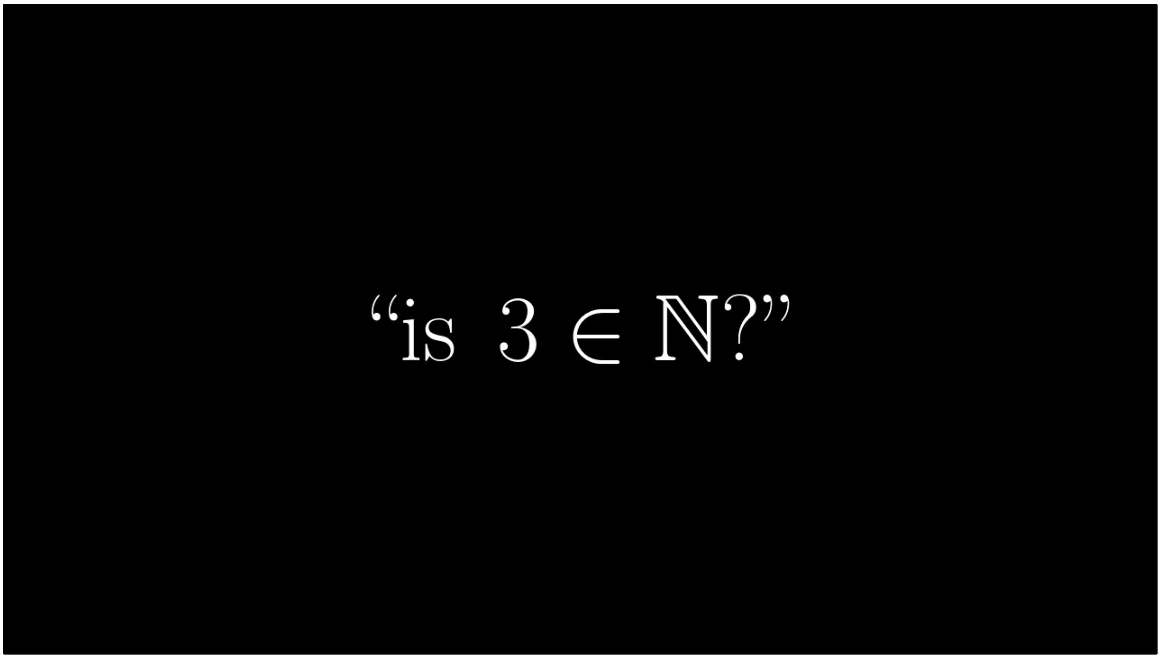
Structural Set Theory in Foundations

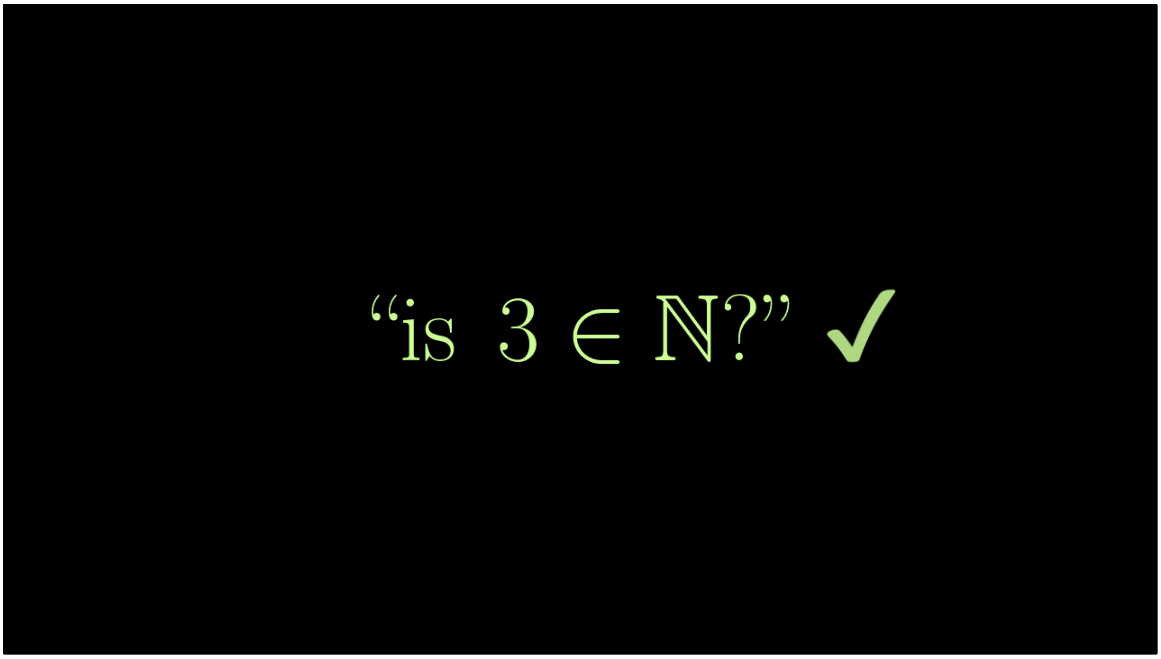
The Elementary Theory of the Category of Sets (ETCS)

Kit L (u2111082@warwick.ac.uk)

