r1 + r2; } )
  )
  .val( function(msg){
    console.log( msg );
  })

```

```
});
```

The iterable sequence expresses a sequential series of (sync or async) steps that looks awfully similar to a Promise chain -- in other words, it's much cleaner looking than just plain nested callbacks, but not quite as nice as the `yield`-based sequential syntax of generators.

But we pass the iterable sequence into `ASQ#runner(...)`, which runs it to completion the same as if it was a generator. The fact that an iterable sequence behaves essentially the same as a generator is notable for a couple of reasons.

First, iterable sequences are kind of a pre-ES6 equivalent to a certain subset of ES6 generators, which means you can either author them directly (to run anywhere), or you can author ES6 generators and transpile/convert them to iterable sequences (or Promise chains for that matter!).

Thinking of an async-run-to-completion generator as just syntactic sugar for a Promise chain is an important recognition of their isomorphic relationship.

Before we move on, we should note that the previous snippet could have been expressed in *asynquence* as:

```
ASQ( "http://some.url.1" )
  .seq( /*STEP 1*/ request )
  .seq( function STEP2(resp){
    return ASQ().gate(
      request( "http://some.url.2/?v=" + resp ),
      request( "http://some.url.3/?v=" + resp )
    );
  })
  .val( function STEP3(r1,r2){ return r1 + r2; } )
  .val( function(msg){
    console.log( msg );
  });
}
```

Moreover, step 2 could have even been expressed as:

```
.gate(
  function STEP2a(done,resp) {
    request( "http://some.url.2/?v=" + resp )
    .pipe( done );
  },
  function STEP2b(done,resp) {
    request( "http://some.url.3/?v=" + resp )
    .pipe( done );
  }
)
```

So, why would we go to the trouble of expressing our flow control as an iterable sequence in a `ASQ#runner(...)` step, when it seems like a simpler/flatter *asynquence* chain does the job well?

Because the iterable sequence form has an important trick up its sleeve that gives us more capability. Read on.

## Extending Iterable Sequences

Generators, normal *asynquence* sequences, and Promise chains, are all **eagerly evaluated** -- whatever flow control is expressed initially is the fixed flow that will be followed.

However, iterable sequences are **lazily evaluated**, which means that during execution of the iterable sequence, you can extend the sequence with more steps if desired.

**Note:** You can only append to the end of an iterable sequence, not inject into the middle of the sequence.

Let's first look at a simpler (synchronous) example of that capability to get familiar with it:

```

function double(x) {
  x *= 2;

  // should we keep extending?
  if (x < 500) {
    isq.then( double );
  }

  return x;
}

// setup single-step iterable sequence
var isq = ASQ iterable().then( double );

for (var v = 10, ret;
     (ret = isq.next( v )) && !ret.done;
  ) {
  v = ret.value;
  console.log( v );
}

```

The iterable sequence starts out with only one defined step (`isq.then(double)`), but the sequence keeps extending itself under certain conditions (`x < 500`). Both *asynquence* sequences and Promise chains technically *can* do something similar, but we'll see in a little bit why their capability is insufficient.

Though this example is rather trivial and could otherwise be expressed with a `while` loop in a generator, we'll consider more sophisticated cases.

For instance, you could examine the response from an Ajax request and if it indicates that more data is needed, you conditionally insert more steps into the iterable sequence to make the additional request(s). Or you could conditionally add a value-formatting step to the end of your Ajax handling.

Consider:

```

var steps = ASQ iterable()

.then( function STEP1(token){
  var url = token.messages[0].url;

  // was an additional formatting step provided?
  if (token.messages[0].format) {
    steps.then( token.messages[0].format );
  }

  return request( url );
} )

.then( function STEP2(resp){
  // add another Ajax request to the sequence?
  if (/x1/.test( resp )) {
    steps.then( function STEP5(text){
      return request(
        "http://some.url.4/?v=" + text
      );
    });
  }
}

return ASQ().gate(
  request( "http://some.url.2/?v=" + resp ),
  request( "http://some.url.3/?v=" + resp )
);
}

.then( function STEP3(r1,r2){ return r1 + r2; } );

```

You can see in two different places where we conditionally extend `steps` with `steps.then(..)`. And to run this `steps` iterable sequence, we just wire it into our main program flow with an *asynquence* sequence (called `main` here) using

