

Lambda Calculus and its Implementation in Computer Science

Joris Klaasse Bos
Zaanlands Lyceum

18th January 2021

Preface

The reason I chose to write about this subject, is that it combines two things I enjoy: abstract mathematics and computer science. I have been programming for about five to six years now. I mainly enjoy low-level programming, so—naturally—C is my most used language and I am most familiar with a simple procedural paradigm. Such a paradigm is, however, not always very easy to use when working with very large and complex systems. I, as many, started out with object-oriented programming (OOP), but I did not like that very much. Therefore, I have been exploring alternative paradigms, including data-oriented programming and functional programming. I am quite familiar with data-oriented programming and the Rust programming language by now, but functional programming isn't something I've ever really got into yet. I did find out about lambda calculus and combinatory logic, which intrigued me, but I haven't got into it beyond a basic level of understanding. That is why I decided to research it for this work.

I have some things to note on the structure and style of this work. Firstly, the structure. The work starts with some introductory information about lambda calculus and its history. It goes on to explain the simple syntax of lambda calculus and how it can be used. The next section goes into lambda calculus in-depth. It really shows how lambda calculus is used as a Turing complete mathematical system. The section after that explains functional programming. It shows how lambda calculus can be turned into a programming language, and explains why it is useful by comparing it to other paradigms, namely procedural imperative programming (what that means isn't really of importance right now). The final few sections are smaller; they really just meant to illustrate what is said in the preceding sections.

Secondly, every reference will be denoted with a number in braces which corresponds with a work found in the back on the references page. Many of my sources, however, aren't on there, because I didn't directly refer to them. Those sources will be displayed on the page after the references, the bibliography (as of now, I haven't yet compiled those sources and created that page).

As for the style in which I explain lambda calculus, I may differ in it from how other literature usually approaches the topic. Lambda calculus is often used in a mathematical context, so people often write in a very mathematical style. People don't usually write pure lambda calculus, they borrow a lot of symbols from other mathematical disciplines, such as logic. Nor do people define it in words, but also with mathematical symbols and definitions. I try to keep the use of these symbols to a minimum so as to lower the barrier to entry. Rather than using mathematical symbols and definitions, I use my words to explain the lambda calculus, thus this work is written in a more prose-like style, rather than a strictly mathematical one. I think that lambda calculus is simple enough that, when explained well, most people should be able to follow it. On a similar note, I think that the way this work is written, most people who are not familiar with mathematical literature should be able to follow it.

Contents

Preface	2
1 Introduction	5
2 Introduction to lambda calculus	5
2.1 A short history	5
2.2 The syntax	6
2.3 Combinatory logic	8
2.3.1 Identity	8
2.3.2 The omega combinator	8
2.3.3 The constant combinator	9
2.3.4 The kite	9
2.3.5 The flip combinator	10
2.3.6 The composition combinator	10
2.3.7 The thrush	11
2.3.8 The vireo	11
2.3.9 The starling	11
2.3.10 S and K	11
2.3.11 To Mock a Mockingbird	12
3 Using lambda calculus for computation	13
3.1 Church encoding	13
3.1.1 Simple boolean operations	13
3.1.2 Natural numbers	17
3.1.3 Boolean comparison	20
3.1.4 Data structures	21
3.1.5 Natural numbers continuation	22
3.1.6 Other kinds of numbers	23
3.2 Recursion	24
3.2.1 Understanding recursion	24
3.2.2 General recursion	26
3.2.3 The Y-combinator	27
4 Functional programming (lambda calculus applied)	30
4.1 Interesting and useful properties	30
4.1.1 Statelessness	30
4.1.2 First class and higher order functions	30
4.1.3 Function currying	30
4.1.4 Laziness	30
4.2 Typed lambda calculus	30
4.2.1 Simply typed lambda calculus	30
4.2.2 Strongly typed lambda calculus (System F)	30
4.3 Category theory	30
4.3.1 A quick look at set theory	30
4.3.2 The jist (of category theory)	31
4.3.3 Composition	31
4.3.4 Functors	31
4.3.5 Monoids	32
4.3.6 Monads	32
4.4 Haskell	32

4.4.1	Basic syntax	32
4.4.2	Monads	32
4.5	Comparison to other paradigms	32
4.5.1	Declarative vs imperative	32
4.5.2	Usefulness vs conceptual purity	32
4.5.3	Meta programming	32
5	A practical example	32
5.1	Scheme interpreter in C	32
5.2	Scheme interpreter in Haskell	32
	Afterword	34
	Unreferenced resources	34
	References	35

1 Introduction

With the decline of OOP (object-oriented programming), many other paradigms are gaining in popularity. One increasingly popular paradigm is functional programming. Functional programming is fundamentally based on lambda calculus and it has been seeping into other paradigms and into mainstream languages. Most of the popular languages now implement lambda functions and have ways to write in a more declarative style of programming. In this work I will look at all the ways lambda calculus has influenced computer science and how it may do so in the future.

2 Introduction to lambda calculus

Lambda calculus is, as its name suggests, a calculus. A calculus is a system of manipulating symbols, which by themselves don't have any semantic meaning, in a way that is somehow meaningful. We all know algebra. Algebra itself doesn't have an innate meaning, but we can use it to represent and solve real world problems. Algebra, however, is limited. Not every problem can be represented in algebra. There are many branches of mathematics that use different systems. One example would be formal logic, which is used for logical operations on booleans. Another such example is lambda calculus.

2.1 A short history

People always trusted mathematics to be true and relied on it heavily. If something was proven true with mathematical logic, then that must be true. However, starting from the late 19th century, people ran into paradoxes. People made a distinction between reasoning that is rigorous and reasoning that isn't—reasoning that is logical, and reasoning that is psychological. The fact that mathematical logic, which people looked at for rigorousness, is infested with paradoxes and self-reference was very troubling for people at the time.

The concept of mathematics and mathematical logic wasn't well defined, so people started to think about the formalisation of mathematical logic to try to solve these issues. People wanted a system that would encapsulate all of mathematical logic. Preferably this system would be simple, clean and intuitive.

Throughout the late 19th and early 20th centuries, people started formally defining and redefining different aspects of mathematics. Frege [4] wrote about propositional calculus and functions as graphs, and in doing so re-evaluated the concept of functions and was already using concepts like Currying functions (more on this in section 2.2) without really giving thought to it. Peano [6] invented the Peano axioms and Peano arithmetic as a way of defining natural numbers. He was not the first to attempt defining natural numbers, but he was the most successful. Schönfinkel [8] invented combinatory logic, which was later rediscovered and improved on by Curry [3], as a way to remove the need for quantified variables in logic.

One major attempt to define all of mathematics was done by Russell and Whitehead [7]. They wrote a book that would become well known in all of mathematics and logic. This book is called *Principia Mathematica*. They did, however, run into a few problems, which arose from self-reference. To solve these problems that this paradoxical self-reference brought with it, they invented an elaborate system, the theory of types, to circumvent/eliminate it. It was a very carefully crafted bastion against self-reference ever coming up in their system, which was not very simple, clean or intuitive.

People praised *PM* as they thought they had finally done it; they had formalised all of mathematical logic, they had realised the dream of grounding all of mathematics in logic. But in Vienna, Gödel was sceptical of this book. He started seeing some cracks, he felt that there was something wrong about this attempt. Gödel felt that self-reference was a fundamental part

of mathematical logic. Then he went out and actually proved that there is no consistent system of axioms whose theorems can be listed by an effective procedure that is capable of proving all truths about arithmetic of natural numbers¹ [5], meaning that such a system is either inconsistent or incomplete², greatly disturbing many mathematicians and upending mathematics as they knew it.

During this time, in this environment of the formalisation of mathematical logic, Church [1] invented the lambda calculus. Lambda calculus is a very simple and minimalistic system of substitution. A little while later, Turing [11, 12] invented Turing machines. Turing machines are conceptual mathematical machines that function based on state—they were state machines. These could perform all kinds of mathematical and logical computations. He was not the first to invent computers, but he was the first to work them out as well as he did (and, as you probably know, he built one which he cracked the German enigma code with).

There was this problem that has a few different names. It is often known as the *halting problem* or the *Entscheidungsproblem*, which is German for *decision problem*. The halting problem and decision problem aren't exactly synonyms, but they come down to the same thing. Basically, it asks whether it is possible to know via an algorithm whether a computation will complete execution or result in an infinite loop. In 1936, Turing [11, 12] spent a long time proving, using his Turing machines, that this isn't possible, but it didn't get published until early 1937. Also in 1936, Church [2] proved the same thing using lambda calculus and happened to publish it before Turing did. When Turing finally got around to publishing his proof, he found out that he was beaten to it by Church. He wasn't too pleased. What is interesting, though, is that lambda calculus and Turing machines take two completely different approaches. Turing machines function entirely on state, while lambda calculus is completely stateless (we'll look at this later). Turing thought this was interesting too, so he researched lambda calculus and how it relates to his Turing machines, and proved that they are formally equivalent [10].

Why do I tell you all this? Well, your main takeaway should be that even though lambda calculus is a very simple system, which at first glance might not seem to be very semantic or seem to have any real world implications, it actually is Turing complete. Lambda calculus and Turing machines take wildly different approaches: one state based, the other stateless. Another difference is that Turing machines can be physically built. We can, however, use lambda calculus on these Turing machines *and* simulate Turing machines with lambda calculus, which is a fundamental part to the thesis of this work. We will look at lambda calculus and how the work of all the previously mentioned mathematicians, and many more, can be applied in lambda calculus to get a useful system.

