

Final Thoughts on the semester

17-313 Fall 2023

Foundations of Software Engineering

<https://cmu-313.github.io>

Andrew Begel and Rohan Padhye and **Michael Hilton**

Administrivia

- Final Exam time and Location: Tuesday, December 12
05:30pm-08:30pm in Person PH 100
- Snack Survey to be released tomorrow

Today- JT Topic Selection:

- I have a set of topics I could present today, but I would like to ask you what you would like to hear about:
 - Grad School
 - ML Fairness Types
 - Dependency Management + SemVer
 - Software Patents
 - Decision Calculus

Grad School

Not All Grad School Programs are created equal

- PhD - Research Focused
- Masters programs - A wide variety of programs with different goals
 - 5th year masters
 - Research Masters
 - Full/Part time Masters
 - Immigration Concerns
 - Breadth vs Depth

The PhD in CS: Getting There and Being Successful

More Great Reasons to Pursue a Ph.D.

Gain Opportunities to Teach and Mentor

Pursuing a Ph.D. provides you with a unique opportunity to teach and mentor

Satisfy Your Intellectual Curiosity

- *Discover new things*
- *Identify new problems*
- *Develop creative solutions*
- *Push the boundary of knowledge*
- *Develop a habit of lifelong learning*



Get Paid to Learn!

Many students don't realize that most C.S. Ph.D. programs pay a comfortable stipend. But note, there are opportunity costs.

The CS Doctorate (Ph.D.) in a Nutshell



Timeline

- Variable; average 6-7 years from Bachelor's
- Depending on school, starting with a Master's degree may shorten timeline



Coursework

- Typically "next level" CS foundations and (more) advanced electives



Research

- Dissertation (aka Doctoral Thesis)
- Oral/Written Exam along the way (e.g., qualifying exam)



Tuition & Stipend (Get paid to learn!)

- Generally ***tuition is waived*** and you typically receive ***a stipend and health insurance*** from a teaching or research assistantship or fellowship



Career Paths

- Academia (e.g., a professor)
- Industrial, NPO, or government researcher or engineer
- Entrepreneur (e.g., a start-up)
- Higher-level development/leadership positions

Typical Timeline for a Ph.D. Program

Foundational coursework to prepare for research.
Join a lab with advisor & initial project.



Year 1



Complete a majority of your coursework.
Take qualifying exam.
Identify research area.
Potentially earn Master's degree "along the way".

Year 2

Obtain preliminary results and publish papers.
Formulate Ph.D. research plan. Identify Ph.D. committee. Begin writing proposal.



Year 3



Complete and defend Ph.D. proposal.
Continue with research and publishing your results. Identify your future career path.

Year 4



Years 5-N

How a Ph.D. differs from an M.S. Degree

Academic Master's Program

3-4 courses/term in first year.

1-2 courses/term with a research and MS thesis or project in second year.

Not always funded, but there is the potential to serve as a teaching assistant.

Professional Master's Program

3-4 courses/term for 1.5– 2 years.

Geared towards industrial careers.

Typically not funded by the school, but could be funded by a company you work for.



Doctoral (Ph.D) Program

Similar to Academic Master's in the first 2 years. In years 3+, primarily research.

Typically includes additional duties such as teaching assistant or research assistant.



The PhD in CS: Getting There and Being Successful

November 1st, 2021

Dr. Michael Hilton
Associate Teaching Professor
Carnegie Mellon University

Ian Ludden
PhD Student
University of Illinois Urbana-Champaign

Dr. Janet Davis
Associate Professor
Whitman College



<https://www.youtube.com/watch?v=Qn7eUeNw6vl>

Masters Programs

MS Diversity - This is just SCS MS Programs in SCS

Language Technologies Institute (LTI)

- [Master of Computational Data Science \(with CSD\)](#)
- [M.S. in Artificial Intelligence and Innovation](#)
- [M.S. in Language Technologies](#)
- [M.S. in Intelligent Information Systems](#)

Machine Learning (ML)

- [M.S. in Machine Learning](#)
- [Fifth Year Master's in Machine Learning](#)

Robotics Institute (RI)

- [M.S. in Computer Vision](#)
- [M.S. in Robotics](#)
- [M.S. in Robotic Systems Development](#)

Software and Societal Systems Department (S3D)

- [Master of Software Engineering](#)
- [Master of Software Engineering Online](#)
- [Master of Software Engineering - Embedded Systems](#)
- [Master of Software Engineering - Scalable Systems](#)
- [M.S. in Information Technology - Privacy Engineering](#)

Computational Biology Department

- [M.S. in Automated Science: Biological Experimentation](#)
- [M.S. in Computational Biology](#)

Computer Science (CSD)

- [M.S. in Computer Science](#)
- [Master of Computational Data Science \(with LTI\)](#)
- [Fifth-Year Master's in Computer Science](#)

Human-Computer Interaction Institute (HCII)

- [Master of Human-Computer Interaction](#)
- [Master of Educational Technology and Applied Learning Science \(with Dietrich College\)](#)
- [M.S. in Product Management \(with Tepper School of Business\)](#)
- [Accelerated Master of Human-Computer Interaction](#)

Some Programs of Interest

- MSE - Masters of Software Engineering
- MSE - AMP
- SCS 5th Year Master's Program
- 5th-Year Master's in ML
- MBA/Part Time

Reasons to Consider a Master's Degree

- Entry into the US Labor Force
- Extension of Student Visa
- Breadth
- Depth
- Credentials
- Curiosity of Learning
- ???

Master's of Software Engineering



MSE Degree Programs

The Masters of Software Engineering (MSE) is five programs:

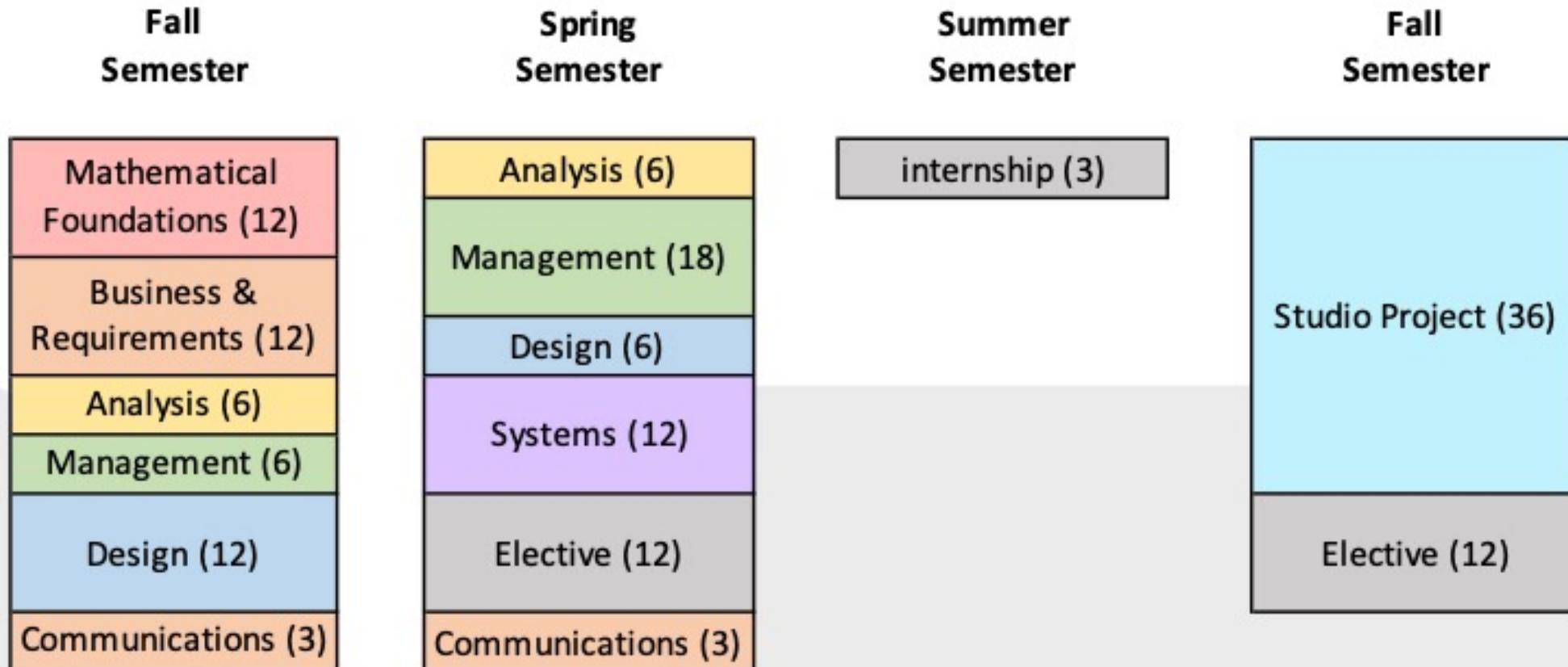
- **MSE for professionals** with >2 years of software engineering experience
 - MSE Campus is a four semester full-time program, graduates in 1.5 years
 - MSE/MBA dual degree program with Tepper School of Business
 - MSE Online is a 6-8 semester part-time program, graduates in 2-3 years
- **MSE for recent grads** with <2 years of related experience, consisting of two programs or tracks:
 - Scalable Systems (SS) – distributed and data-intensive systems
 - Embedded Systems (ES) – robotics, sensors and wearables

	Masters of Science in Computer Science	Masters of Software Engineering Programs
Job Roles	Data Scientist Machine Learning Engineer Product Engineer Research Engineer Software Engineer	Product Manager Software Engineer II Software Engineer III Senior Software Engineer Staff Software Engineer
Top Six Employers	Google Apple Microsoft Amazon Facebook Salesforce	Google Amazon Microsoft Facebook Uber Apple





MSE-SS/-ES – Plan of Study

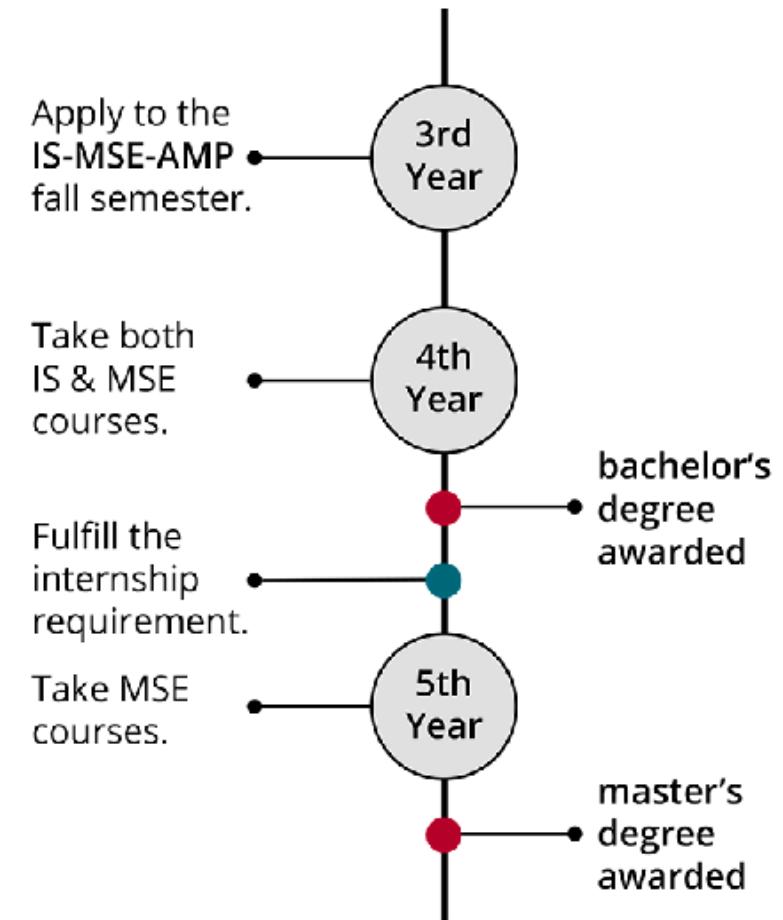


* 60 unit core, 6 units communication, 36 units project, required internship, 24 units electives

IS-MSE Accelerated Master's Program (AMP)

IS-MSE Accelerated Master's Program (AMP)

- 5th year master's program for undergraduate Information Systems majors in the Dietrich College of Humanities and Social Sciences at Carnegie Mellon.
- <https://mse.isri.cmu.edu/applicants/mse-amp.html>
- Application Deadline: December 12, 2022



CSD Fifth-year Master's

CS 5th year's Master

- Fifth Year Master's program usually lasts 12 months, including one normal academic year and one summer
- Students will start working on a research project during the summer after their senior year, and continue that project while taking classes during the academic year.
- Students can only start the Fifth Year Master's program after they have received their B.S. in computer science or A& from CMU.
- “Students typically apply in their senior year”
- <https://csd.cmu.edu/academics/masters/requirements>

M.S. in Computer Science

M.S. in Computer Science

- Pass* 96 units in qualifying master's courses from the curriculum list in the MSCS Handbook. This is typically eight courses.
- Pass 12 free-elective units.
- Pass one course from the available Systems courses. (May be used a qualifying course.)
- Pass one course from the available Theoretical Foundations courses. (May be used a qualifying course.)
- Pass one course from the available Theoretical Foundations courses. (May be used a qualifying course.)
- <https://csd.cmu.edu/academics/masters/requirements>

5th Year Master's Machine Learning

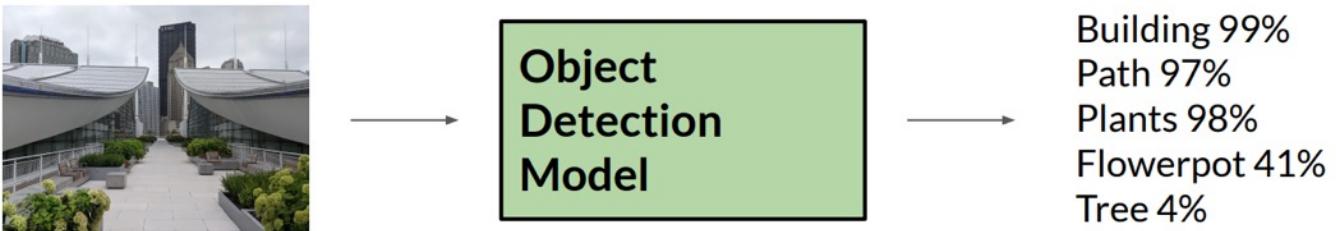
5th-Year Master's in Machine Learning

- The 5th-Year Master's in Machine Learning allows CMU students to complete a MS in Machine Learning in one additional year by taking some of the required courses as an undergraduate.
- Interested students apply earlier in Senior year than the standard application deadline and receive the response earlier as well.
- Must take precisely 3 of the MS courses during their undergraduate years, passed with a B or better. (These courses may also count towards the Bachelor's degree.)
- <https://www.ml.cmu.edu/academics/5th-year-ms.html>

Questions?