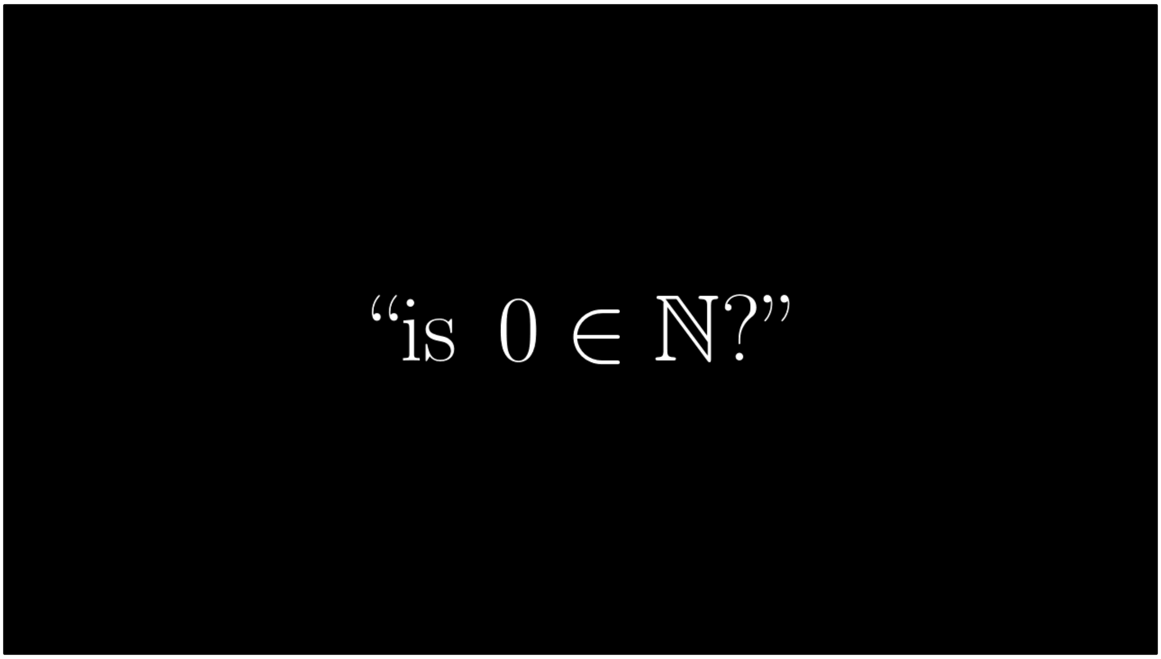
Introduction



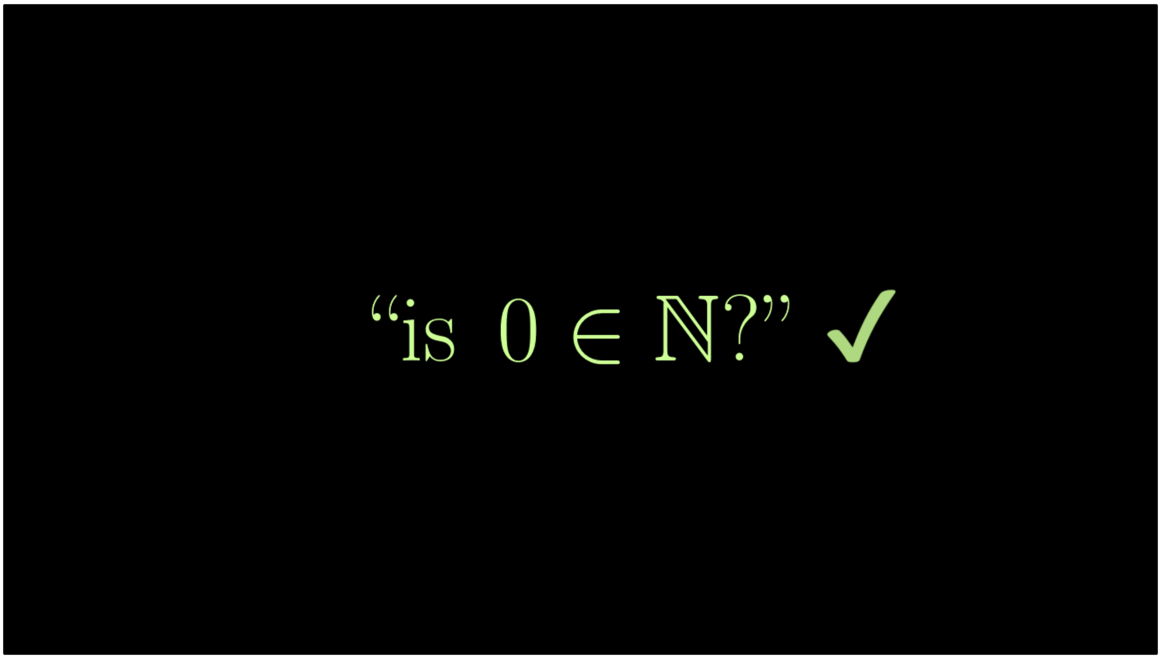
Suppose for a moment that you were asked, is 3 in N? As 3 is a natural number, the answer is simply, “yes”.