2.2 The syntax

Lambda calculus is all about first-class higher-order pure (anonymous³) unary functions. Such a function takes a single input, and returns a single expression that is only dependent on the input, so it doesn't have any outside state. Such a function can take and return any expression, which in lambda calculus is always a function. A simple function definition in lambda calculus looks as follows:

$$\lambda a.a$$

The lambda signifies a function. Everything following it will be part of that function's definition. The a before the $.$ is the name of the argument. There is only one, because, as I said before, all functions in lambda calculus are unary. Everything following the $.$ is part of the function body, which is the return expression. The function above is the identity function in lambda calculus; it just returns the input. This is the equivalent of multiplying by one, or

¹This is a definition of the first incompleteness theorem I got from Wikipedia [14]

²Incompleteness means that there are things that are true, but are not provable.

³The core lambda calculus has no way of naming functions.

defining a function like $f(x) = x$, or multiplying a vector with the identity matrix; it does nothing.

But how do we use this function? Well, just like defining a function, it is quite simple. If you want to apply this function to a symbol, you just put it in front of the symbol. Something like this:

$$(\lambda a.a)x$$

Which evaluates to x , because you remove the x and then replace all the a 's in the function body with x and then remove the function signifier and argument list. It all comes down to a simple process of substitution.

In this case you need parentheses around the function, otherwise x would be considered part of the function's body, which it isn't. It's also important to note that lambda calculus is left-associative, that is, it evaluates an expression from left to right. This means that the function on the far left of an expression gets invoked first.

I have now basically explained the entire lambda calculus, it is really that simple. I have explained abstraction (functions), application (applying functions), and grouping (parentheses), which is basically all we need. You can also give names to expressions. We could name our identity function I as follows:

$$I := \lambda a.a$$

But this isn't really part of the core lambda calculus any more, just some syntactic sugar. This way, instead of constantly having to write $\lambda a.a$, we can just write I . So instead of writing:

$$(\lambda a.a)x = x$$

We could use our previous definition of I and write:

$$Ix = x$$

We have now covered identifiers too.

But if this is all there is, how can this possibly be Turing complete? How do we do boolean logic, or algebra? How can we do things with only unary functions? What are a and x supposed to represent? If there is no concept of value, how do we even use this meaningfully? Well, the key is this: a function can return any expression (remember?), which is always a function⁴, not just a single symbol. We can start combining these simple functions into more complex functions. Let's say that we wanted to have a function that takes two arguments, and then applies the first argument to the second one. You are probably asking yourself a few questions. For example, what does it mean for one argument to be applied to another? Well, as I said, everything is a function. But the biggest question you are probably asking yourself is: how can you have a function that takes two arguments?

We actually can't, but what we can do is to have a function that takes one argument and returns another function that takes one argument. We can define that function as follows:

$$\lambda a.\lambda b.ab$$

We currently have a function definition inside the body of another function. If we now apply this function to a symbol like x , we get this:

$$(\lambda a.\lambda b.ab)x = \lambda b.xb$$

We get a new function that takes an argument and applies x to it. If we now apply this function to a symbol like y , we get this:

$$(\lambda b.xb)y = xy$$

⁴Everything is.

Alternatively, we could write it all on one line:

$$(\lambda a.\lambda b.ab)xy = (\lambda b.xb)y = xy$$

xy in this case is what we would call the β -normal form of the preceding expressions. That just means that it is in the simplest form and isn't able to be evaluated any further. Reducing a lambda expression to the β -normal form is called β -reduction.

You can start to see how we can combine unary functions to create more complex functions⁵. In this example we used two nested unary functions to get the same result you would with a binary function. Such a nested function is often called a *Curried function*. You might think to yourself that having this many nested functions can be quite convoluted and not very readable, and you're quite right. That's why people often use a shorthand notation. They would basically write it as if it is a single binary function (as with any n-ary function). They would write the example function above as:

$$\lambda ab.ab$$

Do keep in mind, that even though this looks and, for the most part, acts as if it is a single binary function, it really isn't. It still is a curried function that feeds in the arguments one by one, but this way the expression becomes more readable and easier to think about conceptually. We will use this notation from now on.

Congratulations, you now know the very basics of lambda calculus. You may still not see how this is Turing complete or how this can be useful and meaningful. You might also already see some of the intrigues of lambda calculus; how simple it is, how it doesn't have a concept of value or data, how everything is an expression, how it is stateless, etc. But we'll get to all of that eventually. If you get this, everything else will follow naturally (mostly).

2.3 Combinatory logic

Combinatory logic is a notation to eliminate the need for quantified variables in mathematical logic [13]. That basically means a form of logic without values, just like with lambda calculus, but just pure logical expressions, using so called *combinators*. The idea of combinators first came from Schönfinkel [8], and was later rediscovered by Curry [3]. Combinators are just symbols, in this case letters, that perform operations on symbols that succeed it. We've actually looked at one of these combinators already.

We will be using Curry's names for combinators, since his names are most widely used.

2.3.1 Identity

The first combinator we'll cover is I . It does exactly the same thing as our I function we defined in lambda calculus in the previous subsection ($I := \lambda a.a$). In fact, all combinators can be defined in lambda calculus. Lambda calculus is really just 90% combinatory logic, but without identifiers. This combinator may seem quite useless, but it is actually quite useful when composing combinators, which we'll come to soon.

2.3.2 The omega combinator

The next combinator we'll cover is M . All it does is repeat its one argument twice. It can be defined in lambda calculus as follows:

$$M := \lambda f.f f$$

We could, for example, look at what happens when you apply M to I . We get:

$$MI = II = I$$

⁵That's what makes them higher-order (and first-class).

Or, written out in lambda calculus:

$$(\lambda f.f f)\lambda a.a = (\lambda a.a)\lambda a.a = \lambda a.a$$

What happens if you apply M to M ? You get:

$$MM = MM = MM = \dots$$

ad infinitum. Or in lambda calculus:

$$(\lambda f.f f)\lambda f.f f = (\lambda f.f f)\lambda f.f f = (\lambda f.f f)\lambda f.f f = \dots$$

This expression cannot be evaluated. We say that it doesn't have a β -normal form. In lambda calculus and combinatory logic not every expression is reducible. As we've seen in the second to last paragraph of section 2.1: there is no single algorithm to decide whether a lambda expression has a β -normal form⁶.

MM is sometimes called the Ω combinator. Omega, because it is the end of the Greek alphabet. The M combinator is sometimes called the ω combinator because of this. Combinators often have many different names. Sometimes because scientists discovered them separately, unaware of each other, sometimes because they preferred a different name, sometimes because scientists like to give them pet names⁷.

The omega combinator will be discussed further in section 3.2.2.

2.3.3 The constant combinator

The next combinator we'll cover is K . It is a combinator that takes two arguments and returns the first. We can easily define it in lambda calculus as follows:

$$K := \lambda ab.a$$

Remember that we defined this as a curried function. This means we can give it just one argument and get a new function out of it. Let's say we apply K to 5:

$$K5 = (\lambda ab.a)5 = \lambda b.5$$

Our new function, $K5$, is a function that takes an argument and returns 5. This means that whatever we apply this function to, we always get 5. K gets its name from the German word *Konstant*, meaning constant. You can probably see why.

Just like with the previous combinators, it'll prove very useful, much more so than you'd expect.

2.3.4 The kite

Here is where things get a little spicier. Our next combinator is KI . It takes two arguments and returns the latter. We can define it in lambda calculus as follows:

$$KI := \lambda ab.b$$

You may already be thinking about its name. Why does it have two letters? And why are they two letters we've talked about already? Well, the answer is very simple. If you apply K to I , you get KI . Don't believe me? Let's try!

If we use our definition of KI and apply it to xy we get:

$$KIxy = (\lambda ab.b)xy = y$$

⁶This is not exactly what is written, but it means the same thing.

⁷I have a theory they are just trying to throw us off

But if we use the K and I combinators separately, we get the following:

$$K I x y = (\lambda a b. a) I x y = (\lambda b. I) x y = I y = y$$

If you think about it, it is very logical. If K takes two arguments and returns the first, then, in this case, it uses up both I and x and returns I , which will just return the next argument, in this case y . $K I$ will always return the second symbol after the I , because—again—the first gets used up by K .

We can also just see what function we get when we apply K to I :

$$K I = (\lambda a b. a) I = \lambda b. I = \lambda b. \lambda a. a = \lambda b a. a$$

We get our definition of $K I$ (except the names of the arguments are switched).

We're starting to define combinators as combinations of other combinators. Every combinator, in fact, can be defined as a combination of other combinators. That's why they are called combinators.

2.3.5 The flip combinator

The next combinator is the C combinator. The C combinator is definable in lambda calculus as:

$$C := \lambda f a b. f b a$$

What it basically does is switch the arguments to the next combinator around.

If we apply C to K and two random symbols, we get the same result we would get if we had applied $K I$ to those same symbols:

$$C K x y = K y x = y$$

$$K I x y = y$$

$$C K = K I$$

Let's see what happens when we apply C to K in lambda calculus (I have changed the names of K 's arguments as to avoid confusion with those of C):

$$(\lambda f a b. f b a) \lambda x y. x = \lambda a b. (\lambda x y. x) b a$$

We don't get our exact definition of $K I$. But we can see that for every input, $C K$ and $K I$ *always* produce the same output. We say that these functions are *extensionally equal*—they have been defined separately and we cannot rewrite one to the other, but we know that they produce the same results, so they must be equal.

You can do the same thing to find out that $C K I = K$. It really does make sense. K and $K I$ both "select" one of two arguments. One selects the first, the other selects the latter. Flipping their arguments make them select the opposite of what they normally would, so they select the argument that the other combinator usually would.