ML Fairness

ML Model = Unreliable Function

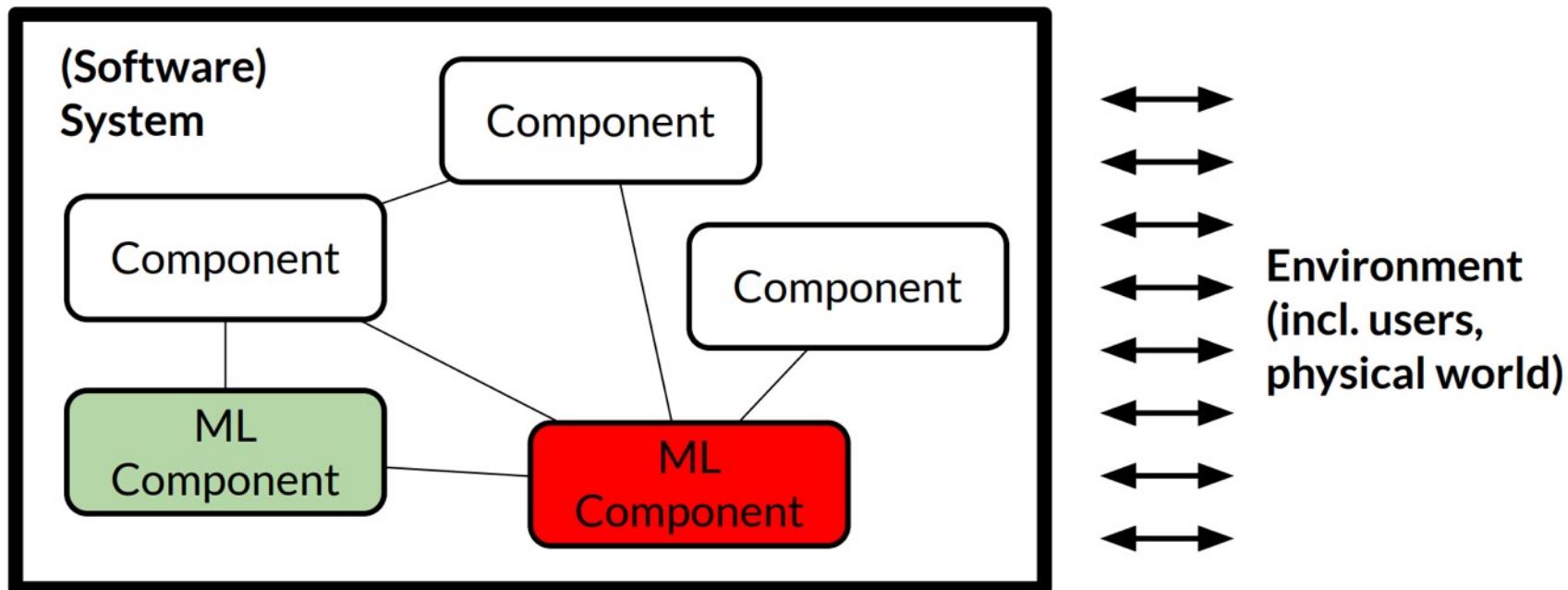


No guarantees, may make mistakes, confidence unreliable

Model often inscrutable, opaque

Evaluated in terms of accuracy, not correctness

What to do when the ML component makes mistake?

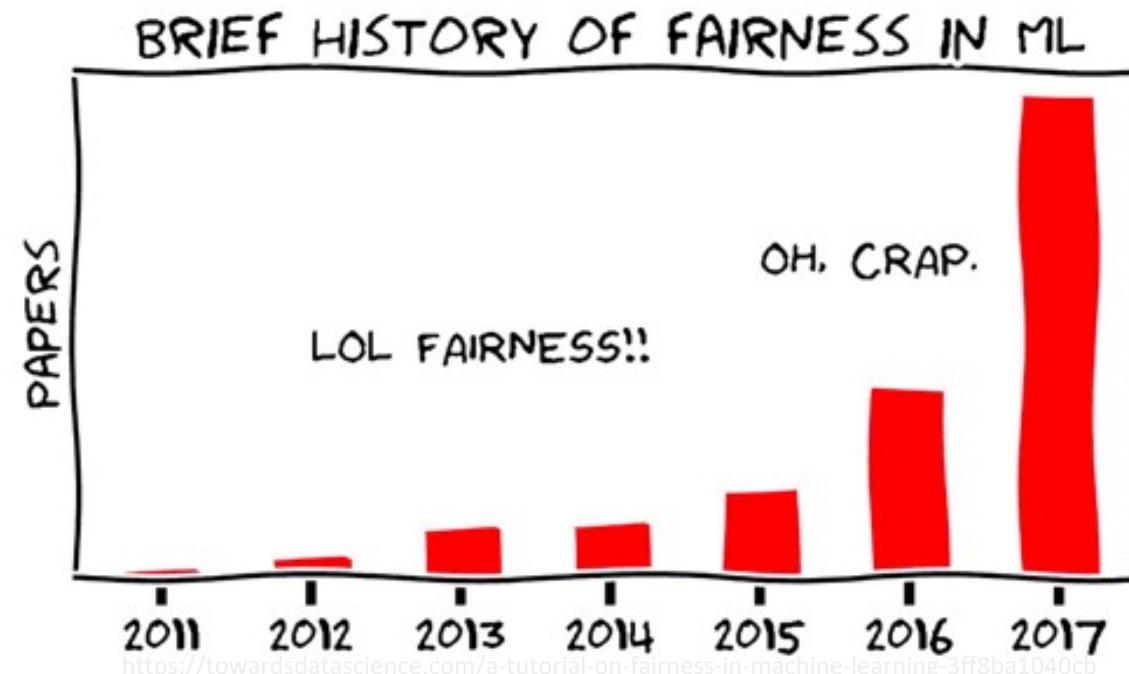


39

Fairness

ML Fairness

- Getting answers is the easy part... Asking the right questions is the hard part.



Perception:



JAKE-CLARK.TUMBLR

Life is often not this simple...



Fairness

- Is a deeply technical topic, but we will discuss it at a higher level of abstraction.
- The formulas are important, but knowing which formula to apply is MUCH more important
- This is a special case of how to test when the desired outcome is hard to measure.

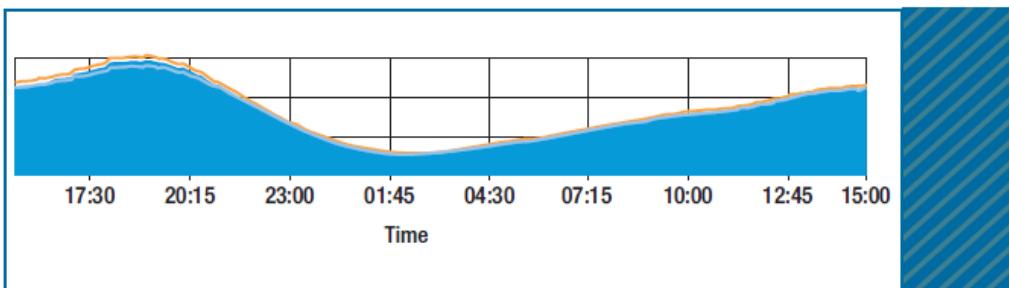


FIGURE 2. A graph of SPS ([stream] starts per second) over a 24-hour period. This metric varies slowly and predictably throughout a day. The orange line shows the trend for the prior week. The y-axis isn't labeled because the data is proprietary.

VS



What does “fair” mean?

What is Fairness?

- Law
 - fairness includes protecting individuals and groups from discrimination or mistreatment with a focus on prohibiting behaviors, biases and basing decisions on certain protected factors or social group categories.
- Social Science
 - “often considers fairness in light of social relationships, power dynamics, institutions and markets.”³ Members of certain groups (or identities) that tend to experience advantages.

What is Fairness? continued

- Quantitative Fields
 - (i.e. math, computer science, statistics, economics): questions of fairness are seen as mathematical problems. Fairness tends to match to some sort of criteria, such as equal or equitable allocation, representation, or error rates, for a particular task or problem.
- Philosophy:
 - ideas of fairness “rest on a sense that what is fair is also what is morally right.” Political philosophy connects fairness to notions of justice and equity.

Fairness as QA

How can we define “fair”

- For the purposes of creating an oracle
- We must have a better definition than infamous 1964 Supreme Court obscenity test:
 - I shall not today attempt further to define [obscene material], and perhaps I could never succeed in intelligibly doing so. But I know it when I see it, and the motion picture involved in this case is not that.[\[1\]](#)

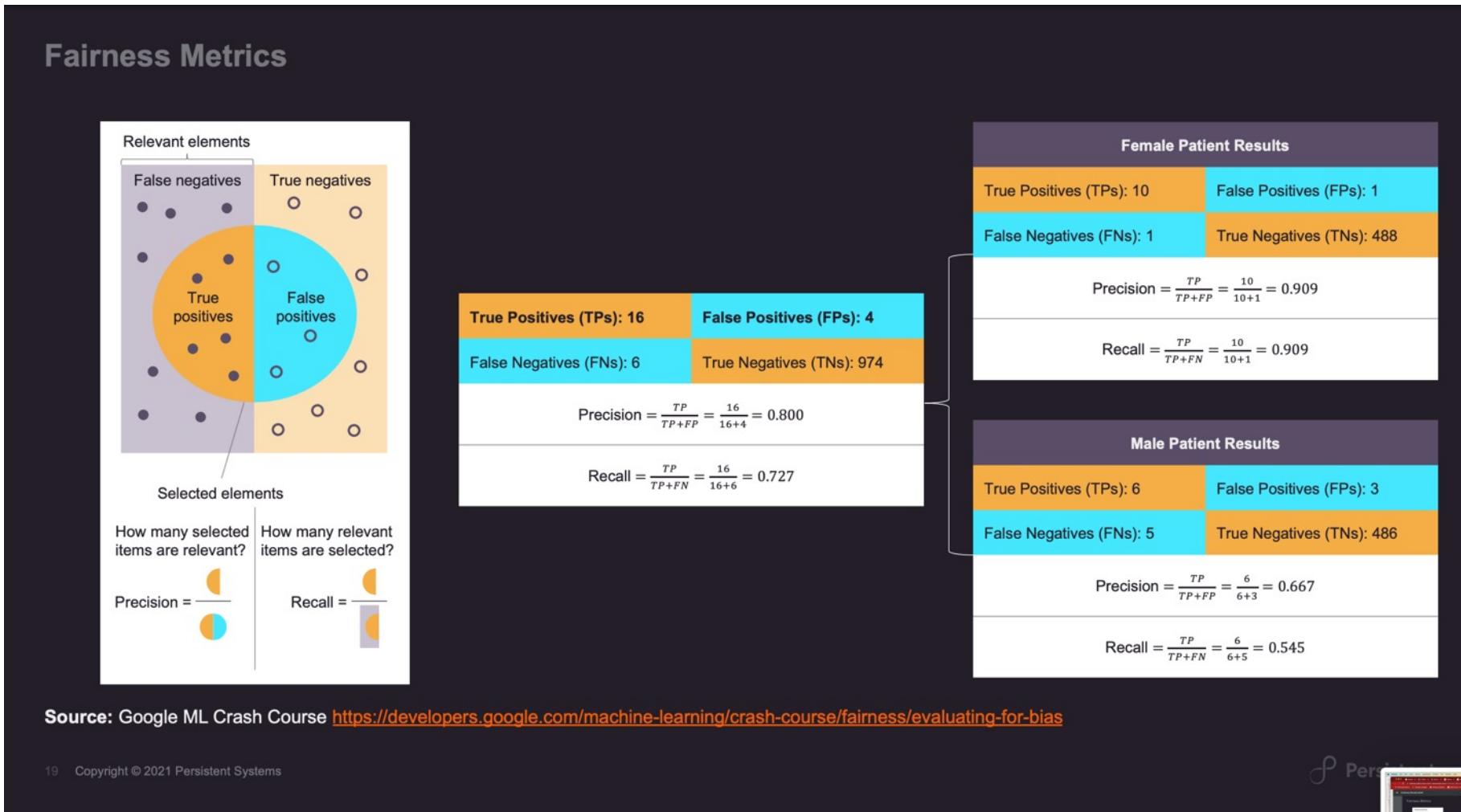
We don't need to start from scratch...

What can we do?

What can we do?

- We can evaluate with different criteria (e.g., different admissions score thresholds).
- We can observe the outcome of changing thresholds, and we can set different thresholds for different groups. (e.g., different SAT scores for in-state or out-of-state admissions)
- We can observe the impact of these different thresholds across a variety of metrics for each group.

First, some definitions:



Varieties of fairness (names vary)

- **Group unaware**
 - Ignore group data (one group could get excluded)
- **Group thresholds**
 - Different rules per group (rules differ by group)
- **Demographic parity**
 - Same percentage in pool as outcomes (might result in random selection)
- **Equal opportunity**
 - Equal chance out positive outcomes regardless of groups (focus on individual, rules differ per group)

Group unaware

- We use some criteria that is independent of the categories we are considering for fairness.
- Guarantees about outcomes: None. One group may be completely excluded

Group thresholds

- We create different criteria per group
- Guarantees about outcomes: candidates inside a group are evaluated by the same standard as others inside the same group.
- By definition, groups are evaluated to a different standard (e.g., different fitness standards by gender in US Military)

Demographic parity

- We create different criteria per group, with a goal of similar outcomes in a certain dimension.
- Guarantees about outcomes: The same percentage of each group will have a positive outcome. e.g., 25 % accepted from group A, 25% accepted from group B.
- However, can result in different true positive rates, (e.g., more “worthy” candidates denied in group A than group B.