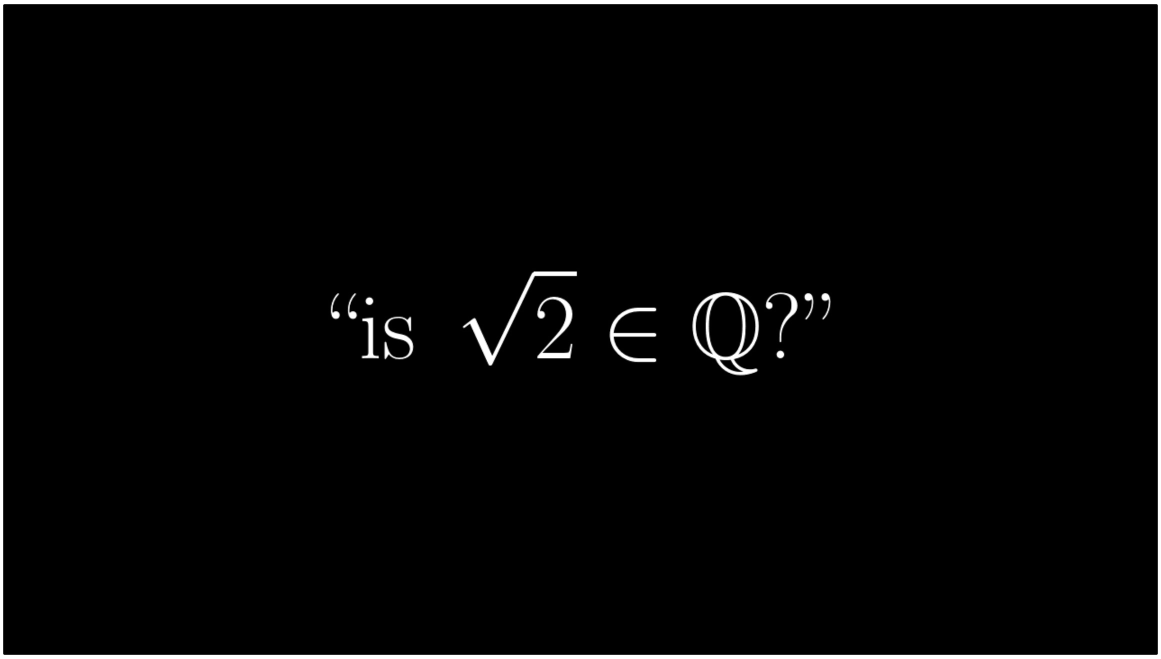
A little tricker is the question, is 0 in N?



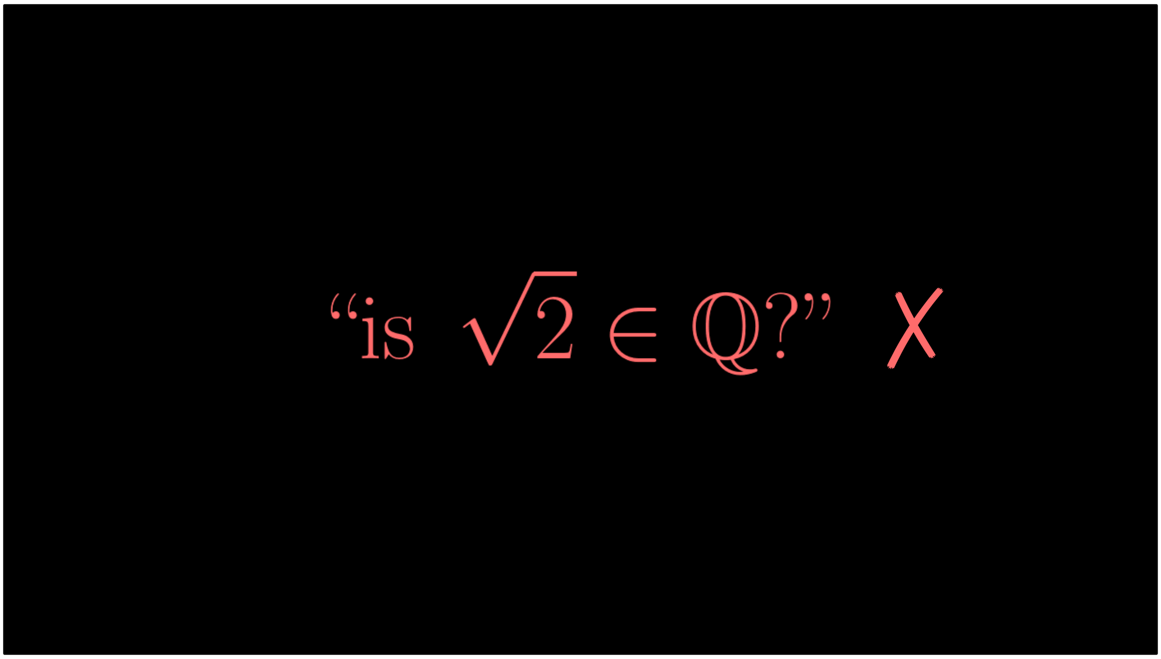
but as long as we are clear about the meaning of this symbol, this statement at least has a clear and unambiguous answer.



For our purposes, we will take the answer to be "yes".



On the other hand, something like, “is sqrt(2) in the set of rationals?” would quickly receive an answer of “no”.