```
ASQ#runner(..) :
```

```
var main = ASQ( {
  url: "http://some.url.1",
  format: function STEP4(text){
    return text.toUpperCase();
  }
} )
.runner( steps )
.val( function(msg){
  console.log( msg );
} );
```

Can the flexibility (conditional behavior) of the `steps` iterable sequence be expressed with a generator? Kind of, but we have to rearrange the logic in a slightly awkward way:

```
function *steps(token) {
  // **STEP 1**
  var resp = yield request( token.messages[0].url );

  // **STEP 2**
  var rvals = yield ASQ().gate(
    request( "http://some.url.2/?v=" + resp ),
    request( "http://some.url.3/?v=" + resp )
  );

  // **STEP 3**
  var text = rvals[0] + rvals[1];

  // **STEP 4**
  // was an additional formatting step provided?
  if (token.messages[0].format) {
    text = yield token.messages[0].format( text );
  }

  // **STEP 5**
  // need another Ajax request added to the sequence?
  if (/foobar/.test( resp )) {
    text = yield request(
      "http://some.url.4/?v=" + text
    );
  }

  return text;
}

// note: `*steps()` can be run by the same `ASQ` sequence
// as `steps` was previously
```

Setting aside the already identified benefits of the sequential, synchronous-looking syntax of generators (see Chapter 4), the `steps` logic had to be reordered in the `*steps()` generator form, to fake the dynamicism of the extendable iterable sequence `steps`.

What about expressing the functionality with Promises or sequences, though? You *can* do something like this:

```
var steps = something( .. )
.then( .. )
.then( function(..){
  // ..

  // extending the chain, right?
```

```

steps = steps.then( ... );

// ...
})
.then( ... );

```

The problem is subtle but important to grasp. So, consider trying to wire up our `steps` Promise chain into our main program flow -- this time expressed with Promises instead of *asynquence*:

```

var main = Promise.resolve( {
  url: "http://some.url.1",
  format: function STEP4(text){
    return text.toUpperCase();
  }
} )
.then( function(..){
  return steps;           // hint!
} )
.val( function(msg){
  console.log( msg );
} );

```

Can you spot the problem now? Look closely!

There's a race condition for sequence `steps` ordering. When you `return steps`, at that moment `steps` *might* be the originally defined promise chain, or it might now point to the extended promise chain via the `steps = steps.then(..)` call, depending on what order things happen.

Here are the two possible outcomes:

- If `steps` is still the original promise chain, once it's later "extended" by `steps = steps.then(..)`, that extended promise on the end of the chain is **not** considered by the `main` flow, as it's already tapped the `steps` chain. This is the unfortunately limiting **eager evaluation**.
- If `steps` is already the extended promise chain, it works as we expect in that the extended promise is what `main` taps.

Other than the obvious fact that a race condition is intolerable, the first case is the concern; it illustrates **eager evaluation** of the promise chain. By contrast, we easily extended the iterable sequence without such issues, because iterable sequences are **lazily evaluated**.

The more dynamic you need your flow control, the more iterable sequences will shine.

**Tip:** Check out more information and examples of iterable sequences on the *asynquence* site (<https://github.com/getify/asynquence/blob/master/README.md#iterable-sequences>).

## Event Reactive

It should be obvious from (at least!) Chapter 3 that Promises are a very powerful tool in your async toolbox. But one thing that's clearly lacking is in their capability to handle streams of events, as a Promise can only be resolved once. And frankly, this exact same weakness is true of plain *asynquence* sequences, as well.

Consider a scenario where you want to fire off a series of steps every time a certain event is fired. A single Promise or sequence cannot represent all occurrences of that event. So, you have to create a whole new Promise chain (or sequence) for *each* event occurrence, such as:

```

listener.on( "foobar", function(data){

  // create a new event handling promise chain
  new Promise( function(resolve,reject){
    // ...
  })
  .then( ... );
}

```

```

    } )
    .then( ... )
    .then( ... );
}

);

```

The base functionality we need is present in this approach, but it's far from a desirable way to express our intended logic. There are two separate capabilities conflated in this paradigm: the event listening, and responding to the event; separation of concerns would implore us to separate out these capabilities.

The carefully observant reader will see this problem as somewhat symmetrical to the problems we detailed with callbacks in Chapter 2; it's kind of an inversion of control problem.

Imagine uninverting this paradigm, like so:

```

var observable = listener.on( "foobar" );

// later
observable
.then( ... )
.then( ... );

// elsewhere
observable
.then( ... )
.then( ... );

```

The `observable` value is not exactly a Promise, but you can *observe* it much like you can observe a Promise, so it's closely related. In fact, it can be observed many times, and it will send out notifications every time its event ( "foobar" ) occurs.

**Tip:** This pattern I've just illustrated is a **massive simplification** of the concepts and motivations behind reactive programming (aka RP), which has been implemented/expounded upon by several great projects and languages. A variation on RP is functional reactive programming (FRP), which refers to applying functional programming techniques (immutability, referential integrity, etc.) to streams of data. "Reactive" refers to spreading this functionality out over time in response to events. The interested reader should consider studying "Reactive Observables" in the fantastic "Reactive Extensions" library ("RxJS" for JavaScript) by Microsoft (<http://rxjs.codeplex.com/>); it's much more sophisticated and powerful than I've just shown. Also, Andre Staltz has an excellent write-up (<https://gist.github.com/staltz/868e7e9bc2a7b8c1f754>) that pragmatically lays out RP in concrete examples.

## ES7 Observables

At the time of this writing, there's an early ES7 proposal for a new data type called "Observable" (<https://github.com/jhusain/asyncgenerator#introducing-observable>), which in spirit is similar to what we've laid out here, but is definitely more sophisticated.

The notion of this kind of Observable is that the way you "subscribe" to the events from a stream is to pass in a generator -- actually the `iterator` is the interested party -- whose `next(..)` method will be called for each event.

You could imagine it sort of like this:

```

// `someEventStream` is a stream of events, like from
// mouse clicks, and the like.

var observer = new Observer( someEventStream, function*(){
    while (var evt = yield) {
        console.log( evt );
    }
});

```

The generator you pass in will `yield` pause the `while` loop waiting for the next event. The `iterator` attached to the generator instance will have its `next(..)` called each time `someEventStream` has a new event published, and so that event data will resume your generator/`iterator` with the `evt` data.

In the subscription to events functionality here, it's the `iterator` part that matters, not the generator. So conceptually you could pass in practically any iterable, including `ASQ.observable()` iterable sequences.

Interestingly, there are also proposed adapters to make it easy to construct Observables from certain types of streams, such as `fromEvent(..)` for DOM events. If you look at a suggested implementation of `fromEvent(..)` in the earlier linked ES7 proposal, it looks an awful lot like the `ASQ.react(..)` we'll see in the next section.

Of course, these are all early proposals, so what shakes out may very well look/behave differently than shown here. But it's exciting to see the early alignments of concepts across different libraries and language proposals!

## Reactive Sequences

With that crazy brief summary of Observables (and F/RP) as our inspiration and motivation, I will now illustrate an adaptation of a small subset of "Reactive Observables," which I call "Reactive Sequences."

First, let's start with how to create an Observable, using an *asynquence* plug-in utility called `react(..)` :

```
var observable = ASQ.react( function setup(next){
  listener.on( "foobar", next );
});
```

Now, let's see how to define a sequence that "reacts" -- in F/RP, this is typically called "subscribing" -- to that `observable` :

```
observable
  .seq( ... )
  .then( ... )
  .val( ... );
```

So, you just define the sequence by chaining off the Observable. That's easy, huh?

In F/RP, the stream of events typically channels through a set of functional transforms, like `scan(..)`, `map(..)`, `reduce(..)`, and so on. With reactive sequences, each event channels through a new instance of the sequence. Let's look at a more concrete example:

```
ASQ.react( function setup(next){
  document.getElementById( "mybtn" )
    .addEventListener( "click", next, false );
}
  .seq( function(evt){
    var btnID = evt.target.id;
    return request(
      "http://some.url.1/?id=" + btnID
    );
}
  .val( function(text){
    console.log( text );
});
```

The "reactive" portion of the reactive sequence comes from assigning one or more event handlers to invoke the event trigger (calling `next(..)`).

The "sequence" portion of the reactive sequence is exactly like the sequences we've already explored: each step can be whatever asynchronous technique makes sense, from continuation callback to Promise to generator.

Once you set up a reactive sequence, it will continue to initiate instances of the sequence as long as the events keep firing. If you want to stop a reactive sequence, you can call `stop()`.

If a reactive sequence is `stop()`'d, you likely want the event handler(s) to be unregistered as well; you can register a teardown handler for this purpose:

```
var sq = ASQ.react( function setup(next,registerTeardown){
  var btn = document.getElementById( "mybtn" );

  btn.addEventListener( "click", next, false );

  // will be called once `sq.stop()` is called
  registerTeardown( function(){
    btn.removeEventListener( "click", next, false );
  } );
} )
.seq( ... )
.then( ... )
.val( ... );

// later
sq.stop();
```

**Note:** The `this` binding reference inside the `setup(..)` handler is the same `sq` reactive sequence, so you can use the `this` reference to add to the reactive sequence definition, call methods like `stop()`, and so on.

Here's an example from the Node.js world, using reactive sequences to handle incoming HTTP requests:

```
var server = http.createServer();
server.listen(8000);

// reactive observer
var request = ASQ.react( function setup(next,registerTeardown){
  server.addListener( "request", next );
  server.addListener( "close", this.stop );

  registerTeardown( function(){
    server.removeListener( "request", next );
    server.removeListener( "close", request.stop );
  } );
});

// respond to requests
request
.seq( pullFromDatabase )
.val( function(data,res){
  res.end( data );
} );

// node teardown
process.on( "SIGINT", request.stop );
```

The `next(..)` trigger can also adapt to node streams easily, using `onStream(..)` and `unStream(..)`:

```
ASQ.react( function setup(next){
  var fstream = fs.createReadStream( "/some/file" );

  // pipe the stream's "data" event to `next(..)`
  next.onStream( fstream );

  // listen for the end of the stream
  fstream.on( "end", function(){
    next.unStream( fstream );
  } );
```

```

} )
.seq( ... )
.then( ... )
.val( ... );

```

You can also use sequence combinations to compose multiple reactive sequence streams:

```

var sq1 = ASQ.react( ... ).seq( ... ).then( ... );
var sq2 = ASQ.react( ... ).seq( ... ).then( ... );

var sq3 = ASQ.react( ... )
.gate(
  sq1,
  sq2
)
.then( ... );

```

The main takeaway is that `ASQ.react(..)` is a lightweight adaptation of F/RP concepts, enabling the wiring of an event stream to a sequence, hence the term "reactive sequence." Reactive sequences are generally capable enough for basic reactive uses.

**Note:** Here's an example of using `ASQ.react(..)` in managing UI state (<http://jsbin.com/rozipaki/6/edit?js,output>), and another example of handling HTTP request/response streams with `ASQ.react(..)` (<https://gist.github.com/getify/bba5ec0de9d6047b720e>).

## Generator Coroutine

Hopefully Chapter 4 helped you get pretty familiar with ES6 generators. In particular, we want to revisit the "Generator Concurrency" discussion, and push it even further.

We imagined a `runAll(..)` utility that could take two or more generators and run them concurrently, letting them cooperatively `yield` control from one to the next, with optional message passing.

In addition to being able to run a single generator to completion, the `ASQ#runner(..)` we discussed in Appendix A is a similar implementation of the concepts of `runAll(..)`, which can run multiple generators concurrently to completion.

So let's see how we can implement the concurrent Ajax scenario from Chapter 4:

```

ASQ(
  "http://some.url.2"
)
.runner(
  function*(token){
    // transfer control
    yield token;

    var url1 = token.messages[0]; // "http://some.url.1"

    // clear out messages to start fresh
    token.messages = [];

    var p1 = request( url1 );

    // transfer control
    yield token;

    token.messages.push( yield p1 );
  },
  function*(token){
    var url2 = token.messages[0]; // "http://some.url.2"

    // message pass and transfer control
  }
)

```

```

    token.messages[0] = "http://some.url.1";
    yield token;

    var p2 = request( url2 );

    // transfer control
    yield token;

    token.messages.push( yield p2 );

    // pass along results to next sequence step
    return token.messages;
}

)