2.3.6 The composition combinator

Our next combinator, B , is defined as follows:

$$B := \lambda f g a. f (g a)$$

It applies g to a before applying f to the result. This combinator is used for function composition. When a function applies g to a and f to the result, we say that this function

composes f with g . Function composition is used to apply a sequence of operations on a variable (but the order in which you write them is reversed).

In mathematics we usually write function composition as follows:

$$f \circ g$$

So

$$(f \circ g) a = f(ga)$$

If we define a function h as

$$h = f_4 \circ f_3 \circ f_2 \circ f_1$$

it means that h applies f_1 , f_2 , f_3 and f_4 in that order. It is really important to keep in mind this reversed order, or you are in for a confusing ride.

In combinatory logic we write function composition of f and g as follows:

$$Bfg$$

which reduces to

$$(\lambda fga.f(ga))fg = \lambda a.f(ga)$$

in lambda calculus.

2.3.7 The thrush

Our next combinator is T_h . It is defined as follows:

$$T_h := \lambda af.f a$$

It swaps around two functions. It is basically a very simple form of storage; when you apply it to an argument, it stores that argument to have some other function applied to it later.

2.3.8 The vireo

The V -combinator is basically the T_h -combinator with an extra argument. If you consider T_h a variable, V could be considered a *pair* (section 3.1.4).

It is defined as

$$V := \lambda abf.fab$$

2.3.9 The starling

Our last combinator is S . It can be defined as follows

$$S := \lambda fga.f a(ga)$$

It applies f to a and its result to the result of the application of g to a .

2.3.10 S and K

As I've said, every combinator can be defined as a combination of other combinators. The question arises: how many combinators do we need to define every other combinator? It turns out you need just two. You can define every other combinator using just S and K .

2.3.11 To Mock a Mockingbird

At the start of this section about combinatory logic, I said that Schönfinkel [8] invented combinatory logic as a way of removing the need for quantified variables. He started with propositional logic and stripped it down until there was a very pure and simple form of logic left. But how can we use this form of logic in the real world, if he even removes things like propositions? You already know that it is Turing complete, so it must be able to do any computation, but I haven't explain how yet (see section 3). But we can use combinatory logic in the real world already.

You may have noticed some of the previous subsections have bird names as titles. This is because they are the names given to the combinators, discussed in the respective subsections, by an author named Smullyan. He is a mathematician who likes to write puzzle books. His book *To Mock a Mockingbird* [9] is practically a large metaphor for combinatory logic. There are some unrelated puzzles in the beginning of the book, but the rest is about a big forest with birds. The birds represent the combinators. The begin letters of the bird names are the names of the combinators. The way the birds interact reflects the way the combinators interact. The reason he chose birds for his metaphor is that Curry was an avid bird watcher. If you feel like you still don't understand the notation of combinatory logic completely, I would recommend you give this book a read, because it explains it very simply and clearly.

In Smullyan's world, there is a forest with birds. These birds represent combinators. If you call out a bird name to a bird, it will give you the name of another bird. If you give a bird a bird name of a bird that is present, it will give you back a name of a bird that is present. Calling out bird names to birds represents application.

I think it would be fun if we had a look at one of the puzzles to see if we can solve it using our newfound knowledge of combinatory logic. I think the first puzzle is sufficiently interesting. So far, Smullyan has only introduced the mockingbird, which is the omega operator (section 2.3.2), and the idea of function composition, but not yet the bluebird, which is the composition operator (section 2.3.6). I have taken the puzzle directly from the book:

It could happen that if you call out B to A , A might call the same bird B back to you. If this happens, it indicates that A is fond of the bird B . In symbols, A is fond of B means that: $AB = B$

We are now given that the forest satisfies the following two conditions.

C_1 (the composition condition): For any two birds A and B (whether the same or different) there is a bird C such that for any bird x , $Cx = A(Bx)$. In other words, for any birds A and B there is a bird C that composes A with B .

C_2 (the mockingbird condition): The forest contains a mockingbird M .

One rumor has it that every bird of the forest is fond of at least one bird. Another rumor has it that there is at least one bird that is not fond of any bird. The interesting thing is that it is possible to settle the matter completely by virtue of the given conditions C_1 and C_2 .

Which of the two rumors is correct?

Do note that in this case, C and B do not refer to the C and B combinators we've looked at previously.

The answer will be shown on the next page.

Because of C_1 and C_2 , we know that for every bird A , there's a bird—we'll call it C —that composes A with M . We can say the following:

$$Cx = A(Mx) = A(xx)$$

If we now fill in C in place of x , we get:

$$A(CC) = CC$$

We thus know that for any bird A , A is fond of the bird CC , where C composes A with M . Therefore rumour one is true and rumour two is false.

The answer Smullyan [9] gives is a bit more verbose, but it comes down to the same thing.

3 Using lambda calculus for computation

Now that we've gone through lambda calculus and combinatory logic, it is finally time to get to the good parts. I hope the way here wasn't too boring or difficult. We will now look at how to use lambda calculus for computation.

3.1 Church encoding

To do computation with lambda calculus, we need to be able to do a few things such as boolean logic and arithmetic. Lambda calculus itself doesn't have these things built in to it, but we can define things such as boolean logic and arithmetic in lambda calculus. In section 2.1, I talked about how mathematicians were defining everything in mathematics. We can build off of their work and see how we can implement their definitions in lambda calculus. This is exactly what Church did.

3.1.1 Simple boolean operations

Let's look at a simple form of computation before jumping into arithmetic. We'll start with binary logic. In it's simplest form, binary logic is really just control flow: *if A then B else C* ($A ? B : C$). We want to have some condition that chooses between two expressions. We know how to do that already (see section 2.3). K chooses the first of two expressions and KI the latter. We can define true to be K and false to be KI :

$$T := K = \lambda ab.a$$

$$F := KI = \lambda ab.b$$

But how can we do logic gates? Let's look at negation. We have already looked at the flip combinator— C (section 2.3.5), which does exactly what we want. Thus, we can say:

$$NOT := C = \lambda fab.fba$$

We could also do it another way. We already know how to do control flow, so we could define a function that chooses the opposite of what the input is. In simple programming terms:

$$\text{if } p \text{ then } F \text{ else } T$$

Or in a C-like expression:

$$p ? F : T$$

Or in lambda calculus:

$$NOT := \lambda p.pFT$$

It depends which one's preferable. The C combinator is a bit more elegant and performant, because it takes less function applications, but you get a function that is only extensionally equal to T or F , while the other definition literally returns T or F .

How do we define AND ? We could define it in simple programming terms, and then translate it to lambda calculus. In simple programming terms, AND would look like this:

if p then (if q then T else F) else f

$p \text{ ? } q \text{ ? } T : F : F$

In lambda calculus we would get:

$AND := \lambda pq.p(qTF)F$

But this is a very naïve way of defining AND . If you look closely at the qTF part, you notice that it actually just returns whatever q is anyway. Thus, we can say:

$AND := \lambda pq.pqF$

You can also replace the remaining F with q if you want, because the F will only be returned if q is F —in other words: q in that case is always false. So we can say:

$AND := \lambda pq.pqp$

This is the equivalent of:

if p then q else p

$p \text{ ? } q : p$

We can define OR in a very similar way. If the first argument is true, we just return it, else we return the second argument:

$OR := \lambda pq.ppq$

Defining XOR is very easy too. If the first argument is true, then we want to return what the second argument is not, else we want to return what the second argument is. In lambda calculus:

$XOR := \lambda pq.p(NOT\ q)q$

Giving:

$XOR := \lambda pq.p(qFT)q$

Alternatively:

$XOR := \lambda pq.p((\lambda fab.fba)q)q = \lambda pq.p(\lambda ab.qba)q$

We can define boolean equality (BEQ) similarly:

$BEQ := \lambda pq.pq(qFT)$

Or:

$BEQ := \lambda pq.pq(\lambda ab.qba)$

Which you can read as: "If p is true, then return what q is, else return what q is not."

Let's look at an example expression and solve it in lambda calculus. We'll solve the following expression:

$!(x \ \&\& \ y) == !x \ || \ !y$

If you're not familiar with C-like expressions: $!$ means NOT, $\&\&$ means AND, $==$ means BEQ and $||$ means OR. $!$ has a higher and $==$ a lower precedence than the rest.

We can write and reduce this expression using the functions/combinators we defined (using $NOT := \lambda p.pFT$, because else we wouldn't be able to write out NOT without using lambda calculus):

$$\begin{aligned}
& BEQ (NOT (AND x y)) (OR (NOT x) (NOT y)) \\
&= BEQ (NOT (xyx)) (OR (xFT) (yFT)) \\
&= BEQ (xqxFT) (xFT (xFT) (yFT)) \\
&= xyxFT (xFT (xFT) (yFT)) (xFT (xFT) (yFT) FT)
\end{aligned}$$

If we've done this correctly, then according to *De Morgan's Law*, we should always get T for any substitution of x and y for T and F in this final expression.