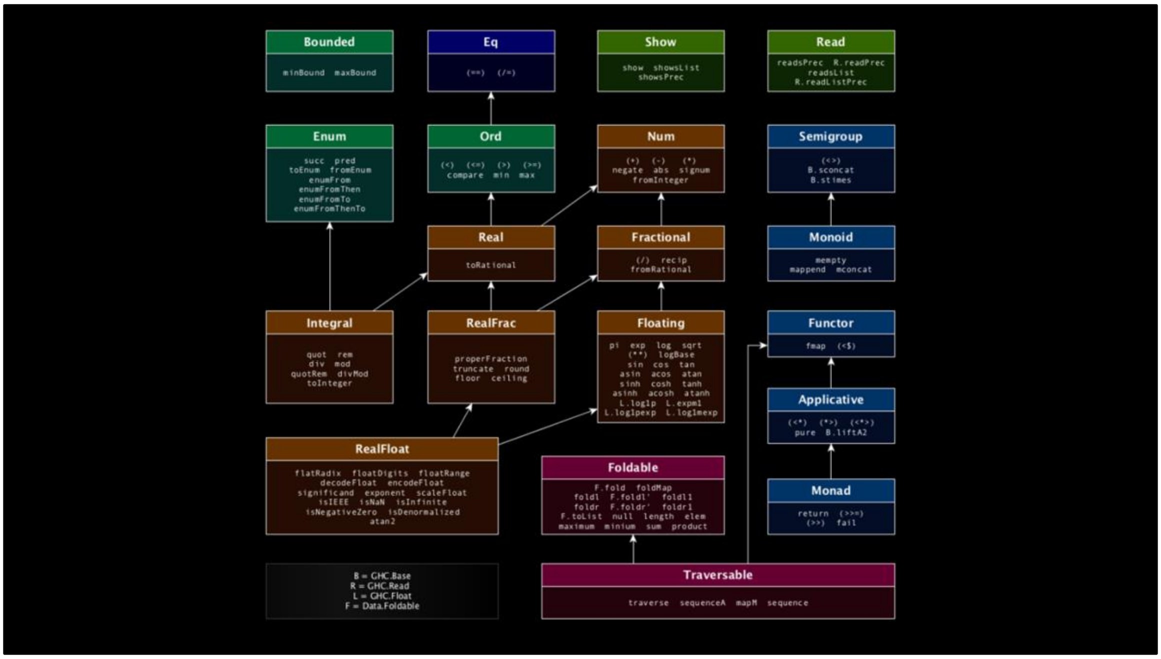
Now, suppose you were then asked,

“is $\pi \in \log$?”

“is pi in log?” We’d might take a second before again saying “no”, but for a different reason than before. After all, log is a function, so pi being a member of log, whatever that means, would be ridiculous.

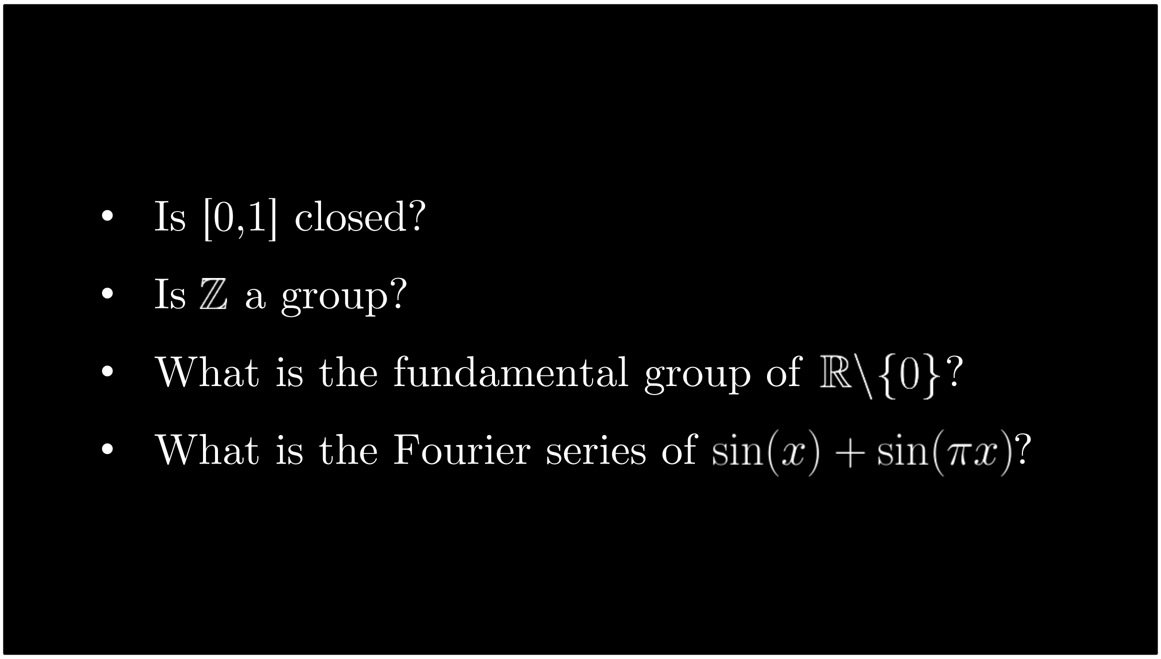
Rather than saying “no”, a better answer might be to say that the question is meaningless.

This illustrates the intuitive notion of type, which may be particularly familiar to programmers.



Many programming languages require you to declare the type of a variable before it can be instantiated, the idea being that strictly enforcing the type of every variable stops you from attempting to perform nonsensical operations, like adding an int to a Bool, or trying to divide by a string.