.val( function(res){
    // `res[0]` comes from "http://some.url.1"
    // `res[1]` comes from "http://some.url.2"
} );

```

The main differences between `ASQ#runner(..)` and `runAll(..)` are as follows:

- Each generator (coroutine) is provided an argument we call `token`, which is the special value to `yield` when you want to explicitly transfer control to the next coroutine.
- `token.messages` is an array that holds any messages passed in from the previous sequence step. It's also a data structure that you can use to share messages between coroutines.
- `yield`ing a Promise (or sequence) value does not transfer control, but instead pauses the coroutine processing until that value is ready.
- The last `return`ed or `yield`ed value from the coroutine processing run will be forward passed to the next step in the sequence.

It's also easy to layer helpers on top of the base `ASQ#runner(..)` functionality to suit different uses.

## State Machines

One example that may be familiar to many programmers is state machines. You can, with the help of a simple cosmetic utility, create an easy-to-express state machine processor.

Let's imagine such a utility. We'll call it `state(..)`, and will pass it two arguments: a state value and a generator that handles that state. `state(..)` will do the dirty work of creating and returning an adapter generator to pass to `ASQ#runner(..)`.

Consider:

```

function state(val,handler) {
    // make a coroutine handler for this state
    return function*(token) {
        // state transition handler
        function transition(to) {
            token.messages[0] = to;
        }

        // set initial state (if none set yet)
        if (token.messages.length < 1) {
            token.messages[0] = val;
        }

        // keep going until final state (false) is reached
        while (token.messages[0] !== false) {
            // current state matches this handler?
            if (token.messages[0] === val) {
                // delegate to state handler
                yield *handler( transition );
            }
        }
    }
}

```

```

        // transfer control to another state handler?
        if (token.messages[0] !== false) {
            yield token;
        }
    };
}

```

If you look closely, you'll see that `state(...)` returns back a generator that accepts a `token`, and then it sets up a `while` loop that will run until the state machine reaches its final state (which we arbitrarily pick as the `false` value); that's exactly the kind of generator we want to pass to `ASQ#runner(...)`!

We also arbitrarily reserve the `token.messages[0]` slot as the place where the current state of our state machine will be tracked, which means we can even seed the initial state as the value passed in from the previous step in the sequence.

How do we use the `state(...)` helper along with `ASQ#runner(...)`?

```

var prevState;

ASQ(
    /* optional: initial state value */
    2
)
// run our state machine
// transitions: 2 -> 3 -> 1 -> 3 -> false
.runner(
    // state `1` handler
    state( 1, function *stateOne(transition){
        console.log( "in state 1" );

        prevState = 1;
        yield transition( 3 );      // goto state `3`
    } ),

    // state `2` handler
    state( 2, function *stateTwo(transition){
        console.log( "in state 2" );

        prevState = 2;
        yield transition( 3 );      // goto state `3`
    } ),

    // state `3` handler
    state( 3, function *stateThree(transition){
        console.log( "in state 3" );

        if (prevState === 2) {
            prevState = 3;
            yield transition( 1 ); // goto state `1`
        }
        // all done!
        else {
            yield "That's all folks!";