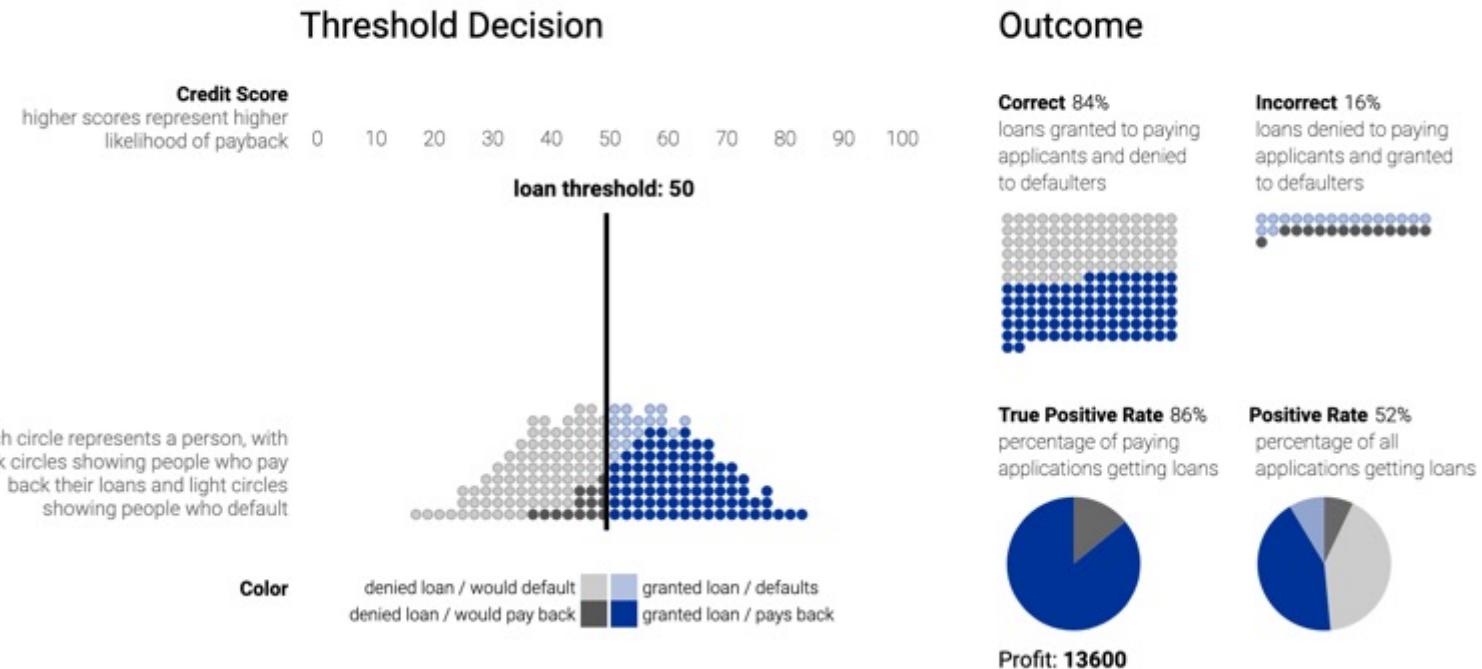
Equal opportunity

- We create different criteria per group, with a goal of similar outcomes for similar individuals across groups.
- Guarantees about outcomes: The same number of true positives per group. e.g., 80% true positives in group A, 80% true positives in group B.
- However, can result in different positive rates across groups.

Explainability

Simulating loan thresholds

Drag the black threshold bars left or right to change the cut-offs for loans.



<https://research.google.com/bigpicture/attacking-discrimination-in-ml/>

Activity

Consider the different approaches to fairness. Can you come up with different scenarios where each fairness approach might or might not be appropriate?

Remember the fairness approaches are:

- Group unaware
- Group thresholds
- Demographic parity
- Equal opportunity

Resources

- Fairness Textbook:
- <https://fairmlbook.org/testing.html>

Dependency Management

Left-pad (March 22, 2016)



The screenshot shows a news article from The Verge. At the top left is the The Verge logo with 'THE VERGE' in white on a black background, followed by 'TECH ▾ REVIEWS ▾'. To the right is a small 'OBSESSIONS' section with the word 'NPM ERR!'. At the top right is the 'QUARTZ' logo. The main title is 'How one programmer broke the internet by deleting a tiny piece of code'. Below the title is a subtitle 'How an irate developer briefly broke JavaScript'. A sub-subtitle reads 'Unpublishing 11 lines of code brought down an open source house of cards'. The author is listed as 'By Paul Miller | @futurepaul | Mar 24, 2016, 4:29pm EDT'. There are social sharing icons for Facebook and Twitter, and a 'SIGN IN' button. The Register logo is visible at the bottom.

How one developer just broke Node, Babel and thousands of projects in 11 lines of JavaScript

Code pulled from NPM – which everyone was using

Left-pad (March 22, 2016)

npmjs.org tells me that left-pad is not available (404 page) #4

 Closed silkentrance opened this issue on Mar 22, 2016 · 193 comments



silkentrance commented on Mar 22, 2016

...

When building projects on travis, or when searching for left-pad on npmjs.com, both will report that the package cannot be found.

Here is an excerpt from the travis build log

```
npm ERR! Linux 3.13.0-40-generic
npm ERR! argv "/home/travis/.nvm/versions/node/v4.2.2/bin/node" "/home/travis/.nvm/versions/node/v4.2.2/bin/npm"
npm ERR! node v4.2.2
npm ERR! npm  v2.14.7
npm ERR! code E404
npm ERR! 404 Registry returned 404 for GET on https://registry.npmjs.org/left-pad
npm ERR! 404
npm ERR! 404 'left-pad' is not in the npm registry.
npm ERR! 404 You should bug the author to publish it (or use the name yourself!)
npm ERR! 404 It was specified as a dependency of 'line-numbers'
npm ERR! 404
npm ERR! 404 Note that you can also install from a
npm ERR! 404 tarball, folder, http url, or git url.
npm ERR! Please include the following file with any support request:
npm ERR!   /home/travis/build/coldrye-es/pingo/npm-debug.log
make: *** [deps] Error 1
```

And here is the standard npmjs.com error page <https://www.npmjs.com/package/left-pad>

However, if I remove left-pad from my local npm cache and then reinstall it using npm it will happily install left-pad@0.0.4.

 88  3

Left-pad (Docs)

left-pad

String left pad

build unknown

Install

```
$ npm install left-pad
```

Usage

```
const leftPad = require('left-pad')

leftPad('foo', 5)
// => " foo"

leftPad('foobar', 6)
// => "foobar"

leftPad(1, 2, '0')
// => "01"

leftPad(17, 5, 0)
// => "00017"
```

Install

```
> npm i left-pad
```

Repository

❖ github.com/stevemao/left-pad

Homepage

🔗 github.com/stevemao/left-pad#readme

Weekly Downloads

2,962,641



Version

1.3.0

License

WTFPL

Unpacked Size

9.75 kB

Total Files

10

Issues

3

Pull Requests

7

Last publish

4 years ago

Left-pad (Source Code)

17 lines (11 sloc) | 222 Bytes

```
1 module.exports = leftpad;
2
3 function leftpad (str, len, ch) {
4     str = String(str);
5
6     var i = -1;
7
8     if (!ch && ch !== 0) ch = ' ';
9
10    len = len - str.length;
11
12    while (++i < len) {
13        str = ch + str;
14    }
15
16    return str;
17 }
```

See also: isArray

5 lines (4 sloc) | 133 Bytes

```
1 var toString = {}.toString;
2
3 module.exports = Array.isArray || function (arr) {
4   return toString.call(arr) === '[object Array]';
5 }
```

isarray

Array#isArray for older browsers and deprecated Node.js versions.

build passing downloads 227M/month



Just use `Array.isArray` directly, unless you need to support those older versions.

Usage

```
var isArray = require('isarray');

console.log(isArray([])); // => true
console.log(isArray({})); // => false
```

Install

```
> npm i isarray
```

Repository

◆ github.com/juliangruber/isarray

Homepage

🔗 github.com/juliangruber/isarray

Weekly Downloads

50,913,317



Version

2.0.5

License

MIT

Unpacked Size

3.43 kB

Total Files

4

Issues

Pull Requests

4

3

Dependency Management

- It's hard
- It's mostly a mess (everywhere)
- But it's critical to modern software development

What is a Dependency?

- Core of what most build systems do
 - “Compile” and “Run Tests” is just a fraction of their job
- Examples: Maven, Gradle, NPM, Bazel, ...
- **Foo->Bar**: To build Foo, you may need to have a built version of Bar
- Dependency Scopes:
 - **Compile**: Foo uses classes, functions, etc. defined by Bar
 - **Runtime**: Foo uses an abstract API whose implementation is provided by Bar (e.g. logging, database, network or other I/O)
 - **Test**: Foo needs Bar only for tests (e.g. JUnit, mocks)
- Internal vs. External Dependencies
 - Is Bar also built/maintained by your org or is it pulled from elsewhere using a package manager?

Dependencies: Example

```
git clone https://github.com/CMU-313/Teedy/blob/main/pom.xml
152 <dependencyManagement>
153   <dependencies>
154     <dependency>
155       <groupId>com.sismics.docs</groupId>
156       <artifactId>docs-core</artifactId>
157       <version>${project.version}</version>
158     </dependency>
159
160     <dependency>
161       <groupId>com.sismics.docs</groupId>
162       <artifactId>docs-web-common</artifactId>
163       <version>${project.version}</version>
164     </dependency>
165
166     <dependency>
167       <groupId>com.sismics.docs</groupId>
168       <artifactId>docs-web-common</artifactId>
169       <type>test-jar</type>
170       <version>${project.version}</version>
171     </dependency>
172
173     <dependency>
174       <groupId>com.sismics.docs</groupId>
175       <artifactId>docs-web</artifactId>
176       <version>${project.version}</version>
177     </dependency>
178
179     <dependency>
180       <groupId>org.eclipse.jetty</groupId>
181       <artifactId>jetty-server</artifactId>
182       <version>${org.eclipse.jetty.jetty-server.version}</version>
183     </dependency>
184
185     <dependency>
186       <groupId>org.eclipse.jetty</groupId>
187       <artifactId>jetty-webapp</artifactId>
188       <version>${org.eclipse.jetty.jetty-webapp.version}</version>
189     </dependency>
```

Package: git (1:2.17.1-1ubuntu0.9)

fast, scalable, distributed revision control system

Other Packages Related to git

- depends ● recommends ■ suggests • enhances
- [git-man](#) (<< 1:2.17.0-. [not amd64, i386]
fast, scalable, distributed revision control system (manual pages)
- [git-man](#) (<< 1:2.17.1-. [amd64, i386]
- [git-man](#) (> 1:2.17.0 [not amd64, i386]
- [git-man](#) (> 1:2.17.1) [amd64, i386]
- [libc6](#) (>= 2.16) [not arm64, ppc64el]
GNU C Library: Shared libraries
also a virtual package provided by [libc6-udeb](#)
- [libc6](#) (>= 2.17) [arm64, ppc64el]
- [libcurl3-gnutls](#) (>= 7.16.2)
easy-to-use client-side URL transfer library (GnuTLS flavour)
- [liberror-perl](#)
Perl module for error/exception handling in an OO-ish way
- [libexpat1](#) (>= 2.0.1)
XML parsing C library - runtime library
- [libpcre3](#)
Old Perl 5 Compatible Regular Expression Library - runtime files
- [perl](#)
Larry Wall's Practical Extraction and Report Language
- [zlib1g](#) (>= 1:1.2.0)
compression library - runtime
- ◆ [less](#)
pager program similar to more
- ◆ [patch](#)
Apply a diff file to an original
- ◆ [ssh-client](#)
virtual package provided by [openssh-client](#)

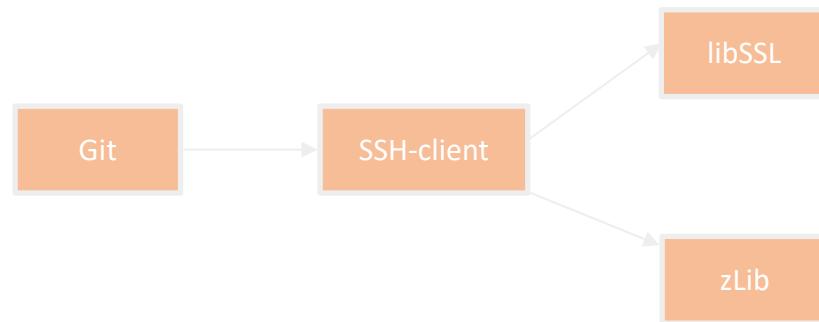
Where are the dependencies hosted?

- Typically downloaded from dependency servers:
 - Maven Central (<https://repo.maven.apache.org/maven2/>)
 - Ubuntu Packages for Apt (<https://packages.ubuntu.com/>)
 - Python Package Index (<https://pypi.org/>)
 - NPM Public Registry (<https://registry.npmjs.org/>)
- Packages need a unique identifier
 - Typically a package name (sometimes owner name) and version
- Custom repositories allowed by most package managers
 - Often used for company-internal packages or cache mirroring
 - Note problems with duplicates (same package name in different repositories; some priority order is needed)
- Somebody needs to manage repositories
 - Availability: Repository needs to be running
 - Access Control: Packages should only be published by owners
 - Integrity: Packages should be signed or otherwise verifiable
 - Uniqueness and archival: Only one artifact per version
 - Traceability: Packages can have metadata pointing to source or tests
 - Security: ???