$x = F$ and $y = F$ gives us:

$$\begin{aligned}
& FFFFT (FFT (FFT) (FFT)) (FFT (FFT) (FFT) FT) \\
&= FFT (TTT) (TTTFT) \\
&= TT (TFT) \\
&= TTF \\
&= T
\end{aligned}$$

$x = F$ and $y = T$ gives us:

$$\begin{aligned}
& FTFFFT (FFT (FFT) (TFT)) (FFT (FFT) (TFT) FT) \\
&= FFT (TTF) (TTFFT) \\
&= TT (TFT) \\
&= TTF \\
&= T
\end{aligned}$$

$x = T$ and $y = F$ gives us:

$$\begin{aligned}
& TFTFT (TFT (TFT) (FFT)) (TFT (TFT) (FFT) FT) \\
&= FFT (FFT) (FFFTFT) \\
&= TT (TFT) \\
&= TTF \\
&= T
\end{aligned}$$

$x = T$ and $y = T$ gives us:

$$\begin{aligned}
& TTTFT (TFT (TFT) (TFT)) (TFT (TFT) (TFT) FT) \\
&= TFT (FFF) (FFFFFT) \\
&= FF (FFT) \\
&= FFT \\
&= T
\end{aligned}$$

We can also try to solve this expression in lambda calculus (using the shorthand notation

and using $NOT := C$, because it is easier to write out):

$$\begin{aligned}
& \lambda xy.(\lambda pq.pq(\lambda cd.qdc))((\lambda fab.fba)((\lambda pq.pqp)xy))((\lambda pq.ppq)((\lambda fab.fba)x)((\lambda fab.fba)y)) \quad (1) \\
& = \lambda xy.(\lambda pq.pq(\lambda cd.qdc))((\lambda fab.fba)(xyx))((\lambda pq.ppq)(\lambda ab.xba)(\lambda ab.yba)) \quad (2) \\
& = \lambda xy.(\lambda pq.pq(\lambda cd.qdc))(\lambda ab.xyxba)((\lambda ab.xba)(\lambda ab.xba)(\lambda ab.yba)) \quad (3) \\
& = \lambda xy.(\lambda pq.pq(\lambda cd.qdc))(\lambda ab.xyxba)(x(\lambda ab.yba)(\lambda ab.xba)) \quad (4) \\
& = \lambda xy.(\lambda ab.xyxba)(x(\lambda ab.yba)(\lambda ab.xba))(\lambda cd.x(\lambda ab.yba)(\lambda ab.xba)dc) \quad (5) \\
& = \lambda xy.xy x(\lambda cd.x(\lambda ab.yba)(\lambda ab.xba)dc)(x(\lambda ab.yba)(\lambda ab.xba)) \quad (6) \\
& \equiv \lambda xy.xy x(\lambda cd.x(\lambda ab.yba)(\lambda ab.a)dc)(x(\lambda ab.yba)(\lambda ab.a)) \quad (7) \\
& \equiv \lambda xy.xy x(xy(\lambda ab.b))(x(\lambda ab.yba)(\lambda ab.a)) \quad (8) \\
& \equiv \lambda xy.xy x(xy(\lambda ab.b))(\lambda ab.a) \quad (9) \\
& \equiv \lambda xy.xy x(\lambda ab.a)(\lambda ab.a) \quad (10) \\
& \equiv \lambda ab.a \quad (11)
\end{aligned}$$

We know that expressions 6 and 7 are extensionally equal;

$$\lambda x.x(\lambda ab.yba)(\lambda ab.xba) \equiv \lambda x.x(\lambda ab.yba)(\lambda ab.a)$$

because x picks between $\lambda ab.yba$ and $\lambda ab.xba$. $\lambda ab.xba$ only gets picked when x is F , so $\lambda ab.xba$ is always $\lambda ab.Fba = \lambda ab.a$. Therefore, a $x(\lambda ab.yba)(\lambda ab.xba)$ that occurs at the start of its grouping (we do need to keep left-associativity in mind) can be replaced with $x(\lambda ab.yba)(\lambda ab.a)$.

We also know that expressions 7 and 8 are extensionally equal. We know that

$$\lambda xcd.x(\lambda ab.yba)(\lambda ab.a)dc$$

selects between

$$\lambda cd.(\lambda ab.yba)dc = \lambda cd.ycd \equiv y$$

and

$$\lambda cd.(\lambda ab.a)dc = \lambda cd.d = \lambda ab.b$$

therefore

$$\lambda xcd.x(\lambda ab.yba)(\lambda ab.a)dc \equiv \lambda x.xy(\lambda ab.b)$$

Expressions 8 and 9 are also extensionally equal, because xyx selects between the expressions $xy(\lambda ab.b)$ and $x(\lambda ab.yba)(\lambda ab.a)$. The expression $x(\lambda ab.yba)(\lambda ab.a)$ only gets selected if $xyx = F$. There are two ways for $xyx = F$ to be true; either $x = F$, or both $x = T$ and $y = F$. $x = F$ gives

$$x(\lambda ab.yba)(\lambda ab.a) = F(\lambda ab.yba)(\lambda ab.a) = \lambda ab.a$$

and $x = T$ and $y = F$ gives

$$x(\lambda ab.yba)(\lambda ab.a) = T(\lambda ab.Fab)(\lambda ab.a) = \lambda ab.Fab = \lambda ab.a$$

meaning that for all the possibilities of

$$xyx(xy(\lambda ab.b))(x(\lambda ab.yba)(\lambda ab.a)) = x(\lambda ab.yba)(\lambda ab.a)$$

also

$$x(\lambda ab.yba)(\lambda ab.a) \equiv \lambda ab.a$$

meaning

$$\lambda xy.xy x(xy(\lambda ab.b))(x(\lambda ab.yba)(\lambda ab.a)) \equiv \lambda xy.xy x(xy(\lambda ab.b))(\lambda ab.a)$$

We can apply the same reasoning to prove that equations 9 and 10 are extensionally equal. We know xyx picks between $xy(\lambda ab.a)$ and $\lambda ab.a$. For xyx to pick the first, xyx must equal T , which is only possible if both $x = T$ and $y = T$. Which gives

$$xy(\lambda ab.a) = TT(\lambda ab.a) = T = \lambda ab.a$$

The last step should be quite self-explanatory; xyx can only select between $\lambda ab.a$ and $\lambda ab.a$.

3.1.2 Natural numbers

Now that we've covered boolean logic, it is finally time to move on to arithmetic. Remember that in section 2.1, I said that Peano [6] defined natural numbers? He basically defined zero and then defined every number as a successor of the previous number, starting from zero. He also defined the properties of natural numbers. We can do exactly the same in lambda calculus. The key concept is function composition (section 2.3.6).

We'll define functions that compose a given function f , n times with themselves, where n is the natural number the function is supposed to represent. Thus we'll say

$$\begin{aligned} N0 &:= \lambda f a. a \\ N1 &:= \lambda f a. f a \\ N2 &:= \lambda f a. f(f a) \\ N3 &:= \lambda f a. f(f(f a)) \\ N4 &:= \lambda f a. f(f(f(f a))) \end{aligned}$$

and so on, or in a more standard mathematical notation

$$\begin{aligned} 0f &:= 0 \\ 1f &:= f \\ 2f &:= f \circ f \\ 3f &:= f \circ f \circ f \\ 4f &:= f \circ f \circ f \circ f \end{aligned}$$

You can also say

$$\begin{aligned} N0 &:= F \\ N1 &:= \lambda f. f = I \\ N2 &:= \lambda f. B f f \\ N3 &:= \lambda f. B(B f f) f \\ N4 &:= \lambda f. B(B(B f f) f) f \end{aligned}$$

We can't define an infinite number of functions by hand, so we'll use this shorthand definition:

$$n := \lambda f a. f^{\circ n} a$$

meaning: applying a number n to a function f is the same as composing f with f , n times. This is meaningful, because it allows us to do something a given number of times.

We can use this in an example. Let's say we wanted to do a negation multiple times. We can use our newly defined numbers:

$$\begin{aligned} N0 \text{ NOT } T &= T \\ N1 \text{ NOT } T &= \text{NOT } T = F \\ N2 \text{ NOT } T &= \text{NOT } (\text{NOT } T) = T \\ N3 \text{ NOT } T &= \text{NOT } (\text{NOT } (\text{NOT } T)) = F \end{aligned}$$

Numbers are just function composition. Function composition, and thus the B combinator (section 2.3.6), is fundamental to all arithmetic in lambda calculus. What happens when we

compose two numbers, say $N3$ with $N3$?

$$\begin{aligned}
& B \ N3 \ N3 \\
&= (\lambda f g h. f(gh)) \ N3 \ N3 \\
&= \lambda h. N3 \ (N3 \ h) \\
&= \lambda h. N3 \ ((\lambda f b. f(f(b)))h) \\
&= \lambda h. N3 \ (\lambda b. h(h(hb))) \\
&= \lambda h. (\lambda f a. f(f(fa))) (\lambda b. h(h(hb))) \\
&= \lambda h. \lambda a. (\lambda b. h(h(hb))) ((\lambda b. h(h(hb))) ((\lambda b. h(h(hb))) a)) \\
&= \lambda h a. (\lambda b. h(h(hb))) ((\lambda b. h(h(hb))) (h(h(ha)))) \\
&= \lambda h a. (\lambda b. h(h(hb))) (h(h(h(h(h(h(ha))))))) \\
&= \lambda h a. h(h(h(h(h(h(h(h(h(h(ha)))))))) \\
&= N9
\end{aligned}$$

We can tell that composition with natural numbers is the same as multiplication, which makes a lot of sense, because function composition is associative. The ninefold composition of f is the same as the threefold composition of the threefold composition of f :

$$f \circ f \circ f \circ f \circ f \circ f \circ f \circ f \circ f = (f \circ f \circ f) \circ (f \circ f \circ f) \circ (f \circ f \circ f)$$