            prevState = 3;
            yield transition( false ); // terminal state
        }
    } )
)
// state machine complete, so move on
.val( function(msg){
    console.log( msg );      // That's all folks!
} );

```

It's important to note that the `*stateOne(..)`, `*stateTwo(..)`, and `*stateThree(..)` generators themselves are reinvoked each time that state is entered, and they finish when you `transition(..)` to another value. While not shown here, of course these state generator handlers can be asynchronously paused by `yield`ing Promises/sequences/thunks.

The underneath hidden generators produced by the `state(..)` helper and actually passed to `ASQ#runner(..)` are the ones that continue to run concurrently for the length of the state machine, and each of them handles cooperatively `yield`ing control to the next, and so on.

**Note:** See this "ping pong" example (<http://jsbin.com/qutabu/1/edit?js,output>) for more illustration of using cooperative concurrency with generators driven by `ASQ#runner(..)`.

## Communicating Sequential Processes (CSP)

"Communicating Sequential Processes" (CSP) was first described by C. A. R. Hoare in a 1978 academic paper (<http://dl.acm.org/citation.cfm?doid=359576.359585>), and later in a 1985 book (<http://www.usingcsp.com/>) of the same name. CSP describes a formal method for concurrent "processes" to interact (aka "communicate") during processing.

You may recall that we examined concurrent "processes" back in Chapter 1, so our exploration of CSP here will build upon that understanding.

Like most great concepts in computer science, CSP is heavily steeped in academic formalism, expressed as a process algebra. However, I suspect symbolic algebra theorems won't make much practical difference to the reader, so we will want to find some other way of wrapping our brains around CSP.

I will leave much of the formal description and proof of CSP to Hoare's writing, and to many other fantastic writings since. Instead, we will try to just briefly explain the idea of CSP in as un-academic and hopefully intuitively understandable a way as possible.

### Message Passing

The core principle in CSP is that all communication/interaction between otherwise independent processes must be through formal message passing. Perhaps counter to your expectations, CSP message passing is described as a synchronous action, where the sender process and the receiver process have to mutually be ready for the message to be passed.

How could such synchronous messaging possibly be related to asynchronous programming in JavaScript?

The concreteness of relationship comes from the nature of how ES6 generators are used to produce synchronous-looking actions that under the covers can indeed either be synchronous or (more likely) asynchronous.

In other words, two or more concurrently running generators can appear to synchronously message each other while preserving the fundamental asynchrony of the system because each generator's code is paused (aka "blocked") waiting on resumption of an asynchronous action.

How does this work?

Imagine a generator (aka "process") called "A" that wants to send a message to generator "B." First, "A" `yield`s the message (thus pausing "A") to be sent to "B." When "B" is ready and takes the message, "A" is then resumed (unblocked).

Symmetrically, imagine a generator "A" that wants a message **from** "B." "A" `yield`s its request (thus pausing "A") for the message from "B," and once "B" sends a message, "A" takes the message and is resumed.

One of the more popular expressions of this CSP message passing theory comes from ClojureScript's core.async library, and also from the go language. These take on CSP embody the described communication semantics in a conduit that is opened between processes called a "channel."

**Note:** The term *channel* is used in part because there are modes in which more than one value can be sent at once into the "buffer" of the channel; this is similar to what you may think of as a stream. We won't go into depth about it here, but it can be a very powerful technique for managing streams of data.

In the simplest notion of CSP, a channel that we create between "A" and "B" would have a method called `take(..)` for blocking to receive a value, and a method called `put(..)` for blocking to send a value.

This might look like:

```
var ch = channel();

function *foo() {
  var msg = yield take( ch );

  console.log( msg );
}

function *bar() {
  yield put( ch, "Hello World" );

  console.log( "message sent" );
}

run( foo );
run( bar );
// Hello World
// "message sent"
```

Compare this structured, synchronous(-looking) message passing interaction to the informal and unstructured message sharing that `ASQ#runner(..)` provides through the `token.messages` array and cooperative `yield`ing. In essence, `yield put(..)` is a single operation that both sends the value and pauses execution to transfer control, whereas in earlier examples we did those as separate steps.

Moreover, CSP stresses that you don't really explicitly "transfer control," but rather you design your concurrent routines to block expecting either a value received from the channel, or to block expecting to try to send a message on the channel. The blocking around receiving or sending messages is how you coordinate sequencing of behavior between the coroutines.

**Note:** Fair warning: this pattern is very powerful but it's also a little mind twisting to get used to at first. You will want to practice this a bit to get used to this new way of thinking about coordinating your concurrency.

There are several great libraries that have implemented this flavor of CSP in JavaScript, most notably "js-csp" (<https://github.com/ubolonton/js-csp>), which James Long (<http://twitter.com/jlongster>) forked (<https://github.com/jlongster/js-csp>) and has written extensively about (<http://jlongster.com/Taming-the-Asynchronous-Beast-with-CSP-in-JavaScript>). Also, it cannot be stressed enough how amazing the many writings of David Nolen (<http://twitter.com/swannodette>) are on the topic of adapting ClojureScript's go-style core.async CSP into JS generators (<http://swannodette.github.io/2013/08/24/es6-generators-and-csp>).

## asynquence CSP emulation

Because we've been discussing async patterns here in the context of my *asynquence* library, you might be interested to see that we can fairly easily add an emulation layer on top of `ASQ#runner(..)` generator handling as a nearly perfect porting of the CSP API and behavior. This emulation layer ships as an optional part of the "asynquence-contrib" package alongside *asynquence*.

Very similar to the `state(..)` helper from earlier, `ASQ.csp.go(..)` takes a generator -- in go/core.async terms, it's known as a goroutine -- and adapts it to use with `ASQ#runner(..)` by returning a new generator.

Instead of being passed a `token`, your goroutine receives an initially created channel (`ch` below) that all goroutines in this run will share. You can create more channels (which is often quite helpful!) with `ASQ.csp.chan(..)`.

In CSP, we model all asynchrony in terms of blocking on channel messages, rather than blocking waiting for a Promise/sequence/thunk to complete.

So, instead of `yield`ing the Promise returned from `request(..)`, `request(..)` should return a channel that you `take(..)` a value from. In other words, a single-value channel is roughly equivalent in this context/usage to a Promise/sequence.

Let's first make a channel-aware version of `request(..)`:

```
function request(url) {
  var ch = ASQ.csp.channel();
  ajax( url ).then( function(content){
    // `putAsync(..)` is a version of `put(..)` that
    // can be used outside of a generator. It returns
    // a promise for the operation's completion. We
    // don't use that promise here, but we could if
    // we needed to be notified when the value had
    // been `take(..)`n.
    ASQ.csp.putAsync( ch, content );
  } );
  return ch;
}
```

From Chapter 3, "promisory" is a Promise-producing utility, "thunkory" from Chapter 4 is a thunk-producing utility, and finally, in Appendix A we invented "sequory" for a sequence-producing utility.

Naturally, we need to coin a symmetric term here for a channel-producing utility. So let's unsurprisingly call it a "chanory" ("channel" + "factory"). As an exercise for the reader, try your hand at defining a `channelify(..)` utility similar to `Promise.wrap(..)` / `promisify(..)` (Chapter 3), `thunkify(..)` (Chapter 4), and `ASQ.wrap(..)` (Appendix A).

Now consider the concurrent Ajax example using *asyquence*-flavored CSP:

```
ASQ()
.runner(
  ASQ.csp.go( function*(ch){
    yield ASQ.csp.put( ch, "http://some.url.2" );

    var url1 = yield ASQ.csp.take( ch );
    // "http://some.url.1"

    var res1 = yield ASQ.csp.take( request( url1 ) );

    yield ASQ.csp.put( ch, res1 );
  } ),
  ASQ.csp.go( function*(ch){
    var url2 = yield ASQ.csp.take( ch );
    // "http://some.url.2"

    yield ASQ.csp.put( ch, "http://some.url.1" );

    var res2 = yield ASQ.csp.take( request( url2 ) );
    var res1 = yield ASQ.csp.take( ch );

    // pass along results to next sequence step
    ch.buffer_size = 2;
    ASQ.csp.put( ch, res1 );
    ASQ.csp.put( ch, res2 );
  } )
)
.val( function(res1,res2){
  // `res1` comes from "http://some.url.1"
  // `res2` comes from "http://some.url.2"
} );
```

The message passing that trades the URL strings between the two goroutines is pretty straightforward. The first goroutine makes an Ajax request to the first URL, and that response is put onto the `ch` channel. The second goroutine makes an Ajax request to the second URL, then gets the first response `res1` off the `ch` channel. At that point, both responses `res1` and `res2` are completed and ready.

If there are any remaining values in the `ch` channel at the end of the goroutine run, they will be passed along to the next step in the sequence. So, to pass out message(s) from the final goroutine, `put(..)` them into `ch`. As shown, to avoid the blocking of those final `put(..)`s, we switch `ch` into buffering mode by setting its `buffer_size` to `2` (default: `0`).

**Note:** See many more examples of using *asynquence*-flavored CSP here (<https://gist.github.com/getify/e0d04f1f5aa24b1947ae>).

## Review

Promises and generators provide the foundational building blocks upon which we can build much more sophisticated and capable asynchrony.

*asynquence* has utilities for implementing *iterable sequences*, *reactive sequences* (aka "Observables"), *concurrent coroutines*, and even *CSP goroutines*.

Those patterns, combined with the continuation-callback and Promise capabilities, gives *asynquence* a powerful mix of different asynchronous functionalities, all integrated in one clean async flow control abstraction: the sequence.