Transitive Dependencies

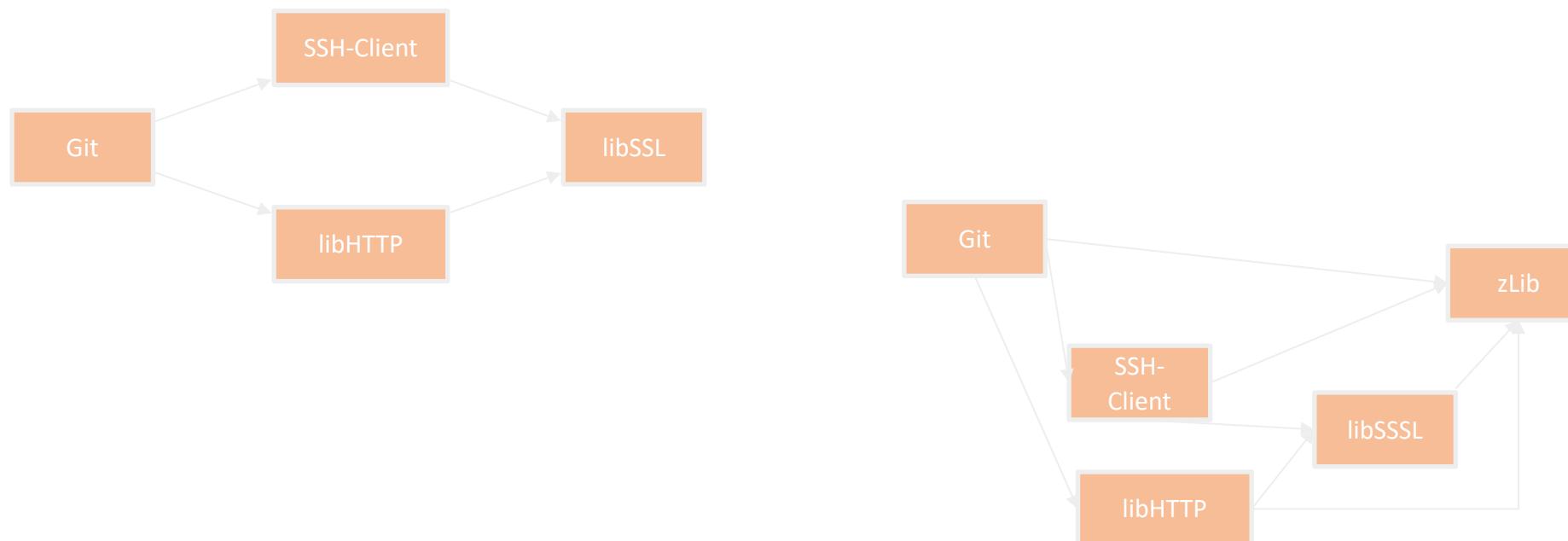
Packages can depend on other packages



Q: Should Git be able to use exports of libSSL (e.g. certificate management) or zLib (e.g. gzip compression)?

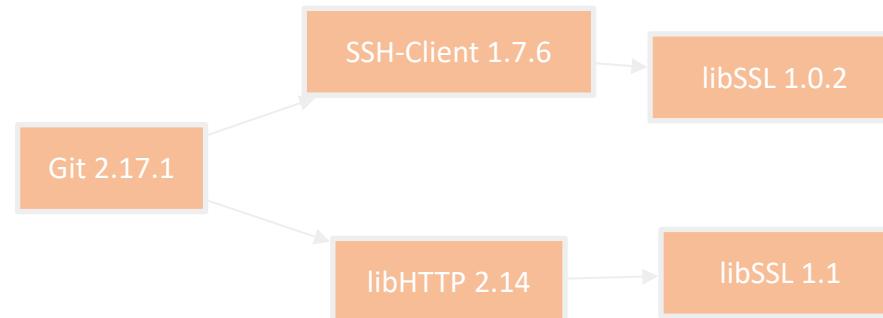
Diamond Dependencies

What are some problems when multiple intermediate dependencies have the same transitive dependency?



Diamond Dependencies

What are some problems when multiple intermediate dependencies have the same transitive dependency?



Resolutions to the Diamond Problem

1. Duplicate it!

- Doesn't work with static linking (e.g. C/C++), but may be doable with Java (e.g. using ClassLoader hacking or package renaming)
- Values of types defined by duplicated libraries cannot be exchanged across

2. Ban transitive dependencies; just use a global list with one version for each

- Challenge: Keeping things in sync with latest
- Challenge: Deciding which version of transitive deps to keep

3. Newest version (keep everything at latest)

- Requires ordering semantics
- Intermediate dependency may break with update to transitive

4. Oldest version (lowest denominator)

- Also requires ordering semantics
- Sacrifices new functionality

5. Oldest non-breaking version / Newest non-breaking version

- Requires faith in tests or semantic versioning contract

Semantic Versioning

- Widely used convention for versioning releases
 - E.g. 1.2.1, 3.1.0-alpha-1, 3.1.0-alpha-2, 3.1.0-beta-1, 3.1.0-rc1
- Format: {MAJOR} . {MINOR} . {PATCH}
- Each component is ordered (numerically, then lexicographically; release-aware)
 - 1.2.1 < 1.10.1
 - 3.1.0-alpha-1 < 3.1.0-alpha-2 < 3.1.0-beta-1 < 3.1.0-rc1 < 3.1.0
- Contracts:
 - MAJOR updated to indicate breaking changes
 - Same MAJOR version => backward compatibility
 - MINOR updated for additive changes
 - Same MINOR version => API compatibility (important for linking)
 - PATCH updates functionality without new API
 - Ninja edit; usually for bug fixes

<https://semver.org/>

[2.0.0](#) [2.0.0-rc.2](#) [2.0.0-rc.1](#) [1.0.0](#) [1.0.0-beta](#)

Semantic Versioning 2.0.0

Summary

Given a version number MAJOR.MINOR.PATCH, increment the:

1. MAJOR version when you make incompatible API changes,
2. MINOR version when you add functionality in a backwards compatible manner, and
3. PATCH version when you make backwards compatible bug fixes.

Additional labels for pre-release and build metadata are available as extensions to the MAJOR.MINOR.PATCH format.

People rely on SemVer contracts

The screenshot shows a GitHub repository page for 'rohanpadhye / JQF'. The 'Issues' tab is selected, showing 10 issues. The first issue, '#150 Clarify versioning schema', is open and was created by sdruskat on Aug 18. A comment from sdruskat on Aug 18 asks about the versioning schema, mentioning SemVer and API changes. Another comment from sdruskat expresses support for semantic versioning.

rohanpadhye / JQF Public

Code Issues 10 Pull requests 3 Discussions Actions Wiki Security Insights Settings

Clarify versioning schema #150

Open sdruskat opened this issue on Aug 18 · 3 comments

sdruskat commented on Aug 18

Hi, and thanks for a great project.

I'm wondering what the versioning schema for this project is. Seeing the tags (containing 1.8, etc.), I was assuming SemVer, but I see that the API has changed between minor increments (e.g., the newly added constructor arguments in ZestGuidance)? Or am I mixing up things?

FWIW, I think that following semantic versioning would be great, and make it easier for forks to contribute back to the upstream.

Dependency Constraints

- E.g. Declare dependency on "Bar > 2.1"
 - Bar 2.1.0, 2.1.1, 2.2.0, 2.9.0, etc. all match
 - 2.0.x does NOT match
 - 3.0.x does NOT match
- Diamond dependency problem can be resolved using SAT solvers
 - E.g. Foo 1.0.0 depends on "Bar >= 2.1" and "Baz 1.8.x"
 - Bar 2.1.0 depends on "Qux [1.6, 1.7]"
 - Bar 2.1.1 depends on "Qux 1.7.0"
 - Baz 1.8.0 depends on "Qux 1.5.x"
 - Baz 1.8.1 depends on "Qux 1.6.x"
 - Find an assignment such that all dependencies are satisfied
 - Solution: Use Bar 2.1.0, Baz 1.8.1, and Qux 1.6.{latest}

Semantic Versioning Contracts

- Largely trusting developers to maintain them
- Constrained/range dependencies can cause unexpected build failures
- Automatic validation of SemVer is hard

The screenshot shows a GitHub repository interface. At the top, there's a red 'Build' badge with a minus sign, indicating a failed build. Below it, a green 'README: Add build badge' badge with a checkmark. The repository has a 'master' branch. Two commits are listed: one from 9 days ago and one from 2 months ago. The 9-day-old commit is a failed build, while the 2-month-old commit is a successful README update.

The screenshot shows a GitHub pull request titled 'Pin jsonschema version to avoid swagger bugs'. The pull request has 70 reviews and 70 comments. It was last updated 9 days ago by user rohanpadhye. The diff shows changes in the file 'mayan/apps/rest_api/dependencies.py'. A specific line of code is highlighted in green, indicating a new addition: 'module=__name__, name='jsonschema', version_string='==3.2.0'.

```
diff --git a/mayan/apps/rest_api/dependencies.py b/mayan/apps/rest_api/dependencies.py
@@ -59,6 +59,9 @@ PythonDependency(
 60   60     module=__name__, name='flex', version_string='==6.14.1'
 61   61   )
+62 + PythonDependency(
+63 +   module=__name__, name='jsonschema', version_string='==3.2.0'
+64 + )
```

Cyclic Dependencies

- A very bad thing
- Avoid at all costs
- Sometimes unavoidable or intentional
 - E.g. GCC is written in C (needs a C compiler)
 - E.g. Apache Maven uses the Maven build system
 - E.g. JDK tested using JUnit, which requires the JDK to compile



Cyclic Dependencies

- Bootstrapping: Break cycles over time
- Assume older version exists in binary (pre-built form)
- Step 1: Build A using an older version of B
- Step 2: Build B using new (just built) version of A
- Step 3: Rebuild A using new (just built) version of B
- Now, both A and B have been built with new versions of their dependencies
- Doesn't work if both A and B need new features of each other at the same time (otherwise Step 1 won't work)
 - Assumes incremental dependence on new features
- How was the old version built in the first place? (it's turtles all the way down)
 - Assumption: cycles did not exist in the past
 - Successfully applied in compilers (e.g. GCC is written in C)

Dependency Security

- Will you let strangers execute arbitrary code on your laptop?
 - Think about this every time you do “pip install” or “npm install” or “apt-get upgrade” or “brew upgrade” or whatever (esp. with sudo)
 - Scary, right? Who are you trusting? Why?
- Typo squatting (“pip install numpi”)
- Outright malice (remember the *event-stream* incident?)
- Genuine security vulnerabilities due to software bugs

Dependabot alerts / #74

Dismiss alert ▾

Deserialization of Untrusted Data in Apache Log4j #74

Open Opened 3 days ago on log4j:log4j (Maven) · pom.xml

Severity	Critical	9.8 / 10
CVSS base metrics		
Attack vector	Network	Low
Attack complexity	None	Low
Privileges required	None	None
User interaction	None	None
Scope	Unchanged	High
Confidentiality	High	High
Integrity	High	High
Availability	High	High

Patch status: None

Package: log4j:log4j (Maven)

Affected versions: <= 1.2.17

CVE-2020-9493 identified a deserialization issue that was present in Apache Chainsaw. Prior to Chainsaw V2.0 Chainsaw was a component of Apache Log4j 1.2.x where the same issue exists.

Users are advised to migrate from `log4j:log4j` to `org.apache.logging.log4j:log4j` for an updated version of the library.

dependabot bot opened this 3 days ago

Weaknesses: CWE-502

CVSS:3.1/AV:N/AC:L/PR:N/UI:N/S:U/C:H/I:H/A:H

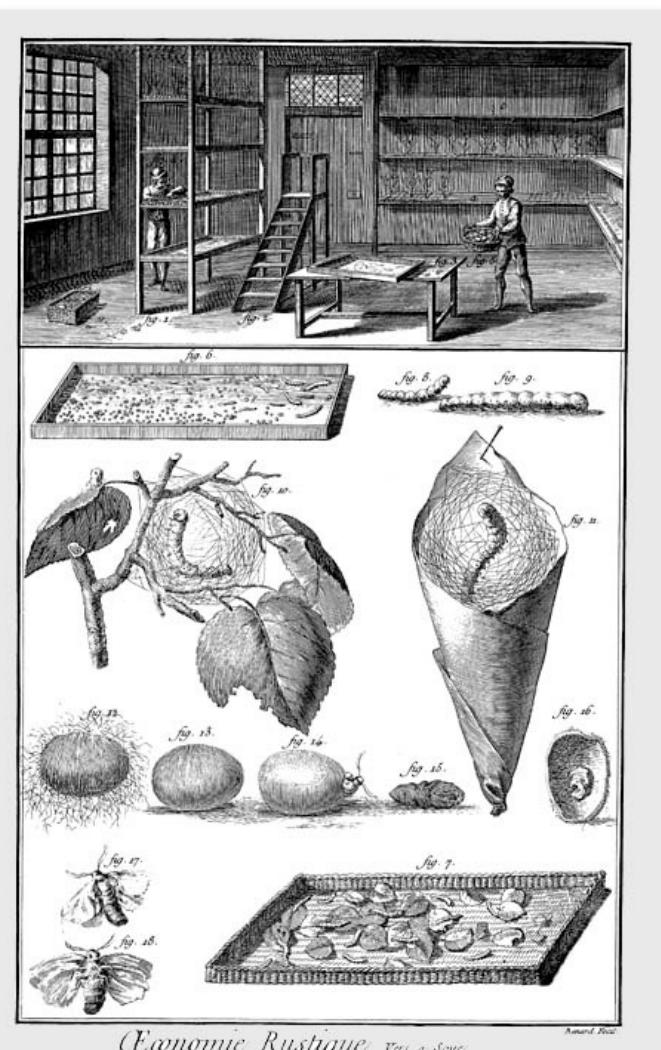
Takeaways

- Dependency management is hard.