We know that if we were to define a *MULT* combinator, the following should be true:

$$MULT \ N3 \ N3 \ f \ a = (f \circ f \circ f \circ f \circ f \circ f \circ f \circ f \circ f) a$$

and in extension

$$\begin{aligned}
MULT \ N3 \ N3 \ f \ a &= ((f \circ f \circ f) \circ (f \circ f \circ f) \circ (f \circ f \circ f)) a \\
&= ((N3 \ f) \circ (N3 \ f) \circ (N3 \ f)) a \\
&= N3 \ (N3 \ f) \ a \\
MULT \ N3 \ N3 \ f &= N3 \ (N3 \ f) \\
&= B \ N3 \ N3 \ f \\
MULT &= B
\end{aligned}$$

thus we can say

$$MULT := B = \lambda n k f. n(kf)$$

What happens when we apply $N3$ to $N2$? We get the following:

$$\begin{aligned}
& N3 \ N2 \\
&= \lambda f a. f(f(fa)) \ N2 \\
&= \lambda a. N2(N2(N2 \ a)) \\
&= \lambda a. N2(N4 \ a)) \\
&= \lambda a. N8 \ a \\
&\equiv N8
\end{aligned}$$

in other "words"

$$\begin{aligned}
& 3 \ 2 \\
&= 2 \circ 2 \circ 2 \\
&= 8
\end{aligned}$$

so application of natural numbers is the same exponentiation, but the numbers are reversed. Thus, we can say:

$$POW := \lambda nk.kn$$

which is just our T_h combinator (section 2.3.7).

$$POW := T_h$$

How do we do addition? Let's start with adding one—in other words—finding the successor. All we need to do is compose the function once more. We could define the *SUCC* function as follows:

$$SUCC := \lambda nfa.f(nfa)$$

Alternatively:

$$SUCC := \lambda nf.Bf(nf)$$

which some find prettier, because it makes it clear that we're doing function composition, but it also takes a bit longer to compute, because the second definition reduces down to the first:

$$\begin{aligned} & \lambda nf.Bf(nf) \\ = & \lambda nf.(\lambda gha.g(ha))f(nf) \\ = & \lambda nf.\lambda a.f((nf)a) \\ = & \lambda nfa.f(nfa) \end{aligned}$$

so if you use the second, you need to perform that reduction every time you invoke it.

If we now try to find the successor of $N3$, we get

$$\begin{aligned} & SUCC\ N3 \\ = & (\lambda nfa.f(nfa))\ N3 \\ = & \lambda fa.f(N3\ fa) \\ = & \lambda fa.f((\lambda hb.h(hb)))fa \\ = & \lambda fa.f(f(f(fa))) \\ = & N4 \end{aligned}$$

or

$$\begin{aligned} & SUCC\ N3 \\ = & (\lambda nf.Bf(nf))\ N3 \\ = & \lambda f.Bf(N3\ f) \\ = & \lambda f.Bf((\lambda ha.h(ha)))f \\ = & \lambda f.Bf(\lambda a.f(f(fa))) \\ = & \lambda f.Bf(\lambda a.f(f(fa))) \\ = & \lambda f.(\lambda ghb.g(hb))f(\lambda a.f(f(fa))) \\ = & \lambda f.(\lambda b.f((\lambda a.f(f(fa)))b)) \\ = & \lambda f.\lambda b.f(f(f(fb))) \\ = & \lambda fa.f(f(f(fa))) \\ = & N4 \end{aligned}$$

To add two numbers, we can now just call this successor function multiple times. Say we wanted to add three to four. We can call the successor function three times on four:

$$\begin{aligned}
& N3 \text{ } SUCC \text{ } N4 \\
&= SUCC (SUCC (SUCC \text{ } N4)) \\
&= SUCC (SUCC \text{ } N5) \\
&= SUCC \text{ } N6 \\
&= N7
\end{aligned}$$

Thus we can say

$$ADD := \lambda nk.(n \text{ } SUCC \text{ } k)$$

SUCC is really just an infix *ADD*.

We could also say that if we want to add two numbers n and k , we can compose a function n times and k times and compose the results to get a $(n + k)$ -fold composition of the function. In lambda calculus:

$$ADD := \lambda nkf.B(nf)(kf)$$

or

$$ADD := \lambda nkfa.(nf)(kfa)$$

which looks a lot like our *SUCC* function, just with an extra argument to decide the number of times to compose the function, and it's prefix instead of infix. In a more mathematical notation, you can say

$$ADD \text{ } n \text{ } k \text{ } f := f^{\circ n} \circ f^{\circ k}$$

3.1.3 Boolean comparison

The C programming language has only added booleans in the C99 standard. In C, booleans are really just integers. All boolean expressions are really just arithmetic. In C, *false* is synonymous with 0 and every non-zero number means *true*. In lambda calculus, this is the same. Our definition of *N0* is exactly the same as our definition of *F*. This section talks about things that are important to both boolean logic and arithmetic, which are really just the same.

I want to define a function that takes a church numeral and returns *T* if it's *N0* or *F* if it's not. I will call this function *ISZERO*. To do this, we need to remember what a church numeral really is; it's a function that takes a function and an argument and applies that function a given number of times to that argument. The unique thing about *N0* is that it applies that given function zero times to the given argument, meaning it just returns the argument. That means that if we give *T* as the second argument, *N0* will return *T*. So far, we have this:

$$ISZERO := \lambda n.n...T$$

We still need to fill in the dots. If n isn't *N0*, whatever is on the dots will be applied to *T*, n times. We want this to always return *F*. We know a function that can do this: the constant function (section 2.3.3). Thus, we get:

$$ISZERO := \lambda n.n(KF)T$$

Here is an example for if it is not yet clear to you why we use the constant combinator:

$$ISZERO \text{ } N3 = (\lambda n.n(KF)T)N3 = N3(KF)T = K \text{ } F \text{ } (KF(KF \text{ } T)) = F$$

3.1.4 Data structures

A fundamental concept in programming is that of data structures. Data structures are ways of storing data together. There are different data structures, which are sort of like different data "layouts". You could argue that one combinator we've looked at, the thrush (section 2.3.7), is basically a data structure already; it takes an argument and holds on to it until you give it a function to apply to it. Although this isn't really a data structure yet; it only stores one thing. We can turn it into a data structure by "upgrading" it; let's give it another argument.

$$V := \lambda abf.fab$$

Which we have already seen in section 2.3.8.

We have basically defined something that is known as a (Church) pair. It holds on to two things. We could put two values into it:

$$Vxy = \lambda f.fxy$$

If we want to look at the first value, we can input K :

$$(\lambda f.fxy) K = Kxy = x$$

If we want to look at the second value, we can input KI :

$$(\lambda f.fxy) KI = KIx y = y$$

Thus, we could define the following functions:

$$FST := \lambda n.nK$$

$$SND := \lambda n.n(KI)$$

From now on, we'll also use:

$$PAIR := V$$

Pairs are very powerful, we use them for all kinds of things. One thing we can do is create linked lists. We do this by putting a value in the first index of the pair and another pair in the second. You can continue this pattern to make infinitely large lists. We could, for example, make a list from $N1$ through $N4$:

$$PAIR N1 (PAIR N2 (PAIR N3 N4))$$

If we want to find the first index of the array, we would do:

$$FST (PAIR N1 (PAIR N2 (PAIR N3 N4))) = N1$$

For the second index, we would do:

$$\begin{aligned} & FST (SND (PAIR N1 (PAIR N2 (PAIR N3 N4)))) \\ &= FST (PAIR N2 (PAIR N3 N4)) \\ &= N2 \end{aligned}$$

We can easily write a function that prepends something to a list. The following function prepends a list with $N0$:

$$\lambda p.PAIR N0 p$$

Applying it gives:

$$\begin{aligned} & (\lambda p.PAIR N0 p) (PAIR N1 (PAIR N2 (PAIR N3 N4))) \\ &= PAIR N0 (PAIR N1 (PAIR N2 (PAIR N3 N4))) \end{aligned}$$

Appending to a list is a bit harder; it involves stepping through the entire list and changing the last pair to link to a new pair. We'll talk more about later, when we cover functional programming and meta programming.

3.1.5 Natural numbers continuation

We've covered addition, multiplication and exponentiation, which were quite simple, but what about subtraction, division and finding the square root? Well, in contrast to the operations we've covered, these are actually quite complex. We'll quickly look at subtraction, but it gets very complicated very quickly, and it's really out of the scope of this text to cover all of arithmetic in lambda calculus. I just wanted to show that it's possible and that lambda calculus is Turing complete.

Just like we defined a successor function before defining addition, we'll define a predecessor function before defining subtraction.

The reason subtraction is so much harder than addition, is that applying a function once more is very easy, but removing an application is hard. The trick is this: you don't remove an application, you reapply all the functions from the start, until you get to the number you want the predecessor of. In other words, you reapply all the functions, except for the last.

How do we know when to stop? If we create an algorithm that just reapplies our function until we get to the number we want to know the predecessor of, we will have already overshoot the predecessor when we get to that number. What we actually want to do, is not to create an algorithm that returns a number after each iteration, but to create a pair containing the new number and the previous number carried over from the last iteration. That way, we can compare the second number of the pair to the input, and if they are equal, we return the first number of the pair.

Let's start by defining a function that takes a pair, and returns the next pair. Given a pair, we can make a new pair, of which the first item is the same as the second item of the input pair, and the second item is the successor of the second item of the input pair.

$$NEXT := \lambda p. PAIR (SND p) (SUCC (SND p))$$

It can be reduced to

$$\begin{aligned} & \lambda p. PAIR (SND p) (SUCC (SND P)) \\ &= \lambda p. (\lambda abf. fab) (SND p) (SUCC (SND p)) \\ &= \lambda pf. f (SND p) (SUCC (SND p)) \\ &= \lambda pf. f (p KI) (SUCC (p KI)) \\ &= \lambda pf. f (p KI) ((\lambda nga. g(nga)) (p KI)) \\ &= \lambda pf. f (p KI) (\lambda ga. g((p KI)ga)) \\ &= \lambda pf. f(p(\lambda ab.b))(\lambda ga. g((p(\lambda ab.b))ga)) \end{aligned}$$

in lambda calculus.