Sometimes, we see mathematical questions that don't appear "grammatically correct", so to speak. For instance,

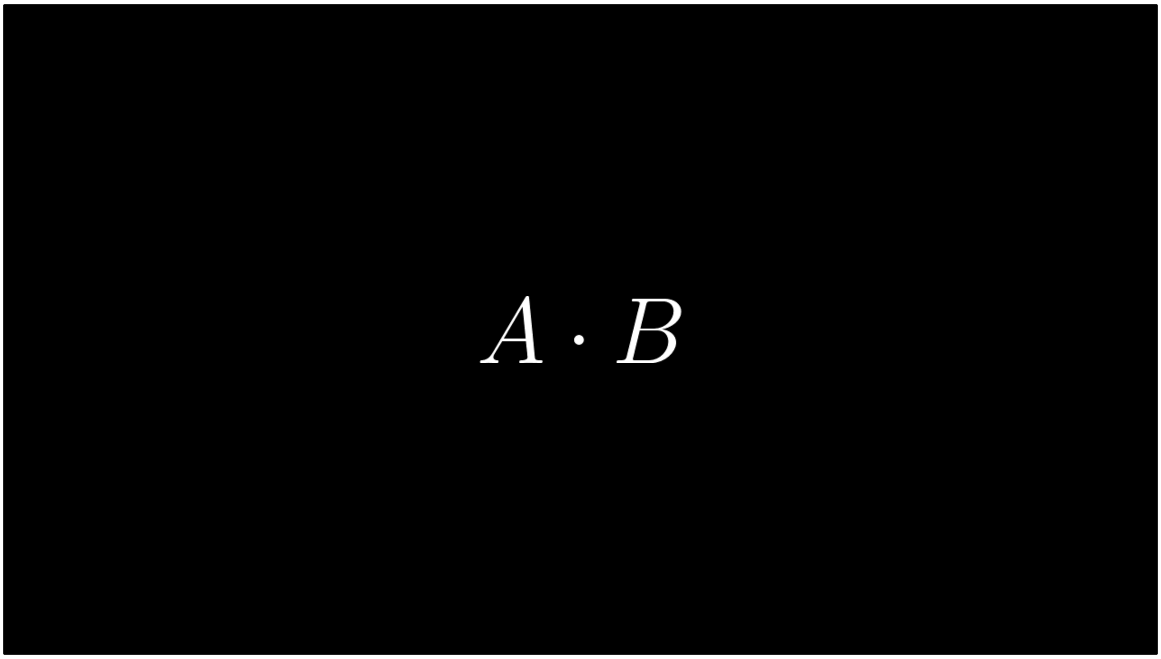


these expressions just don't seem *correct*. Or, for slightly less subtle examples,

- Is [0,1] closed?
- Is $\mathbb{Z}$ a group?
- What is the fundamental group of $\mathbb{R}\backslash\{0\}$?
- What is the Fourier series of $\sin(x) + \sin(\pi x)$?
- Is a rectangle prime?
- Is 3 surjective?
- Does a prime converge?

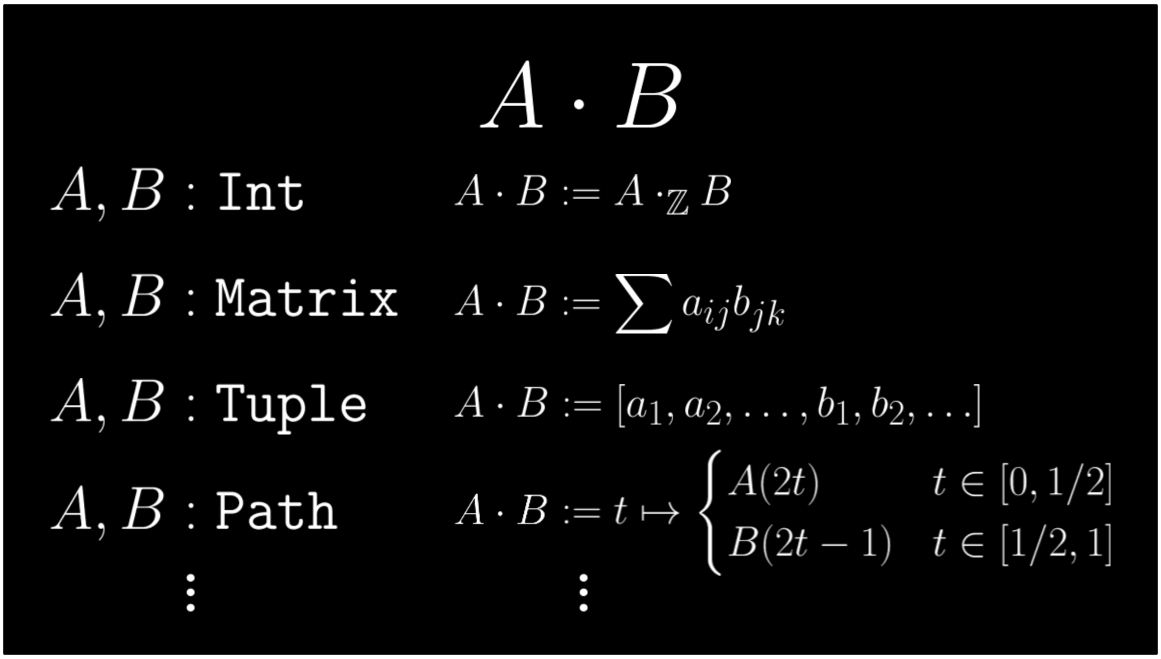
we have these.

The notion of type formalises this idea that these questions aren't grammatically correct – these sentences all have type errors. We all intuitively use types throughout mathematics beyond this as well. For instance,



What does this expression mean?

Well, it depends on what *type* of object A and B are.



If they're integers, then this is multiplication. But if they're say, paths, then it's path concatenation.

The point is, we all unconsciously use the notion of types throughout all of mathematics, and this operator polymorphism is just one simple example of where we do this.

However, in the standard foundational framework of ZFC, Zermelo-Fraenkel set theory with Choice, everything is a set. Moreover, membership is a global relation on sets, so you can ask whether any two sets are members of each other or not.