Software Patents

Software Patents: The Good, The Bad, and The Ugly

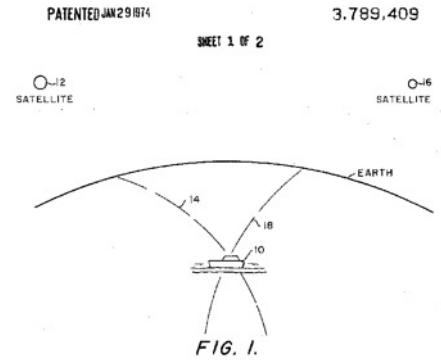
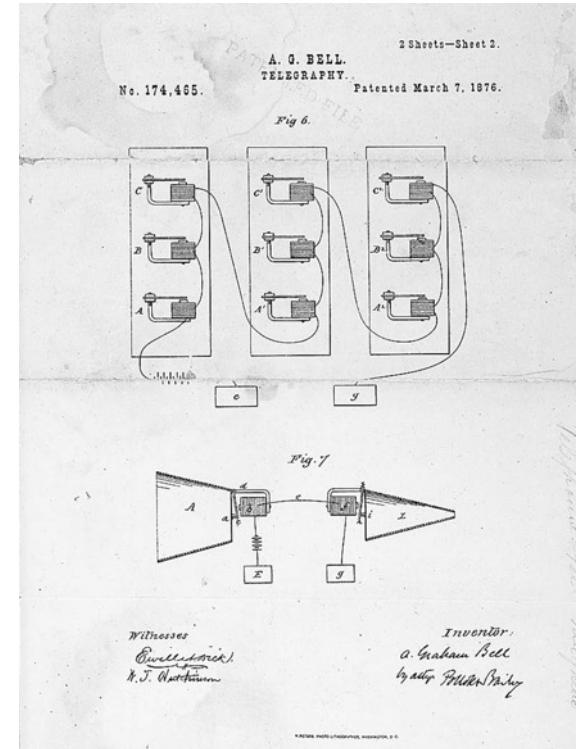
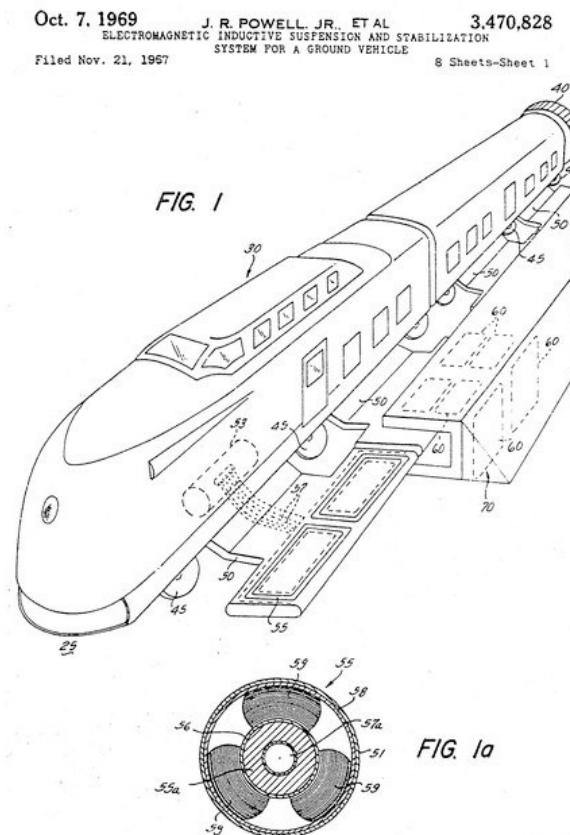
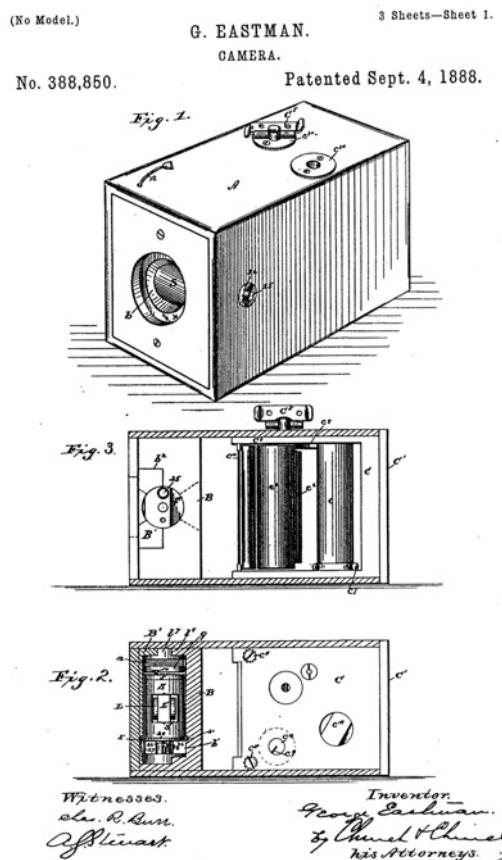
Venice, 1474



England, 1566



Today: USA



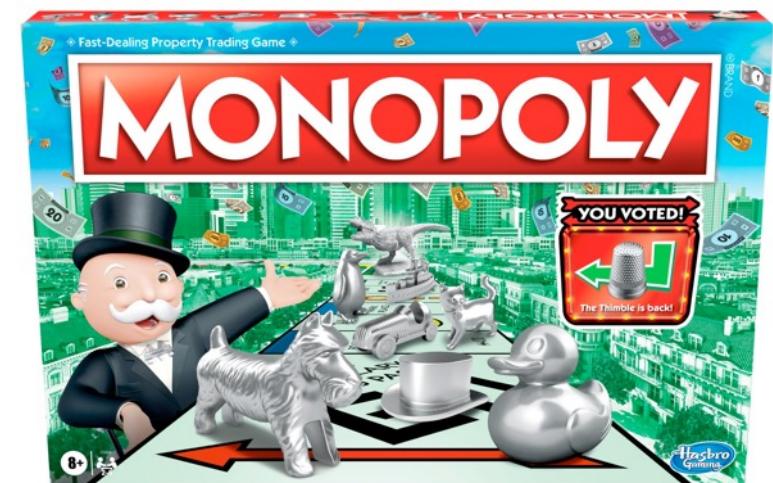
What is a patent? New. Useful. Non-obvious.

"A patent is an exclusive right granted for an invention, which is a product or a process that provides, in general, a new way of doing something, or offers a new technical solution to a problem. To get a patent, technical information about the invention must be disclosed to the public in a patent application."



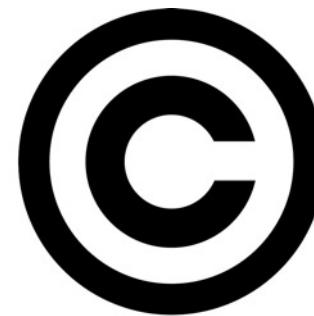
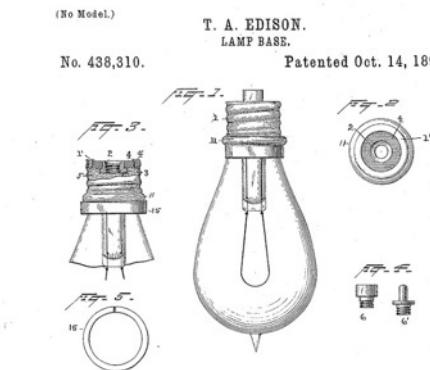
What rights do patents grant?

- Patents don't give you the right to make, use, or sell an invention.
- Patents do give you the right to exclude others from making, using, and selling an invention for the term of a patent (20 years)
 - stop or sue others
 - licensing and royalties



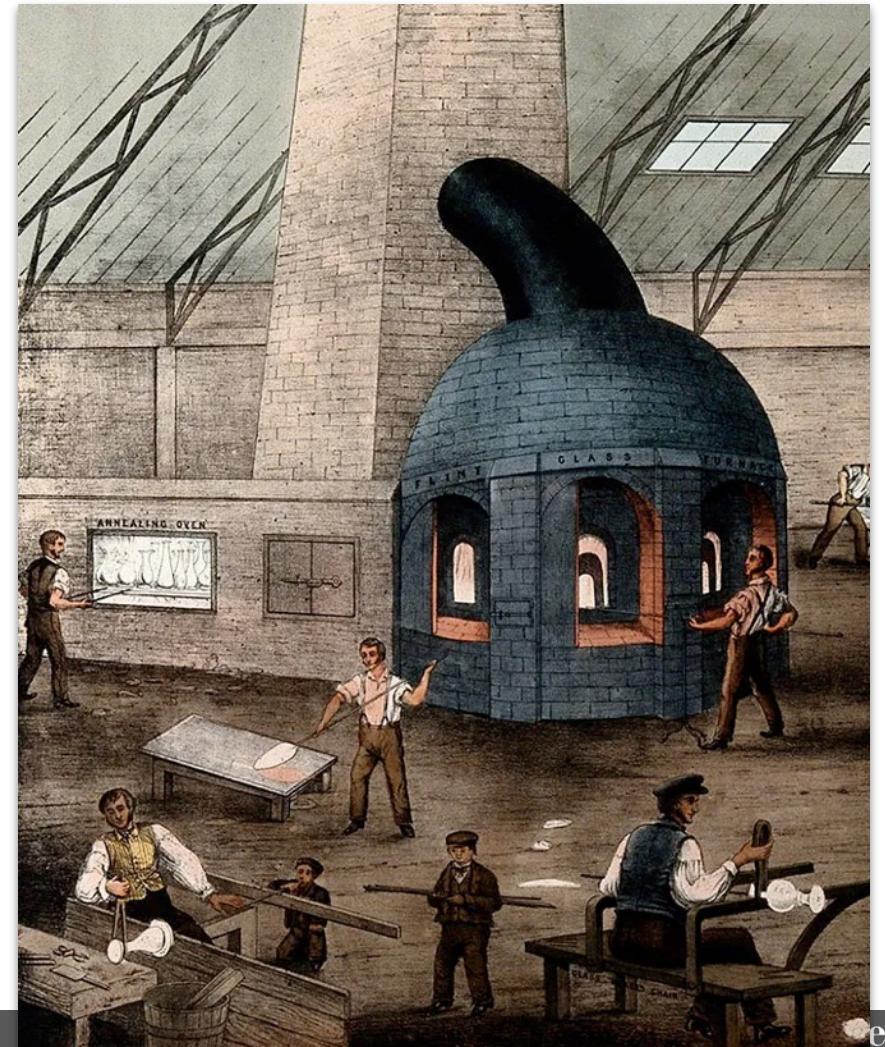
What's the difference? Patents vs. Copyright

- Copyrights cover the details of expression of a work
- Copyrights don't cover any ideas
Patents only cover ideas and the use of ideas
- Copyrights happen automatically.
Patents are issued by a patent office in response to an application.



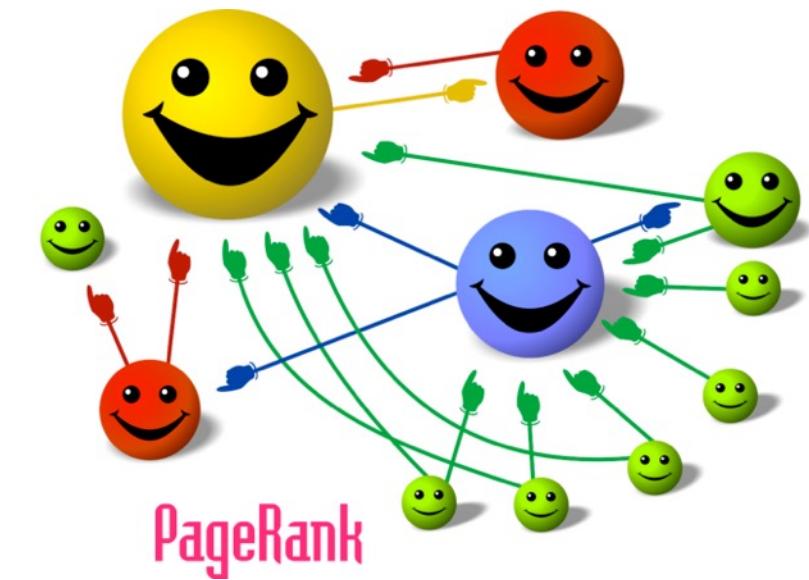
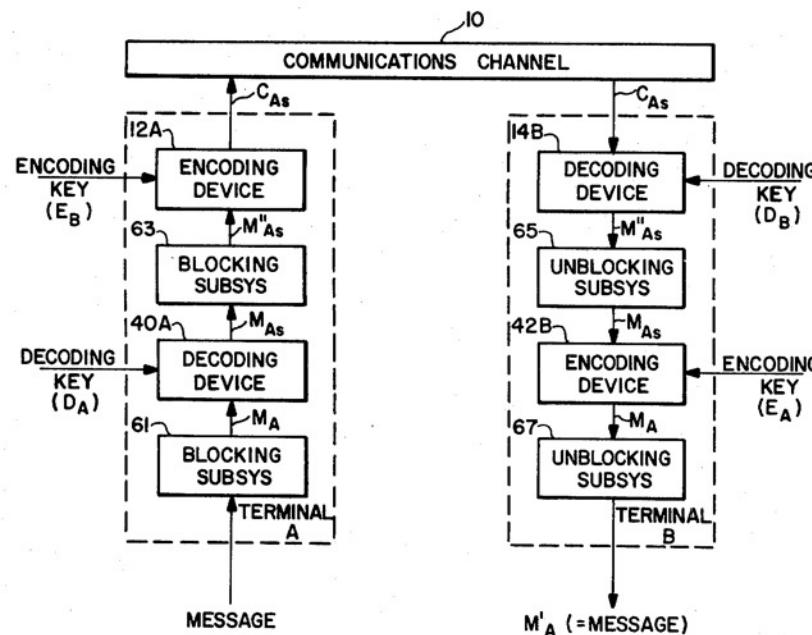
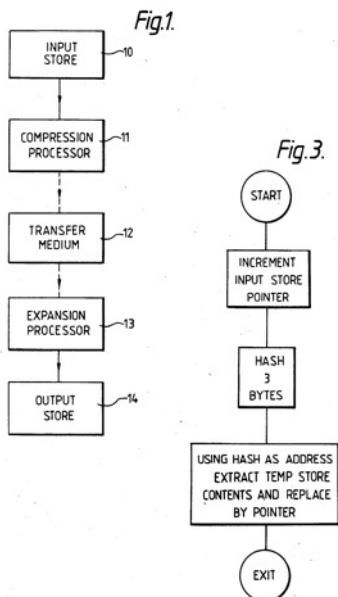
Why do patents exist?

- Encourage disclosure of inventions
- Reward invention and creativity
- Protect investment of capital into R&D
- Encourage the market to “design around”
- Protect small companies from large ones



Software Patents

U.S. Patent Oct. 20, 1987 Sheet 1 of 6 4,701,745



Patent or not?

Patent or not?

1. Running bingo on a computer
2. Using a computer to help users plan meals while achieving diet goals
3. Using a computer to order a pizza with customized toppings
4. Prompting a user before establishing a new network connection
5. Automatically notifying users when an item is picked up or delivered
6. Using a computer network to ask people to complete tasks and then wait for them to do them
7. Using SMS to perform tasks (e.g., checking bank balance)
8. Selecting ALL images in a CAPTCHA that match a given text

The software patent system is broken!