Using this, we can quite easily define a predecessor function. If we want to find the predecessor of n , we can apply $NEXT$, n times to the pair $(N0, N0)$. The second item be equal to n , so the first item will be n 's predecessor.

$$PRED := \lambda n. FST (n NEXT (PAIR N0 N0))$$

which reduces to

$$\begin{aligned} & \lambda n. FST (n NEXT (PAIR N0 N0)) \\ &= \lambda n. n NEXT (PAIR N0 N0) K \\ &= \lambda n. n NEXT (\lambda f. f N0 N0) K \\ &= \lambda n. n NEXT (\lambda f. f(\lambda ba.a)(\lambda ba.a)) K \\ &= \lambda n. n(\lambda pf. f(p(\lambda ab.b))(\lambda ga. g((p(\lambda ab.b))ga)))(\lambda f. f(\lambda ba.a)(\lambda ba.a))(\lambda ab.a) \end{aligned}$$

in lambda calculus.

Just like how addition is just the successor function applied multiple times, subtraction is just the predecessor function applied multiple times.

$$SUB := \lambda nk.n \text{ PRED } k$$

which reduces to

$$\lambda nk.n(\lambda m.m(\lambda pf.f(p(\lambda ab.b))(\lambda ga.g((p(\lambda ab.b))ga)))(\lambda f.f(\lambda ba.a)(\lambda ba.a))(\lambda ab.a))k$$

There are many alternative ways of going about defining subtraction, but this is probably the most straightforward.

3.1.6 Other kinds of numbers

We have as of yet only covered natural numbers. They are the easiest to define and use. There are, however, many other types of numbers, such as integers, rationals, reals, complex numbers, etc. I will quickly cover how you'd go about defining them, but I won't go into it deeply. I have shown how natural numbers work, and using them with pairs you can create all sorts of numbers, but it can be slightly difficult to wrap your head around, and it isn't in the scope of this text. I just wanted to show how simple arithmetic works in lambda calculus and show it is possible. I am not, however, trying to define all arithmetic. We are just touching on the topic.

That said, the way you define these other, more complicated types of numbers is very elegant in my opinion, and I really wanted to show it. Another reason I wanted to show it, is that it makes use of pairs extensively. By showing these elegant solutions to defining more complicated types of numbers, I can really show the usefulness and power of pairs.

Let's start with the easiest of the bunch: integers. They are just like natural numbers, except they can be negative. There are different ways of achieving this. The simplest would be to define an integer as a pair of a boolean and a natural number, where the boolean decides whether the integer is positive or negative. You can then define special arithmetic operators for integers, based on those for natural numbers, to take into account the sign of the number (whether it's positive or negative).

$$(sign, n)$$

This makes multiplication very straightforward; you just perform a regular multiplication on the natural numbers in the pairs, and xor the signs.

$$MULT_s := \lambda ab.PAIR (XOR (FST a) (FST b)) (MULT (SND a) (SND b))$$

Addition is a bit more complicated, since it requires a lot of conditionals (section 3.1.3).

$$ADD_s := \lambda ab. \begin{cases} (T, ADD (SND a) (SND b)) & \text{if } a \geq 0 \wedge b \geq 0 \\ (F, ADD (SND a) (SND b)) & \text{if } a \leq 0 \wedge b \leq 0 \\ (T, SUB (SND a) (SND b)) & \text{if } a \geq 0 \wedge b \leq 0 \wedge |a| \geq |b| \\ (F, SUB (SND b) (SND a)) & \text{if } a \geq 0 \wedge b \leq 0 \wedge |a| < |b| \\ (T, SUB (SND b) (SND a)) & \text{if } a \leq 0 \wedge b \geq 0 \wedge |a| < |b| \\ (F, SUB (SND a) (SND b)) & \text{if } a \leq 0 \wedge b \geq 0 \wedge |a| \geq |b| \end{cases}$$

Converting this to pure lambda calculus will be left as an exercise to the reader.

This makes subtraction very simple though, it is just addition, but you negate the sign of one of the numbers.

$$\begin{aligned} NEG_s &:= \lambda a.PAIR (NOT (FST a)) (SND a) \\ SUB_s &:= \lambda ab.ADD_s a (NEG_s b) \end{aligned}$$

However, this is not the only way of defining integers. Instead of defining an integer as a pair of a boolean and a natural number, you define it as a pair of two natural numbers so that if you subtract the second from the first, you get your integer. In other words, an integer k is represented as a pair (a, b) , where a and b are natural numbers and $k = a - b$. This means there are multiple ways (infinite in fact) of representing a given integer, but it also means that our operators are greatly simplified. We can define our operators as follows

$$\begin{aligned} ADD_i &:= \lambda ab. PAIR (ADD (FST a) (FST b)) (ADD (SND a) (SND b)) \\ NEG_i &:= \lambda a. PAIR (SND a) (FST a) \\ SUB_i &:= \lambda ab. ADD a (NEG_i b) \\ MULT_i &:= \lambda ab. PAIR \\ &\quad (ADD (MULT (FST a) (FST b)) (MULT (SND a) (SND b))) \\ &\quad (ADD (MULT (FST a) (SND b)) (MULT (SND a) (FST b))) \end{aligned}$$

Multiplication is relatively complex, but it's still simpler and more elegant than our previous definition of addition using a cubic ton of convoluted conditionals.

We can use integers to create rationals. How you ask? Well with pairs of course! Just like how we used a pair of naturals to define integers, we can use a pair of integers to define rationals. You define a rational number q as a pair (k, a) , where k is an integer and a a natural number such that $q = k/(1 + a)$.

3.2 Recursion

We already know how to do some control flow using conditionals, but how about looping? We can achieve this using *recursion*.

3.2.1 Understanding recursion

To understand recursion, we must first understand recursion. Recursion is when we make use of self-reference, usually through defining something in terms of itself. A common example is the following factorial function (in C):

```
int fac(unsigned int n)
{
    if (i == 0)
        return 1;
    else
        return n * fac(n - 1);
}
```

or if you prefer ternary operators:

```
int fac(unsigned int i)
{
    return n == 0 ? 1 : n * fac(n - 1);
}
```

Our definition of `fac()` contains calls to `fac()`, i.e. we have defined it in terms of itself. This is a very effective (although sometimes tricky, if you're not used to it) way of looping. Recursion, in fact, completely eliminates the need for things such as *for* or *while* loops commonly found in programming languages.

We've covered boolean logic and arithmetic in section 3. Thus, we can also define our C-function in pseudo lambda calculus⁸.

$$FAC := \lambda n.ISZERO\ n\ N1\ (MULT\ n\ (FAC\ (PRED\ n)))$$

or if you want to eliminate some braces using function composition (section 2.3.6)

$$FAC := \lambda n.ISZERO\ n\ N1\ (MULT\ n\ (B\ FAC\ PRED\ n))$$

where n is a natural number as defined in section 3.1.2.

Because of something called *laziness* (section 4.1.4), we don't have to be afraid that computation won't finish. In short, laziness means that the computer won't try to evaluate something until it has to. If it were to try to evaluate this function, it would become an infinitely long expression

$$\begin{aligned} FAC &:= \lambda n.ISZERO\ n\ N1\ (MULT\ n\ (FAC\ (PRED\ n))) \\ &= \lambda n.ISZERO\ n\ N1\ (MULT \\ &\quad n \\ &\quad ((\lambda m.ISZERO\ m\ N1\ (MULT\ m\ (FAC\ (PRED\ m)))) \\ &\quad (PRED\ n))) \\ &= \lambda n.ISZERO\ n\ N1\ (MULT \\ &\quad n \\ &\quad ((\lambda m.ISZERO\ m\ N1\ (MULT \\ &\quad m \\ &\quad ((\lambda l.ISZERO\ l\ N1\ (MULT\ l\ (FAC\ (PRED\ l)))) \\ &\quad (PRED\ m)))) \\ &\quad (PRED\ n))) \\ &= \dots \end{aligned}$$

but if we input a value, computation does terminate. Say we want to know the factorial of 0

$$\begin{aligned} FAC\ N0 &= (\lambda n.ISZERO\ n\ N1\ (MULT\ n\ (FAC\ PRED\ n)))\ N0 \\ &= ISZERO\ N0\ N1\ (MULT\ N0\ (FAC\ PRED\ N0)) \\ &= N1 \end{aligned}$$

or the factorial of 1

$$\begin{aligned} FAC\ N1 &= (\lambda n.ISZERO\ n\ N1\ (MULT\ n\ (FAC\ PRED\ n)))\ N1 \\ &= ISZERO\ N1\ N1\ (MULT\ N1\ (FAC\ PRED\ N1)) \\ &= MULT\ N1\ (FAC\ PRED\ N1) \\ &= MULT\ N1\ (FAC\ N0) \\ &= MULT\ N1\ N1 \\ &= N1 \end{aligned}$$

⁸Lambda calculus has no identifiers, so we can't just define an identifier in terms of itself to achieve recursion

or the factorial of 2

$$\begin{aligned}
FAC\ N2 &= (\lambda n.ISZERO\ n\ N1\ (MULT\ n\ (FAC\ PRED\ n)))\ N2 \\
&= ISZERO\ N2\ N1\ (MULT\ N2\ (FAC\ PRED\ N2)) \\
&= MULT\ N2\ (FAC\ PRED\ N2) \\
&= MULT\ N2\ (FAC\ N1) \\
&= MULT\ N2\ N1 \\
&= N2
\end{aligned}$$

or the factorial of 3

$$\begin{aligned}
FAC\ N3 &= (\lambda n.ISZERO\ n\ N1\ (MULT\ n\ (FAC\ PRED\ n)))\ N3 \\
&= ISZERO\ N3\ N1\ (MULT\ N3\ (FAC\ PRED\ N3)) \\
&= MULT\ N3\ (FAC\ PRED\ N3) \\
&= MULT\ N3\ (FAC\ N2) \\
&= MULT\ N3\ N2 \\
&= N6
\end{aligned}$$

We don't need to evaluate the recursive branch if our input is 0, so we don't. In most languages, like C, every expression gets evaluated immediately (not including function calls⁹).