In ZFC:

Everything is a `set`

However, in the standard foundational framework of ZFC, Zermelo-Fraenkel set theory with Choice, everything is a set. Moreover, membership is a global relation on sets, so you can ask whether any two sets are members of each other or not.

These things in combination imply that "is pi in log" has a definite true or false answer.

Because of this, the way ZFC uses the word "set" is very different from what mathematicians usually mean when they say "set". In ZFC, pi is a set, as is log – but ask any working mathematician to list some of its elements, and you'll likely have difficulties in receiving an answer.

In ZFC:

$\in$ is a global relation on sets, so

$$A \in B$$

is a well-formed proposition
for all sets A, B.

However, in the standard foundational framework of ZFC, Zermelo-Fraenkel set theory with Choice, everything is a set. Moreover, membership is a global relation on sets, so you can ask whether any two sets are members of each other or not.

These things in combination imply that "is pi in log" has a definite true or false answer.

Because of this, the way ZFC uses the word "set" is very different from what mathematicians usually mean when they say "set". In ZFC, pi is a set, as is log – but ask any working mathematician to list some of its elements, and you'll likely have difficulties in receiving an answer.

In ZFC:

$\in$ is a global relation on sets, so

$$A \in B$$

is a well-formed proposition
for all sets A, B.

However, in the standard foundational framework of ZFC, Zermelo-Fraenkel set theory with Choice, everything is a set. Moreover, membership is a global relation on sets, so you can ask whether any two sets are members of each other or not.

These things in combination imply that “is pi in log” has a definite true or false answer.

Because of this, the way ZFC uses the word “set” is very different from what mathematicians usually mean when they say “set”. In ZFC, pi is a set, as is log – but ask any working mathematician to list some of its elements, and you’ll likely have difficulties in receiving an answer.

ZFC is *single-sorted*

The benefit of this is simplicity: everything is a set, so we can have a unified set of rules to deal with every different kind of object, and also, everyone already learns how to manipulate sets early on, so it's relatively easy to teach.

On the other hand, we lose this basic notion of type because everything is of type set. As demonstrated by "is pi in log", this isn't always sensible. For instance, take the axiom of regularity:

Axiom of Regularity (ZFC):

Every non-empty set X has an element x disjoint from itself:

$$\forall X\big(X \neq \emptyset \to \exists x(x \in X \wedge x \cap X = \emptyset)\big)$$

But, if we take X to be any ordinary set,

Axiom of Regularity (ZFC)**:**

Every non-empty set X has an element x disjoint from itself:

$$\forall X\big(X \neq \emptyset \to \exists x(x \in X \wedge x \cap X = \emptyset)\big)$$

Is $3 \cap \mathbb{R}$ empty?

What does an element of this even look like?

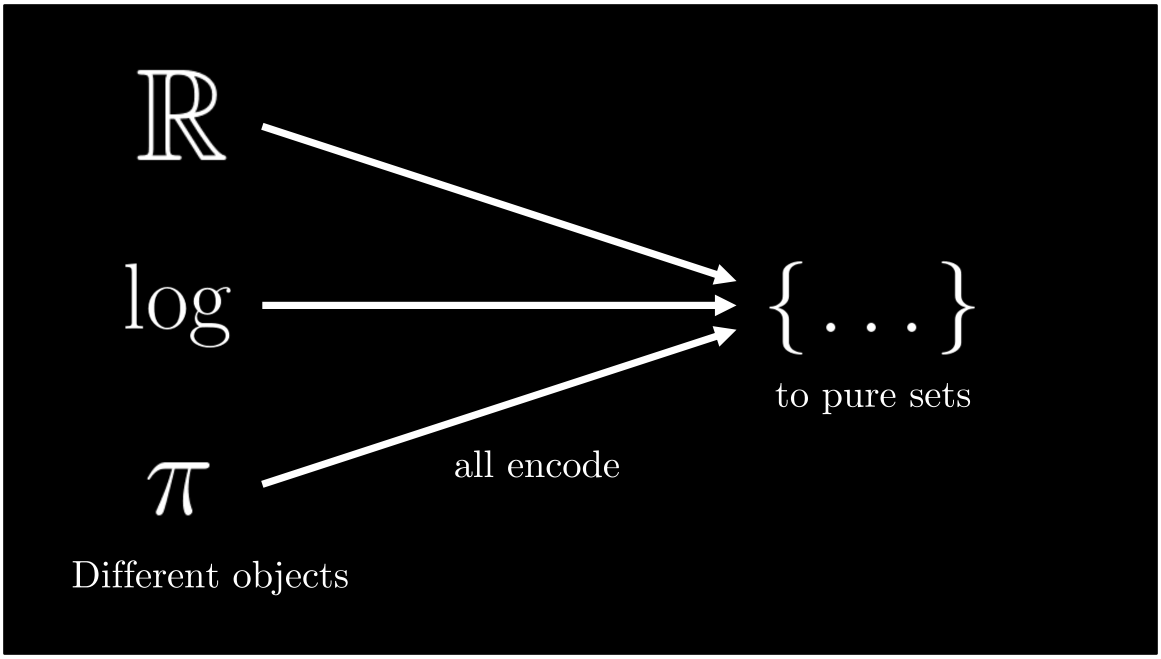
say, the set of real numbers R, then the resulting statement is hard to interpret. The axiom says that there exists a real number x such that x intersect R is empty.

The problem is again that we have an operation – namely intersection – being compatible with any two arbitrary sets, even it that doesn't really make sense.