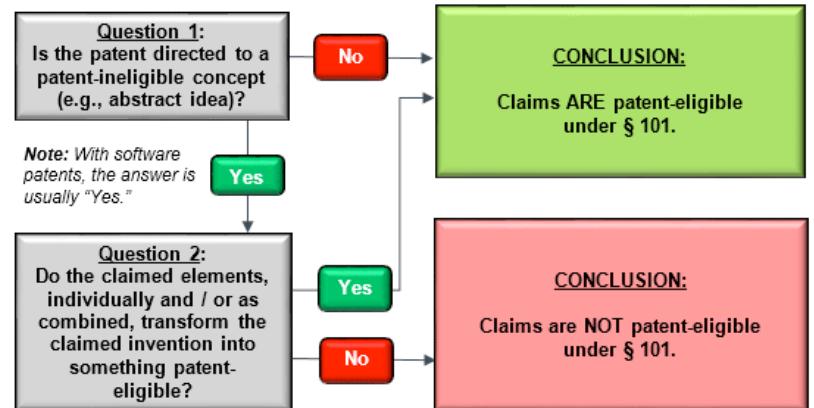
Alice vs. CLS Bank (2014)

Case	Claimed Invention		Result
<i>Alice Corp. v. CLS Bank</i> (June 19, 2014)	Method of computerized risk mitigation in financial settlements	X Step 1 X Step 2	NOT Patent Eligible <i>Why?</i> Risk mitigation is a long-standing “fundamental economic practice” (step 1) and the claims merely required generic computer implementation (step 2)
<i>Digitech</i> (July 11, 2014) 	Method of digital image processing; used “device profiles” to organize devices’ spatial and color properties	X Step 1 X Step 2	NOT Patent Eligible <i>Why?</i> Claimed “device profile” was intangible; method claims covered organization of information untethered to specific structure.
<i>buySAFE v. Google</i> (Sep. 3, 2014)	Online transaction performance guarantee	X Step 1 X Step 2	NOT Patent Eligible <i>Why?</i> The claims are about creating a contractual relationship that is performed by any general purpose computer.
<i>Ultramerical v. Hulu</i> (Nov. 14, 2014)	Internet-distribution of copyright material	X Step 1 X Step 2	NOT Patent Eligible <i>Why?</i> Offering media in exchange for viewing an advertisement is an abstract idea. Implementing it on the internet does not transform it into patent eligible.

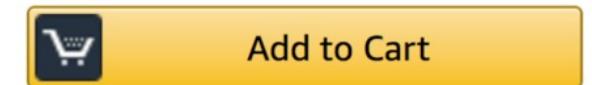
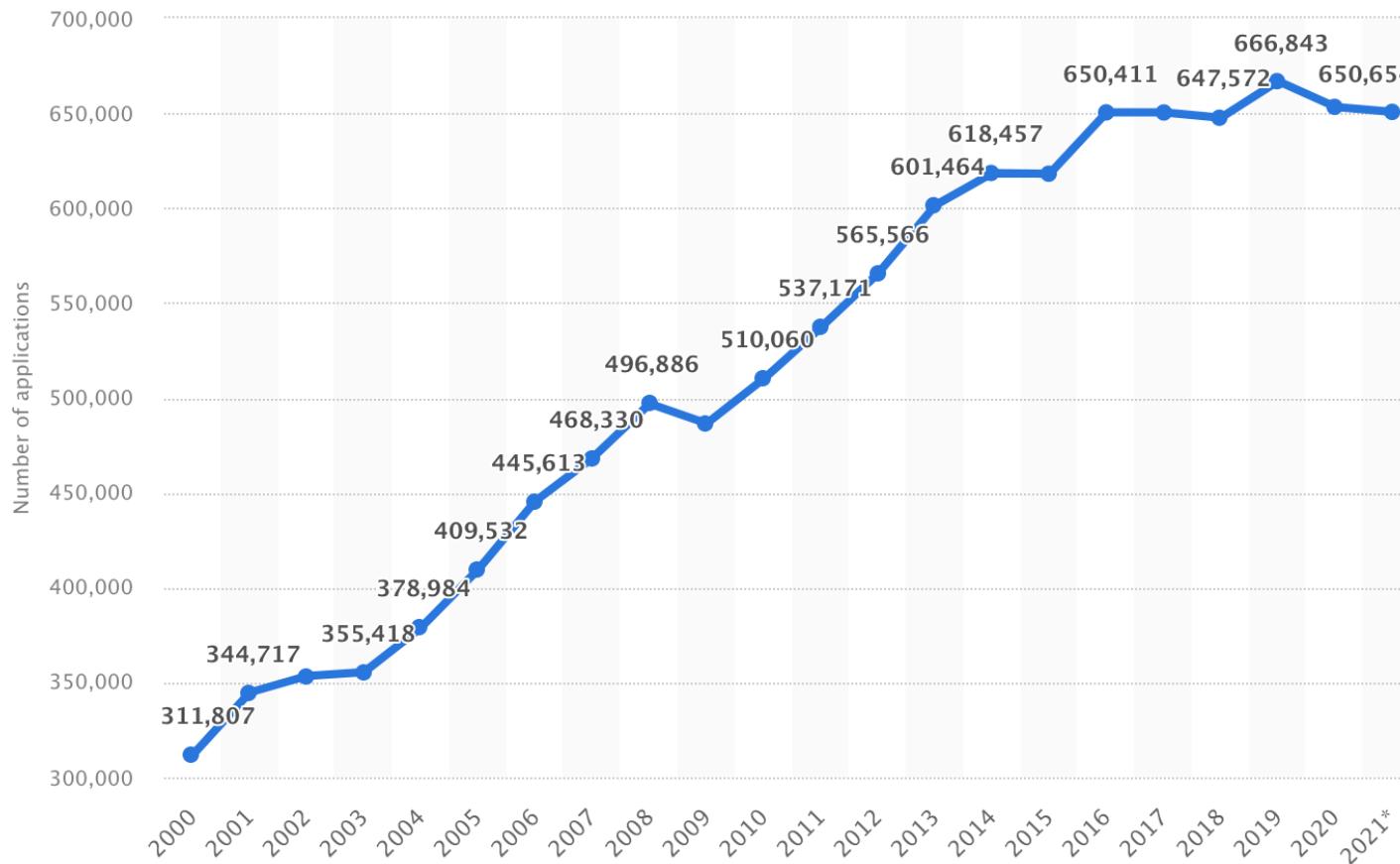
ars TECHNICA SUBSCRIBE  SIGN IN

POLICY — Supreme Court smashes “do it on a computer” patents in 9-0 opinion

Court declines to stop software patents altogether.
JOE MULLIN • 6/19/2014, 12:08 PM



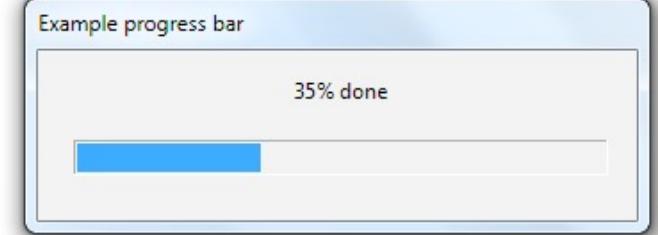
Problem: Inventive step and non-obviousness



or 1-Click Checkout



US5960411A



US5301348A

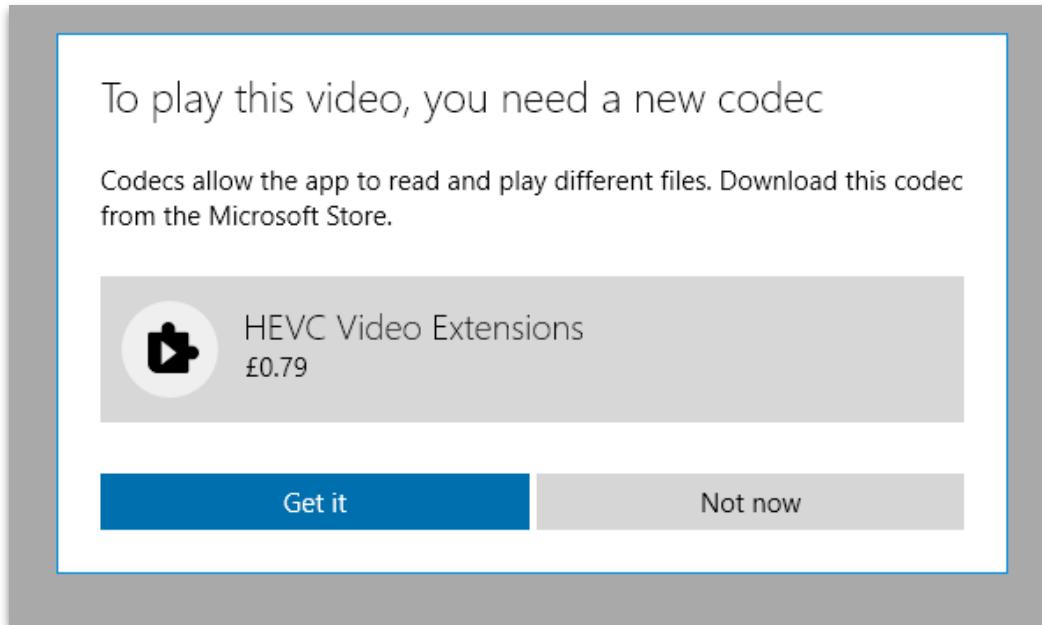
Problem: Long patent pendencies and terms

TABLE 4: PATENT PENDENCY STATISTICS (FY 2021)

Utility, Plant, Reissue Pendency Statistics by Technology Center (in months)	Average First Action Pendency	Total Average Pendency
Total Utility, Plant, and Reissue Pendency	16.9	23.3
Tech Center 1600—Biotechnology and Organic Chemistry	17.0	24.0
Tech Center 1700—Chemical and Materials Engineering	18.8	26.7
Tech Center 2100—Computer Architecture, Software, and Information Security	17.5	25.6
Tech Center 2400—Networks, Multiplexing, Cable, and Security	15.7	22.9
Tech Center 2600—Communications	13.5	19.9
Tech Center 2800—Semiconductor, Electrical, Optical Systems, and Components	15.7	22.3
Tech Center 3600—Transportation, Construction, Agriculture, and Electronic Commerce	18.1	25.9
Tech Center 3700—Mechanical Engineering, Manufacturing, and Products	18.6	26.7

Problem: Incompatibility

- PNG was invented to avoid GIF patent issues
- Opus is a patent-free MP3 alternative
- AV1 vs H265



Problem: Independent discovery doesn't matter!

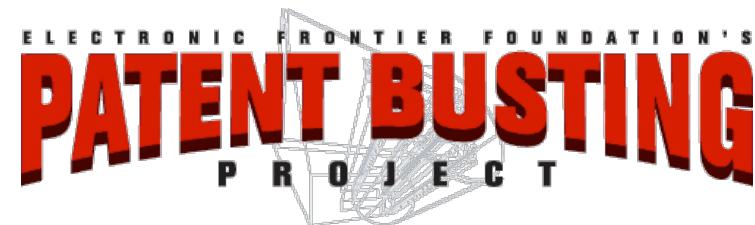
“The idea that I can be presented with a problem, set out to logically solve it with the tools at hand, and wind up with a program that could not be legally used because someone else followed the same logical steps some years ago and filed for a patent on it is horrifying.”

John Carmack



Problem: Only large organizations benefit

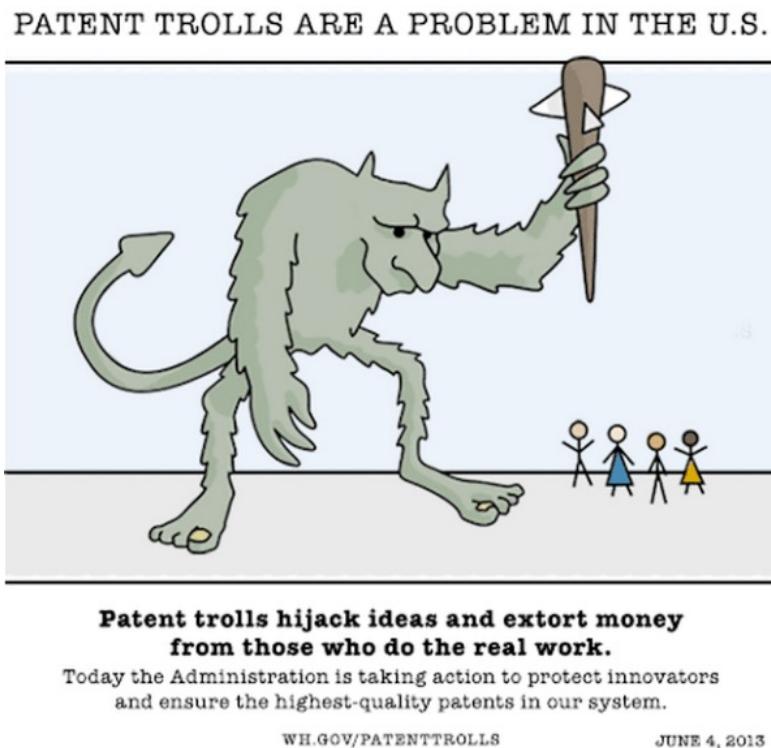
- The patent system relies on people to challenge bad patents
 - requires considerable time, money, and legal expertise
 - the US legal system requires both parties to pay legal fees (c.f., losers pay costs in Europe) *
- US software patents cost between \$15,000 to \$45,000!
 - that's before you even apply for international patents!



<https://www.patenttrademarkblog.com/how-much-patent-costs>

<https://www.eff.org/issues/patent-busting-project>

Problem: Non-Practicing Entities (Patent Trolls)



BBC Home News Sport Reel Worklife

NEWS

Home | War in Ukraine | Coronavirus | Climate | Video | World | US & Canada | UK

Tech

'Patent trolls' cost other US bodies \$29bn last year, says study

© 29 June 2012

Infringement of patents

of _____ which are alleged to be infrin
oss by reason of purchasing and selling _____
erson or persons purchasing and using su
parties that A will sell B

THINKSTOCK

Patent portfolio owners say their actions help incentivise inventors to carry out research

Problem: Innovation is Stifled

"As a developer for a small startup, absurd software patents are a constant worry. Stories abound of people like us getting pressured out of existence over the use of incredibly vague, basic interface elements and system components."