The way we are achieving recursion in this example is by defining an identifier in terms of itself, which doesn't work in pure lambda calculus, since it doesn't have identifiers. What we need to do is to somehow create a combinator that can do recursion without the use of self-referencing identifiers. We will slowly work towards this combinator.

3.2.2 General recursion

Let's start with the simplest recursive combinator you can create in lambda calculus

$$LOOP := (\lambda x.xx)\ \lambda x.xx$$

which will result in an infinite loop if you try to reduce it:

$$\begin{aligned}
&(\lambda x.xx)\ \lambda x.xx \\
&= (\lambda x.xx)\ \lambda x.xx \\
&= (\lambda x.xx)\ \lambda x.xx \\
&= \dots
\end{aligned}$$

We have seen this before, this is the *omega combinator* (section 2.3.2).

This is not very useful, it just hangs without doing anything. We could define a slightly more useful function as

$$REC\ f := f\ (REC\ f)$$

which would be

$$REC := \lambda f.f\ (REC\ f)$$

in pseudo lambda calculus. It can apply any function f recursively. This is called *general recursion* in computer science—it is the simplest way to perform an operation recursively.

We can use this new function to redefine our *LOOP* combinator.

$$LOOP := REC\ \lambda a.a$$

⁹Which is why our factorial function in C works.

But we still haven't achieved general recursion in pure lambda calculus, because we are still defining identifiers—which pure lambda calculus doesn't have—in terms of themselves instead of defining a combinator which can do recursion by itself.

We can also use this new *REC* function to redefine our factorial function

$$\begin{aligned}
FAC &:= REC (\lambda f n. ISZERO\ n\ N1\ (MULT\ n\ (f\ (PRED\ n)))) \\
&= (\lambda f n. ISZERO\ n\ N1\ (MULT\ n\ (f\ (PRED\ n)))) \\
&\quad (REC (\lambda f n. ISZERO\ n\ N1\ (MULT\ n\ (f\ (PRED\ n)))))) \\
&= (\lambda f n. ISZERO\ n\ N1\ (MULT\ n\ (f\ (PRED\ n)))) FAC \\
&= \lambda n. ISZERO\ n\ N1\ (MULT\ n\ (FAC\ (PRED\ n)))
\end{aligned}$$

which reduces to our previous definition as you can see.

3.2.3 The Y-combinator

We have now defined *FAC* not in terms of itself using a specialised recursion combinator. This combinator *is* still defined in terms of itself though. If we succeed to define general recursion in lambda calculus without the use of self-referring identifiers, we can finally write our recursive functions in pure lambda calculus.

Luckily for us, this combinator exists, it's called the *Y-combinator*. Instead of self-reference, it uses self-application, much like how our earlier *LOOP* combinator is really a function applied to itself resulting in that same function being applied to itself. The Y-combinator looks much like our *LOOP* combinator, except it has an extra function as argument which is applied to the application of the arguments of the inner functions to themselves.

$$Y := \lambda f. (\lambda x. f(xx)) \lambda x. f(xx)$$

If we try to evaluate this, we again get an infinitely large expression

$$\begin{aligned}
Y &:= \lambda f. (\lambda x. f(xx)) \lambda x. f(xx) \\
&= \lambda f. f((\lambda x. f(xx)) \lambda x. f(xx)) \\
&= \lambda f. f(f((\lambda x. f(xx)) \lambda x. f(xx))) \\
&= \lambda f. f(f(f((\lambda x. f(xx)) \lambda x. f(xx)))) \\
&= \lambda f. f(f(f(f(f(f(f(f(...))))))))
\end{aligned}$$

We can see it does in fact function as expected.

We have successfully encoded recursion without recursion, and can now implement our *FAC* function in lambda calculus. The lambda expression is really long, so I've tried to break it up across multiple lines in a way that somewhat makes sense.

$$\begin{aligned}
FAC &:= Y \\
&\quad (\lambda f n. ISZERO\ n\ N1 \\
&\quad\quad (MULT\ n \\
&\quad\quad\quad (f \\
&\quad\quad\quad\quad (PRED\ n)))) \\
&= Y \\
&\quad (\lambda f n. ISZERO\ n\ N1 \\
&\quad\quad (MULT\ n \\
&\quad\quad\quad (f
\end{aligned}$$

$$\begin{aligned}
& (FST \\
& \quad (n \ NEXT \\
& \quad \quad (PAIR \ N0 \ N0)))))) \\
= & Y \\
& (\lambda f n. ISZERO \ n \ N1 \\
& \quad (MULT \ n \\
& \quad \quad (f \\
& \quad \quad \quad (FST \\
& \quad \quad \quad \quad (n \\
& \quad \quad \quad \quad \quad (\lambda p. PAIR \ (SND \ p) \ (SUCC \ (SND \ p))) \\
& \quad \quad \quad \quad \quad (PAIR \ N0 \ N0)))))) \\
= & Y \\
& (\lambda f n. n(KF)T \ N1 \\
& \quad (MULT \ n \\
& \quad \quad (f \\
& \quad \quad \quad (n \\
& \quad \quad \quad \quad (\lambda p. V \ (p \ KI) \ (SUCC \ (p \ KI))) \\
& \quad \quad \quad \quad (V \ N0 \ N0) \\
& \quad \quad \quad \quad K)))) \\
= & (\lambda f. (\lambda x. f(xx))(\lambda x. f(xx))) \\
& (\lambda f n. n((\lambda ab. a)(\lambda ab. b))(\lambda ab. a)(\lambda ga. ga) \\
& \quad ((\lambda klg. k(lg))n \\
& \quad \quad (f \\
& \quad \quad \quad (n \\
& \quad \quad \quad \quad (\lambda p. (\lambda abg. gab)(p(\lambda ab. b))((\lambda kga. g(kga))(p(\lambda ab. b)))) \\
& \quad \quad \quad \quad ((\lambda abg. gab)(\lambda ga. a)(\lambda ga. a)) \\
& \quad \quad \quad \quad (\lambda ab. a)))))) \\
= & (\lambda f. (\lambda x. f(xx))(\lambda x. f(xx))) \\
& (\lambda f n. n(\lambda abc. c)(\lambda ab. a)(\lambda ga. ga) \\
& \quad ((\lambda kg. n(kg)) \\
& \quad \quad (f \\
& \quad \quad \quad (n \\
& \quad \quad \quad \quad (\lambda p. (\lambda abg. gab)(p(\lambda ab. b))(\lambda ga. g(p(\lambda bc. c)ga))) \\
& \quad \quad \quad \quad (\lambda g. g(\lambda ab. b)(\lambda ab. b)) \\
& \quad \quad \quad \quad (\lambda ab. a)))))) \\
= & (\lambda f. (\lambda x. f(xx))(\lambda x. f(xx))) \\
& (\lambda f n. n(\lambda abc. c)(\lambda ab. a)(\lambda ga. ga) \\
& \quad ((\lambda kg. n(kg)) \\
& \quad \quad (f \\
& \quad \quad \quad (n \\
& \quad \quad \quad \quad (\lambda pg. g(p(\lambda ab. b))(\lambda ha. h(p(\lambda bc. c)ha))) \\
& \quad \quad \quad \quad (\lambda g. g(\lambda ab. b)(\lambda ab. b))
\end{aligned}$$

$$\begin{aligned}
& (\lambda ab.a)))))) \\
= & (\lambda f.(\lambda x.f(xx))(\lambda x.f(xx))) \\
& (\lambda fn.n(\lambda abc.c)(\lambda ab.a)(\lambda ga.ga) \\
& ((\lambda kg.n(kg)) \\
& (f \\
& (n \\
& (\lambda g.g((\lambda h.h(\lambda ab.b)(\lambda ab.b))(\lambda ab.b))(\lambda ha.h((\lambda i.i(\lambda bc.c)(\lambda bc.c))(\lambda bc.c)ha))) \\
& (\lambda ab.a)))))) \\
= & (\lambda f.(\lambda x.f(xx))(\lambda x.f(xx))) \\
& (\lambda fn.n(\lambda abc.c)(\lambda ab.a)(\lambda ga.ga) \\
& ((\lambda kg.n(kg)) \\
& (f \\
& (n \\
& (\lambda g.g((\lambda ab.b)(\lambda ab.b)(\lambda ab.b))(\lambda ha.h((\lambda bc.c)(\lambda bc.c)(\lambda bc.c)ha))) \\
& (\lambda ab.a)))))) \\
= & (\lambda f.(\lambda x.f(xx))(\lambda x.f(xx))) \\
& (\lambda fn.n(\lambda abc.c)(\lambda ab.a)(\lambda ga.ga) \\
& ((\lambda kg.n(kg)) \\
& (f \\
& (n \\
& (\lambda g.g(\lambda ab.b)(\lambda ha.h((\lambda bc.c)ha))) \\
& (\lambda ab.a)))))) \\
= & (\lambda f.(\lambda x.f(xx))(\lambda x.f(xx))) \\
& (\lambda fn.n(\lambda abc.c)(\lambda ab.a)(\lambda ga.ga) \\
& ((\lambda kg.n(kg)) \\
& (f(n(\lambda g.g(\lambda ab.b)(\lambda ha.ha))(\lambda ab.a)))))) \\
= & (\lambda f.(\lambda x.f(xx))(\lambda x.f(xx))) \\
& (\lambda fn.n(\lambda abc.c)(\lambda ab.a)(\lambda ga.ga) \\
& (\lambda g.n(f(n(\lambda h.h(\lambda ab.b)(\lambda ab.ab))(\lambda ab.a))g)))
\end{aligned}$$

I've suffered through many painful hours reducing this lambda function, so out of pure evil, I'll deliberately not give an example of this reduced lambda function in action, but leave it as an exercise to the reader so as to make him suffer through many painful hours like I have¹⁰.