ZFC also includes a set of standard encodings of mathematical objects

One response to these problems is to say that set theory offers not only a collection of axioms,

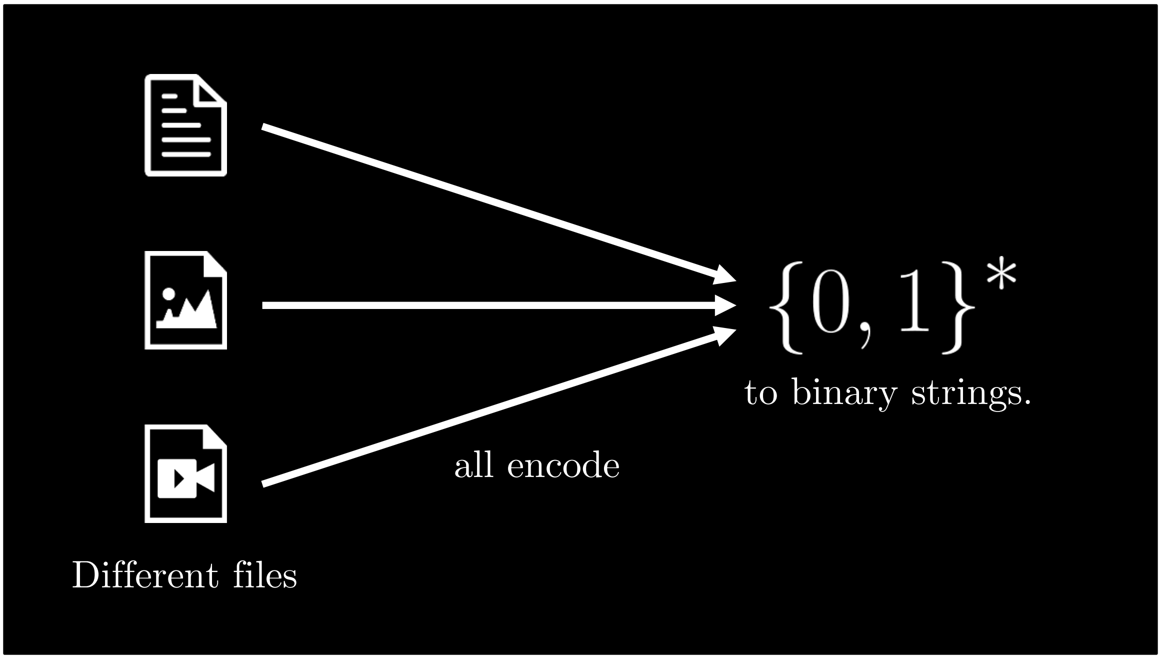
but also a collection of standard encodings of different mathematical objects.



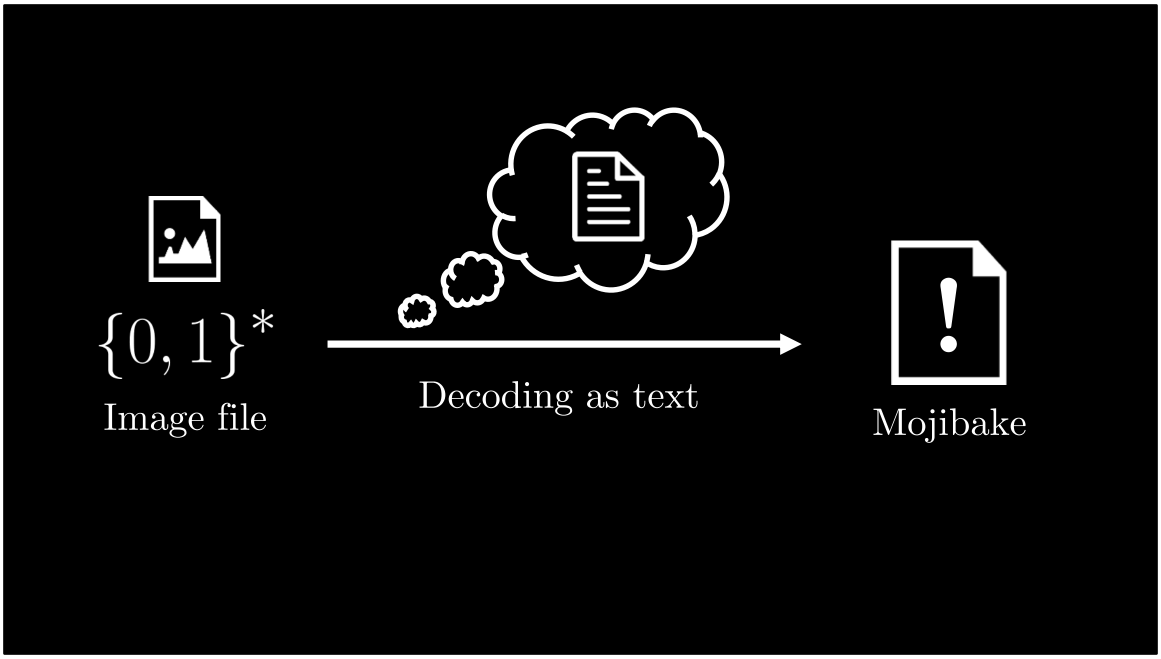
but also a collection of standard encodings of different mathematical objects.

We can again compare this situation with computers:

In a hard drive, every file – text, image, audio, video, etc – is encoded as sequence of bits; ones and zeros.