"Software patents are generally written in vague and nontechnical legal language, which obfuscates the patent in question . . . and also makes it easy to dramatically extend the patent to elements not considered at all when the patent was originally filed."



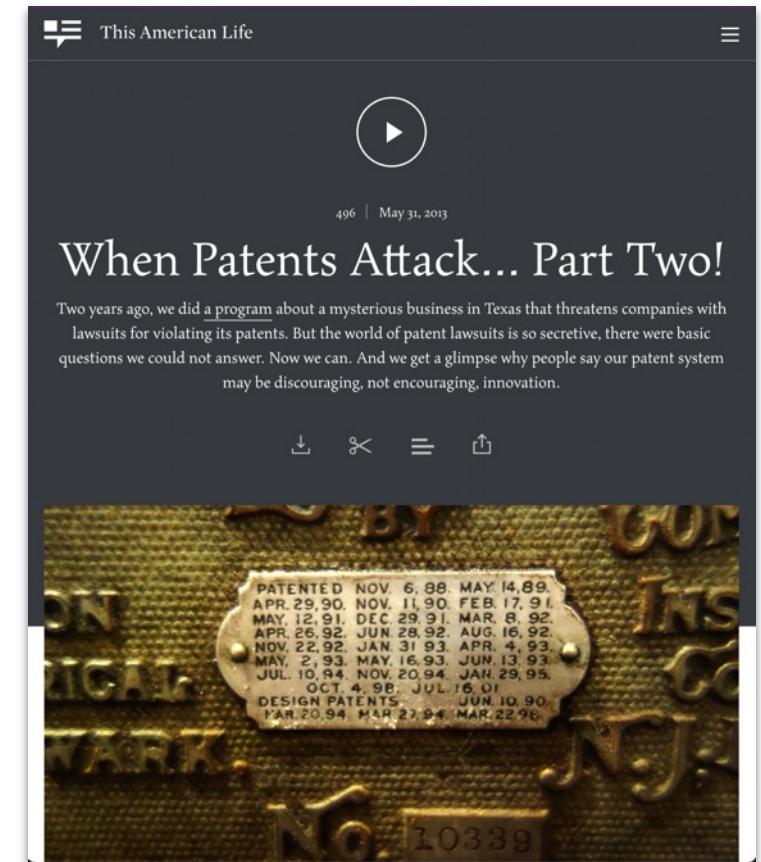
Defend Innovation
How to Fix Our Broken Patent System
by Adi Kamdar, Daniel Nazer, Vera Ranieri
FEBRUARY 2015

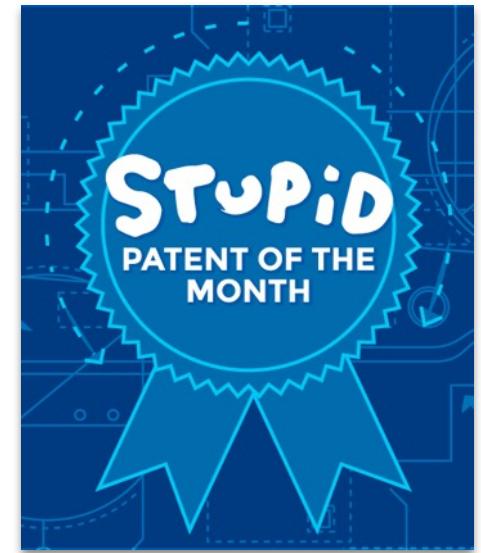
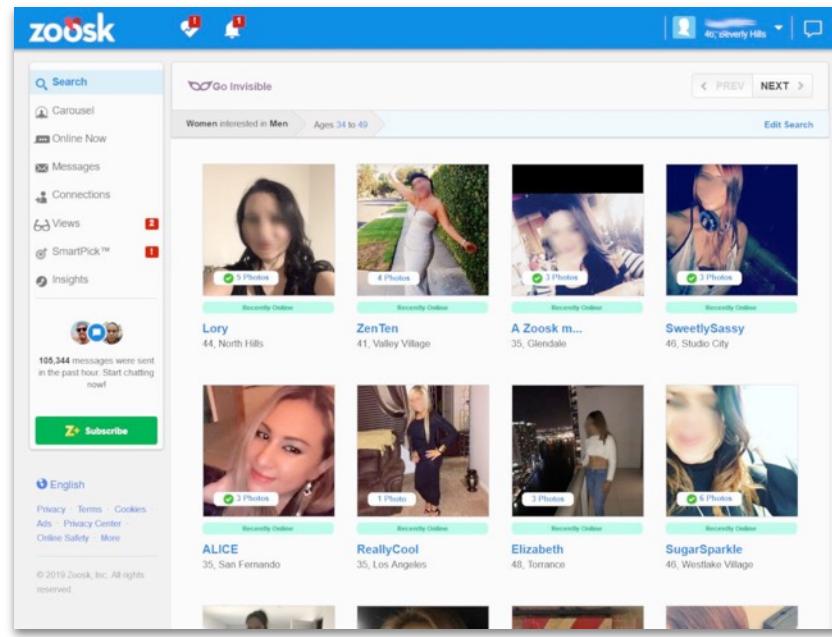
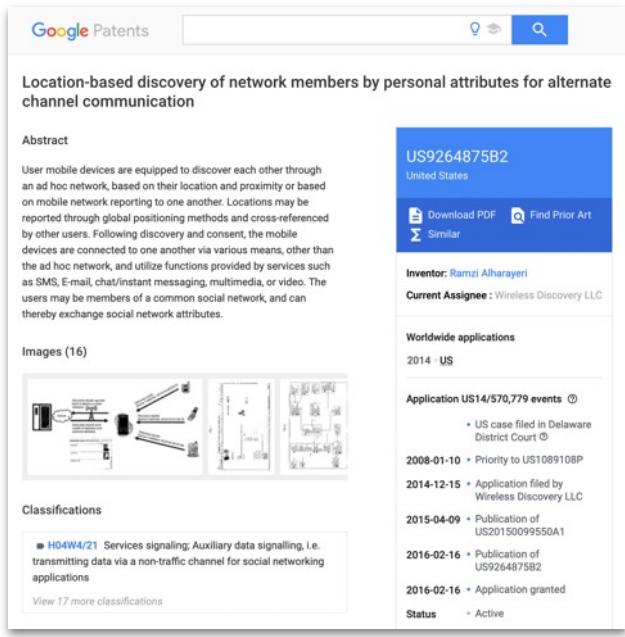


ELECTRONIC FRONTIER FOUNDATION

This American Life: When Patents Attack!

- Innovatio sued libraries and coffee shops for providing WiFi in a public space
- Boadin has sued various media outlets, claiming that its patents are infringed whenever a word or phrase on your computer autocompletes
- NPHJ claims they hold a patent on “scanning and emailing documents”. They tried to sue non-profits for \$1000 per employee in damages.





- Zoosk has a website that mobile devices can connect to
- Zoosk's server collects information from the mobile devices, including location and unique device identifiers
- Zoosk users can send and accept invitations to connect with and send messages to each other.
- Zoosk shares profile information of connected users, who are “members of a same social network” (i.e., they’re on Zoosk)
- Zoosk can connect users who are in the immediate vicinity of each other, or a particular distance away

Problem: Open Source is under attack, too!

THE LINUX FOUNDATION [JOIN](#)

5 MIN READ

Ensuring Patents Foster Innovation in Open Source

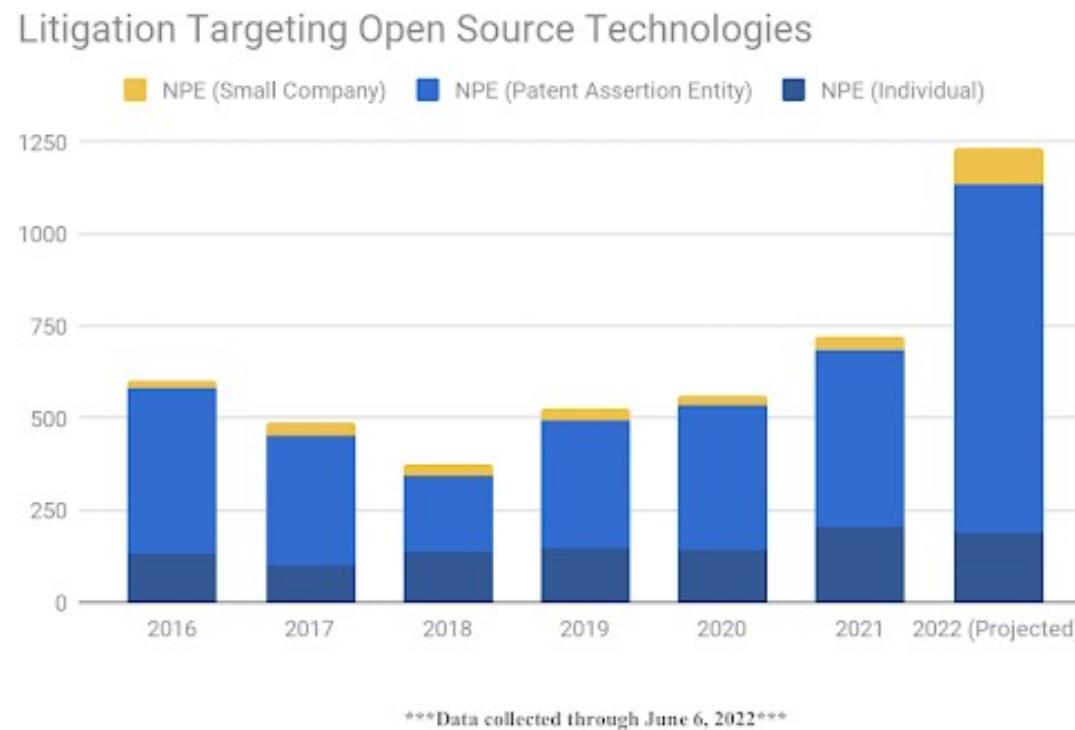
DAN WHITING | 23 JUNE 2022

So, I am old enough to remember when the U.S. Congress temporarily intervened in a patent dispute over the technology that powered BlackBerrys. A U.S. Federal judge ordered the BlackBerry service to shutdown until the matter was resolved, and Congress determined that BlackBerry service was too integral to commerce to be allowed to be turned off. Eventually, RIM settled the patent dispute and the BlackBerry rode off into technology oblivion.

I am not here to argue the merits of this nearly 20-year-old case (in fact, I coincidentally had friends on both legal teams), but it was when I was introduced to the idea of companies that purchase patents with the goal of using this purchased right to extract money from other companies.

Patents are an important legal protection to foster innovation, but, like all systems, it isn't perfect.

At this week's Open Source Summit North America, we heard from Kevin Jakel with Unified Patents. Kevin is a patent attorney who saw damage being done to innovation by patent trolls – more kindly known as non-practicing entities (NPEs).



zdNET [Home](#) / [Business](#) / [Enterprise Software](#)

Patent troll attacks against open source projects are up 100% since last year. Here's why

In recent years, patent trolls have started attacking open-source developers and companies. But, the open-source community is fighting back.

 Written by Steven Vaughan-Nichols, Senior Contributing Editor on Sept. 12, 2022

[D](#) [in](#) [F](#) [T](#)



What next?

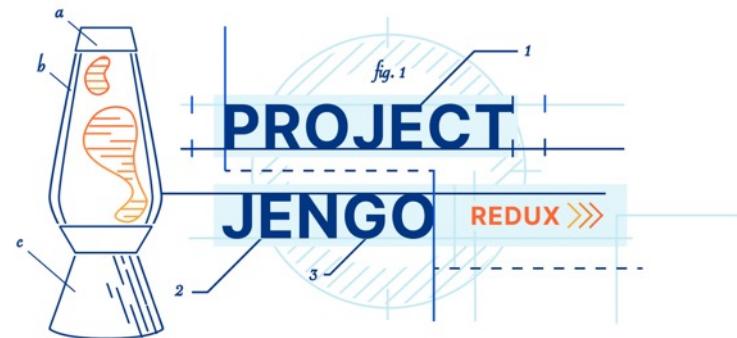
- Alternative licensing models
 - The Defensive Patent License (DPL)
 - The Open Invention Network (OIN)
 - License on Transfer (LOT)
- Bogus patent bounties
- [Unified Patents](#)
- Commonsense reform
- Abolish software patents?

Project Jengo Redux: Cloudflare's Prior Art Search Bounty Returns

04/26/2021



Doug Kramer



Decision Calculus

Life is all about tradeoffs

In this course, we have talked a lot about tradeoffs.

Some tradeoffs we have discussed:

Writing Tests vs writing more features

Choosing a familiar tech stack vs a “trendy one”

Other tradeoffs...?

Think about structured ways to make decisions

Do what is ethical, legal, moral, obvious

Sometimes, there are multiple (legitimate) options, that all seem to have positives and negatives.

What to do?

People have studied this before...