¹⁰Because fuck you.

4 Functional programming (lambda calculus applied)

4.1 Interesting and useful properties

4.1.1 Statelessness

4.1.2 First class and higher order functions

4.1.3 Function currying

4.1.4 Laziness

4.2 Typed lambda calculus

4.2.1 Simply typed lambda calculus

4.2.2 Strongly typed lambda calculus (System F)

4.3 Catagory theory

Catagory theory is very important to functional programming. It really should have a section of its own. The ideas of catagory theory fit really well with lambda calculus. Some ideas, like function composition, we've talked about extensively already. The concepts of catagory theory may be a little too abstract and mathematical for some readers, but it is worth to have some understanding of catagory theory if you want to learn functional programming (although it is not strictly necessary of course). In short, catagory theory is a general theory of functions that fit perfectly with the ideas behind lambda functions. It is usually represented visually in two dimensions, which may remind of Frege's *Begriffsschrift*, which we mentioned in section 2.1.

4.3.1 A quick look at set theory

Mathematicians should be familiar with *set theory*. Set theory is a way of describing things as members of sets of things that share one or more qualities. These things can be members of multiple sets simultaneously and can represent anything, even other sets. Common examples of sets often found in mathematics are the sets of natural numbers, integers, rational numbers, real numbers, imaginary numbers, complex numbers, etc. If something is in the set of natural numbers, it has all the qualities of a natural number, but hasn't have to be limited to those qualities. Sets have some similarities to types, but quite unlike types, something can be part of multiple sets, but always has to have just one type.

It is very possible for paradoxes to come up in set theory. The most common example would be Russell's paradox. It reads as follows: if you define R to be the set of all things not containing themselves, then you can't say wether R is or isn't in R . If R *isn't* in R , it should be in R , but if R *is* in R , then it shouldn't be in R . There are only two possibilities, but both lead to contradiction.

$$\text{Let } R = \{x \mid x \notin x\}, \text{ then } R \in R \iff R \notin R$$

The above equation is Russell's paradox in basic set notation. You should read it as: "*Let R be the set of all things x not containing x , then R is in R if, and only if R is not in R .*" Don't worry about the notation too much, it isn't important right now.

This is also where our good friend Gödel comes in. We have talked about him in section 2.1. There are many parallels between Russell's paradox and so called Gödel sentences. A basic example of such a sentence would be the following:

This sentence is false.

If you try to figure out the truth value of the sentence, you end up with the same contradiction as with Russell’s paradox. This problem arises from self-reference. Self reference is interesting enough in itself, there have been entire books written on it. A very popular example being *Gödel, Escher, Bach*.

Because it is so easy for paradoxes to come up in set theory, this simplest form of set theory is often called *naïve set theory*. Russell, who we have talked about extensively in section 2.1, designed the *theory of types* (not to be confused with *type theory*) to avoid these paradoxes from happening in formal systems such as set theory (as mentioned in section 2.1). I will not go into it, but it is quite interesting to look at. There exist many variations of set theory besides naïve set theory and set theory using Russell’s theory of types.

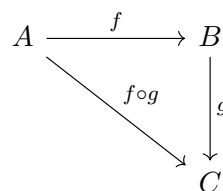
4.3.2 The jist (of catagory theory)

Catagory theory is similar to set theory. It has the same concept of sets, except they’re called *catagories*. The emphasis in catagory theory is not the catagories things belong to, but the relationships between things. We’ll call these relationships *maps*, but other names are also commonly used to indicate these relationships. Maps can exist between simple objects, catagories or even other maps. Maps are often visually represented as arrows between things.

A map between two catagories could be understood as a function where the first catagory is the catagory of possible inputs, and the second catagory being the set of possible outputs. A map *maps* between these two catagories in some formally defined way. This idea of functions is very similar to that in lambda calculus, where there are no side effects and the output is only dependent on the input. This basically means that every function could be written as a (although be it possibly an infinitely long) lookup table. This also reflects the idea of functionality and data being equivalent, which is strongly present in both lambda calculus and catagory theory. There can exist multiple maps between catagories; every arithmetic operation would be a map from numbers to numbers. We can also have maps between maps, which are called *functors*. I will discuss them in section 4.3.4, but first I should talk about composition.

4.3.3 Composition

Probably the most fundamental idea in catagory theory is that of composition. We know what function composition is from lambda calculus. In catagory theory is described as follows: You have two maps, one mapping A to B and another mapping B to C then the composition of those maps maps A to C just as if you were to apply those maps consecutively. The following image, which is one of the first things you’ll see when learning catagory theory, is the following image, which represents function composition and which is almost the “*Hello, world!*”¹¹ of catagory theory:



4.3.4 Functors

As I hinted at earlier, we can have maps between maps, since maps are also just “things”. These maps are called *functors*.

¹¹A “*Hello, world!*” program is the program you usually first write when learning a new language. In most languages it would just output the string “Hello, world!”.

4.3.5 Monoids

A major problem with pure functional programming, which I also talk about in sections 4.1.1 and 4.5.2, is that it is completely useless. There has to be some way to do side effects to write a useful program. You don't want to accidentally fire the rockets when you are just doing arithmetic, but you also want to still be able to fire the rockets when you have to. The solution is monads (see the next section 4.3.6), but we need to first understand monoids.

4.3.6 Monads

4.4 Haskell

4.4.1 Basic syntax

4.4.2 Monads

4.5 Comparison to other paradigms

4.5.1 Declarative vs imperative

4.5.2 Usefulness vs conceptual purity

4.5.3 Meta programming

5 A practical example

Section 4.5 compares functional programming with other paradigms. I thought it would be interesting to do a practical example. In section 4.5.3, I talked about the Lisp programming language family, in particular Scheme. It isn't all too hard to implement a Lisp-like language, but implementing a programming language is very conceptual. Programming languages have many abstract concepts that are more easily expressed in a declarative/functional language like Haskell, rather than a procedural/imperative language like C. This section is really a practical continuation of section 4.5, so I recommend you read it before you read this. This section tries to paint a picture of what is said in section 4.5.

5.1 Scheme interpreter in C

5.2 Scheme interpreter in Haskell

Afterword

Unreferenced resources

References

- [1] Alonzo Church. “A set of postulates for the foundation of logic”. In: *Annals of mathematics* (1932), pp. 346–366.
- [2] Alonzo Church. “An unsolvable problem of elementary number theory”. In: *American journal of mathematics* 58.2 (1936), pp. 345–363.
- [3] Haskell Brooks Curry. “Grundlagen der kombinatorischen Logik”. In: *American journal of mathematics* 52.4 (1930), pp. 789–834.
- [4] Gottlob Frege. *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Nebert, 1879.
- [5] Kurt Gödel. “Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I”. In: *Monatshefte für mathematik und physik* 38.1 (1931), pp. 173–198.
- [6] Giuseppe Peano. *Arithmetices principia: Nova methodo exposita*. Fratres Bocca, 1889.
- [7] Bertrand Russell and Alfred North Whitehead. *Principia mathematica to* 56*. Vol. 2. Cambridge University Press Cambridge, UK, 1997.
- [8] Moses Schönfinkel. “Über die Bausteine der mathematischen Logik”. In: *Mathematische annalen* 92.3-4 (1924), pp. 305–316.
- [9] Raymond Merrill Smullyan. *To Mock a Mockingbird: and other logic puzzles including an amazing adventure in combinatory logic*. Oxford University Press, USA, 2000.
- [10] Alan Mathison Turing. “Computability and λ -definability”. In: *The Journal of Symbolic Logic* 2.4 (1937), pp. 153–163.
- [11] Alan Mathison Turing. “On computable numbers, with an application to the Entscheidungsproblem”. In: *J. of Math* 58.345-363 (1936), p. 5.
- [12] Alan Mathison Turing. *On computable numbers, with an application to the Entscheidungsproblem; a correction*. Royal Society, 1937.
- [13] Wikipedia. *Combinatory logic* — *Wikipedia, The Free Encyclopedia*. 2020. URL: https://web.archive.org/web/20200920164914if_/https://en.wikipedia.org/wiki/Combinatory_logic (visited on 20/09/2020).
- [14] Wikipedia. *Gödel’s incompleteness theorems* — *Wikipedia, The Free Encyclopedia*. 2020. URL: https://web.archive.org/web/20200927072027/https://en.wikipedia.org/wiki/G%C3%B6del%27s_incompleteness_theorems (visited on 27/09/2020).