Economics

Biology + Microbiology

Sociology

Engineering

Computer Science

Strategy Games

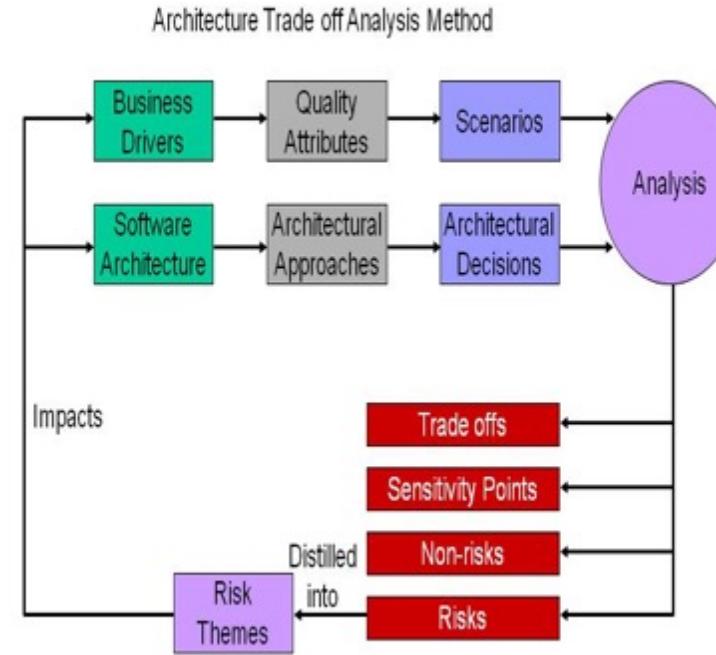
Ethics

Medicine

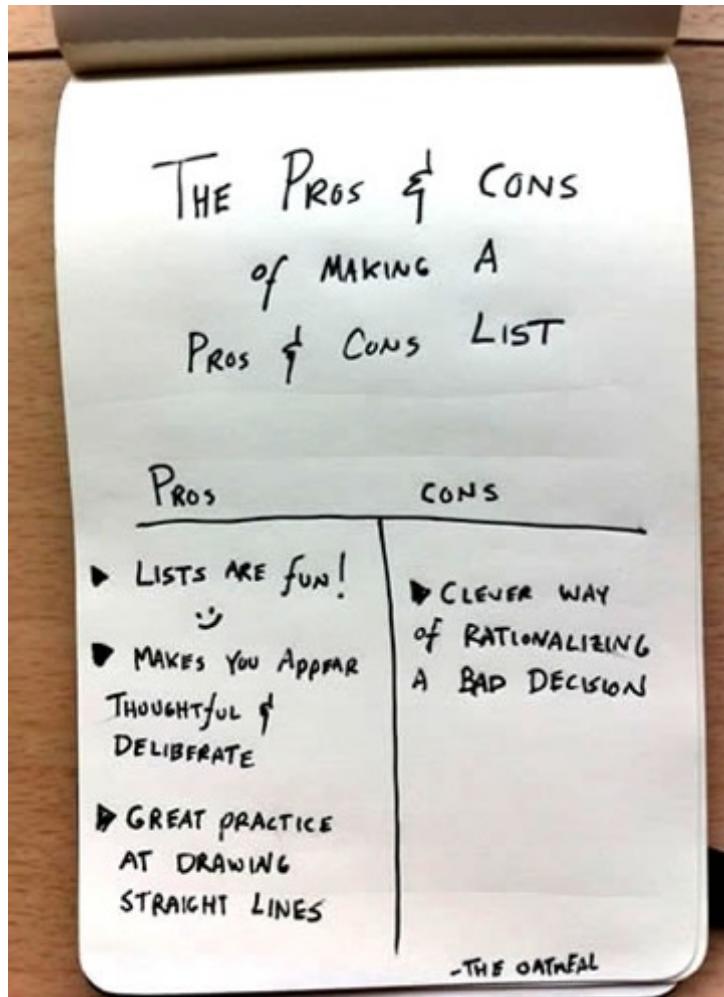
Politics

Architecture Tradeoff Analysis Method

1. Present the ATAM.
2. Present business drivers.
3. Present architecture.
4. Identify architectural approaches.
5. Generate quality attribute utility tree.
6. Analyze architectural approaches.
7. Brainstorm and prioritize scenarios.
[https://concisosoftware.com/architecture-tradeoff-analysis-method-atam/](https://concisesoftware.com/architecture-tradeoff-analysis-method-atam/)

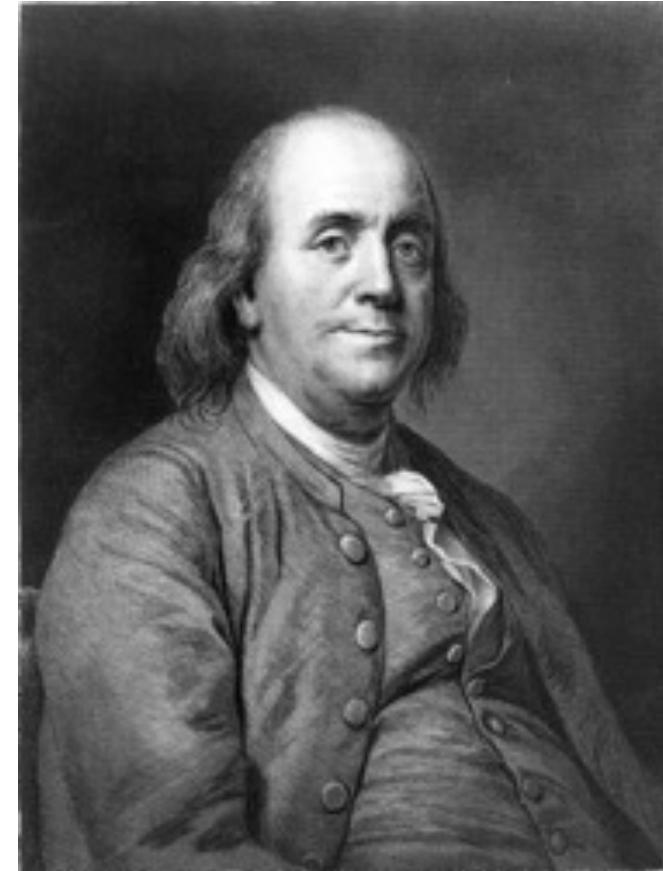


Pros and cons



Moral or Prudential Algebra

1. Make a list of Pros and Cons
2. Take a few days to think, forcing your brain to come up with all of the positive and negative aspects of taking the measure you're considering.
3. Assign weights to each item on your list depending on its importance.
4. When the weight of a item on your “pro” list is equal to the weight of an item on your “con” list, strike both of them out.
5. What is left in the balance is the choice you need to make. Wait a few more days and, if nothing new occurs to you, act on the decision you've made.



Even Swaps

Premise:

It is easy to make decisions when there is only one objective.

But having only one objective, as any decision maker knows, is a rare luxury.

Even swaps provides a practical way of making trade-offs among any set of objectives across a range of alternatives.



THE POLITICS OF E-MAIL PAGE 12

Harvard Business Review

MARCH - APRIL 1998

Illustration: A colorful, stylized illustration of several figures in a dynamic environment. One figure is at a piano, another is near a car, and others are in various poses, suggesting a fast-paced or complex world.

Table of Contents:

- KATHLEEN M. EISENhardt and SHONA T. BROWN
- JOAN MAGRETta
- RICHARD K. NESTER, MICHAEL J. PIORE, and KAMAL M. MALEK
- JEAN-FRANCOIS MAZZONI and JEAN-LOUIS BARROUX
- JOSEPH I. BADARRECCO, JR.
- SHIKHAR GHOSH
- ROBERT Galford
- PAUL SHARPE and TOM KEELIN
- JOHN S. HAMMOND, RALPH L. KEENLY, and HOWARD RAIFEA
- JEFFREY ELTON and JUSTIN ROE
- TIME PACING: COMPETING IN MARKETS THAT WON'T STAND STILL 59
- THE POWER OF VIRTUAL INTEGRATION: AN INTERVIEW WITH DELL COMPUTER'S MICHAEL DELL 72
- INTERPRETIVE MANAGEMENT: WHAT GENERAL MANAGERS CAN LEARN FROM DESIGN 86
- THE SET-UP-TO-FAIL SYNDROME 101
- THE DISCIPLINE OF BUILDING CHARACTER 114
- MAKING BUSINESS SENSE OF THE INTERNET 126
- HBR CASE STUDY: WHY DOESN'T THIS HR DEPARTMENT GET ANY RESPECT? 128
- IDEAS AT WORK: HOW SMITHKLINE BEECHAM MAKES BETTER RESOURCE-ALLOCATION DECISIONS 145
- MANAGER'S TOOL KIT: EVEN SWAPS: A RATIONAL METHOD FOR MAKING TRADE-OFFS 157
- BOOKS IN REVIEW: BRINGING DISCIPLINE TO PROJECT MANAGEMENT 153

BRIEFINGS 12 LETTERS 160 EXECUTIVE SUMMARIES 177

Overview - 3 step process

1. Create a Consequences Table
2. Eliminate “Dominated” Alternatives
3. Make Even Swaps

Running Example from research paper

Example: Alan Miller is a computer scientist who started a technical consulting practice three years ago.

For the first year, he worked out of his home, but as his business grew he decided to sign a two-year lease on some space in the Pierpoint office park.

Now that lease is about to expire. He needs to decide whether to renew it or move to a new location.

Define the objectives

Alan defines five overriding objectives that he needs his office to fulfill:

1. a short commute from home
2. good access to his clients
3. good office services (clerical assistance, copiers and fax machines, and mail service)
4. sufficient space
5. low costs

He finds five viable alternatives: Parkway, Lombard, Baranov, Montana, and his current building, the Pierpoint.

Create Consequences Table

Objectives	Parkway	Lombard	Baranov	Montana	Pierpoint
Commute in Minutes	45	25	20	25	30
Customer Access (%)	50	80	70	85	75
Office Services	A	B	C	A	C
Office Size (Square Feet)	800	700	500	950	700
Monthly Cost (\$)	1850	1700	1500	1900	1750

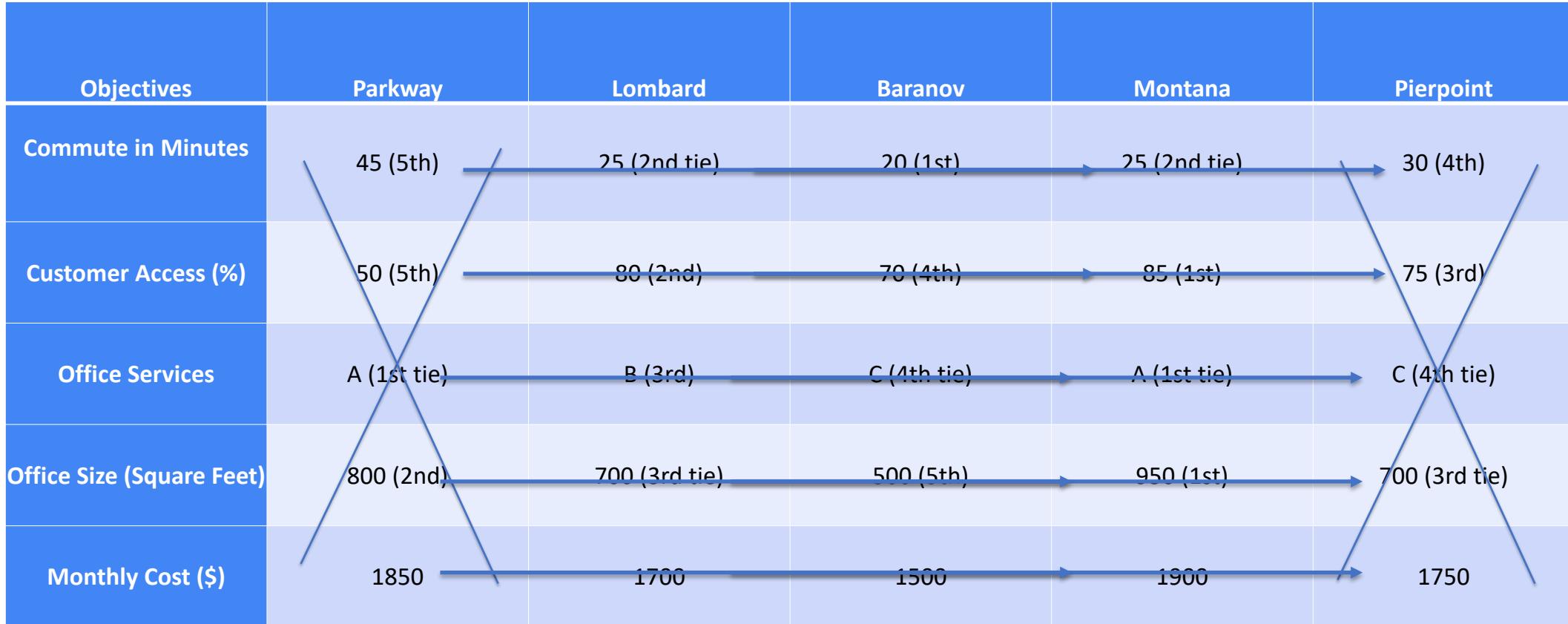
<https://hbr.org/1998/03/even-swaps-a-rational-method-for-making-trade-offs>

Create Ranking Table

Objectives	Parkway	Lombard	Baranov	Montana	Pierpoint
Commute in Minutes	45 (5th)	25 (2nd tie)	20 (1st)	25 (2nd tie)	30 (4th)
Customer Access (%)	50 (5th)	80 (2nd)	70 (4th)	85 (1st)	75 (3rd)
Office Services	A (1st tie)	B (3rd)	C (4th tie)	A (1st tie)	C (4th tie)
Office Size (Square Feet)	800 (2nd)	700 (3rd tie)	500 (5th)	950 (1st)	700 (3rd tie)
Monthly Cost (\$)	1850 (4th)	1700 (2nd)	1500 (1st)	1900 (5th)	1750 (3rd)

<https://hbr.org/1998/03/even-swaps-a-rational-method-for-making-trade-offs>

Eliminate “Dominated” Alternatives



<https://hbr.org/1998/03/even-swaps-a-rational-method-for-making-trade-offs>

Make Even Swaps

1. Determine the change necessary to cancel out an objective.
2. Assess what change in another objective would compensate for the needed change.
3. Make the even swap.
4. Cancel out the now-irrelevant objective.
5. Select the dominant alternative.

First Even Swaps

- For Baranov, swap 5 commute minutes, for 8% customer access.

Objectives	Lombard	Baranov	Montana
Commute in Minutes	25 (1st tie)	20 25 (1st tie)	25 (1st tie)
Customer Access (%)	80	70 78	85
Office Services	B	C	A
Office Size (Square Feet)	700	500	950
Monthly Cost (\$)	1700	1500	1900

Second Even Swaps

- For Baranov, swap C->B, add \$200, for Montana, swap A->B, reduce \$100

Objectives	Lombard	Baranov	Montana
Commute in Minutes	25	25	25
Customer Access (%)	80	78	85
Office Services	B	C B	A B
Office Size (Square Feet)	700	500	950
Monthly Cost (\$)	1700	1500 1700	1900 1800

Second Even Swaps

- For Lombard, swap 250 sq ft for \$250

Objectives	Lombard	Montana
Commute in Minutes	25	25
Customer Access (%)	80	85
Office Services	B	B
Office Size (Square Feet)	700 950	950
Monthly Cost (\$)	1700 1950	1800

Advice for swaps

- Make the easier swaps first
- Concentrate on the amount of the swap, not on the apparent importance of the overall objective.
- Remember that the value of an incremental change depends on what you start with.
- Make consistent swaps.
- Seek out solid information.