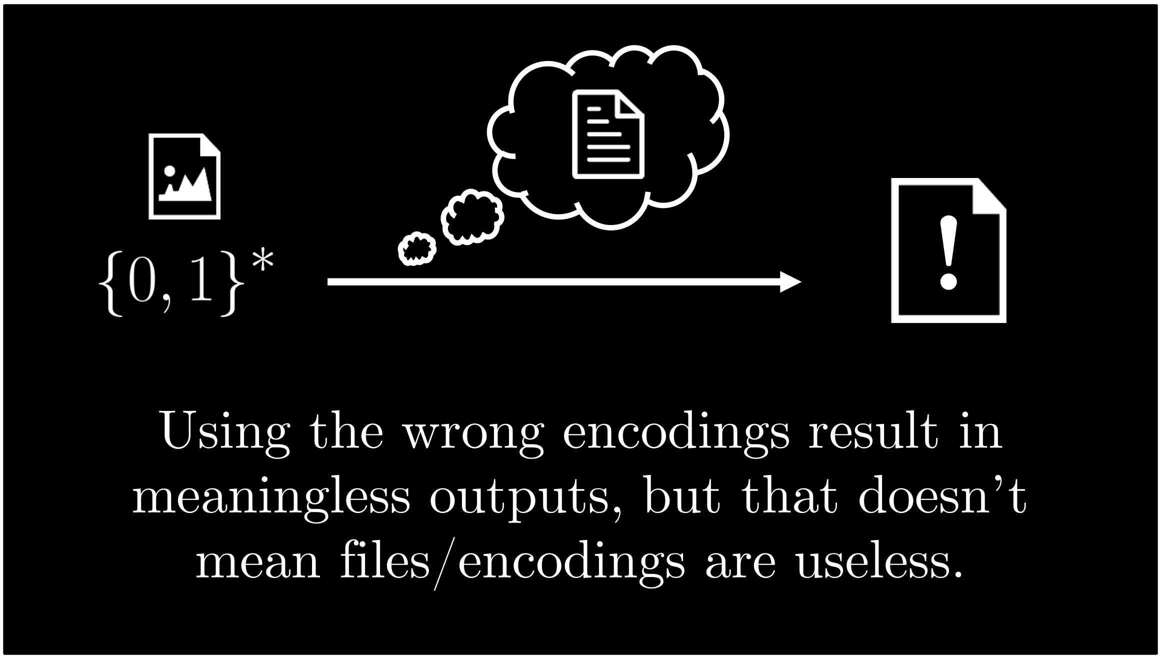


In a hard drive, every file – text, image, audio, video, etc – is encoded as sequence of bits; ones and zeros.



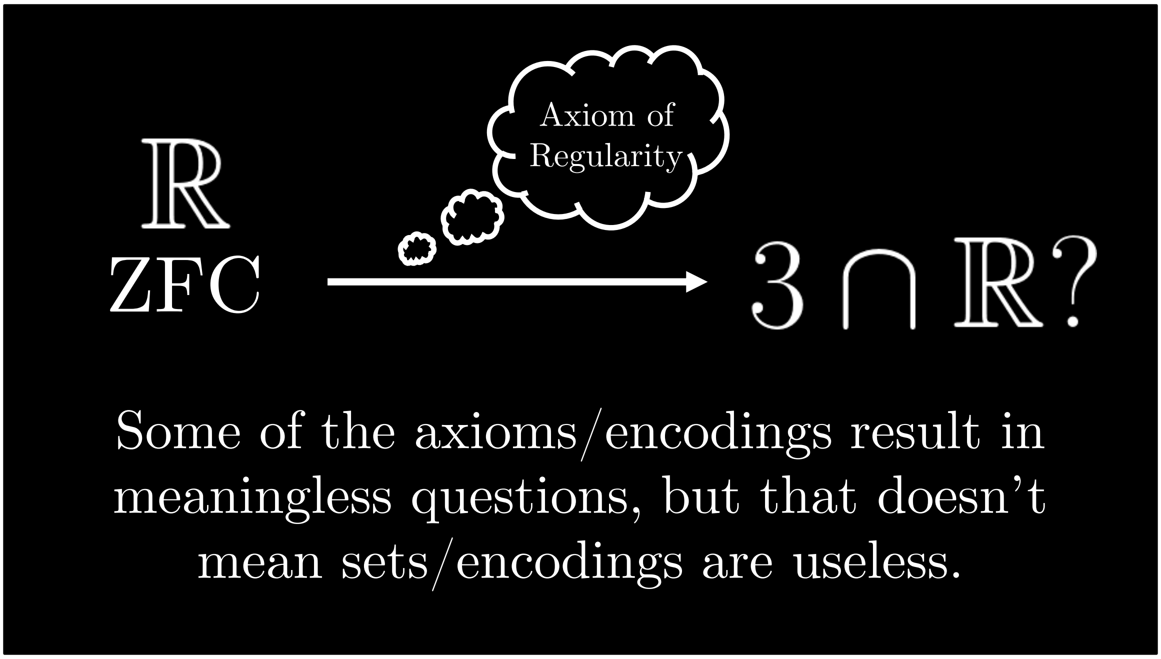
If we decode a file in the wrong way, we get an unrecognisable mess out the other end. But that doesn't mean that files, binary, or encoding is useless.

In the same way, ZFC allowing us to ask weird questions that don't have meaningful answers doesn't stop it from being useful as a foundation.



If we decode a file in the wrong way, we get an unrecognisable mess out the other end. But that doesn't mean that files, binary, or encoding is useless.

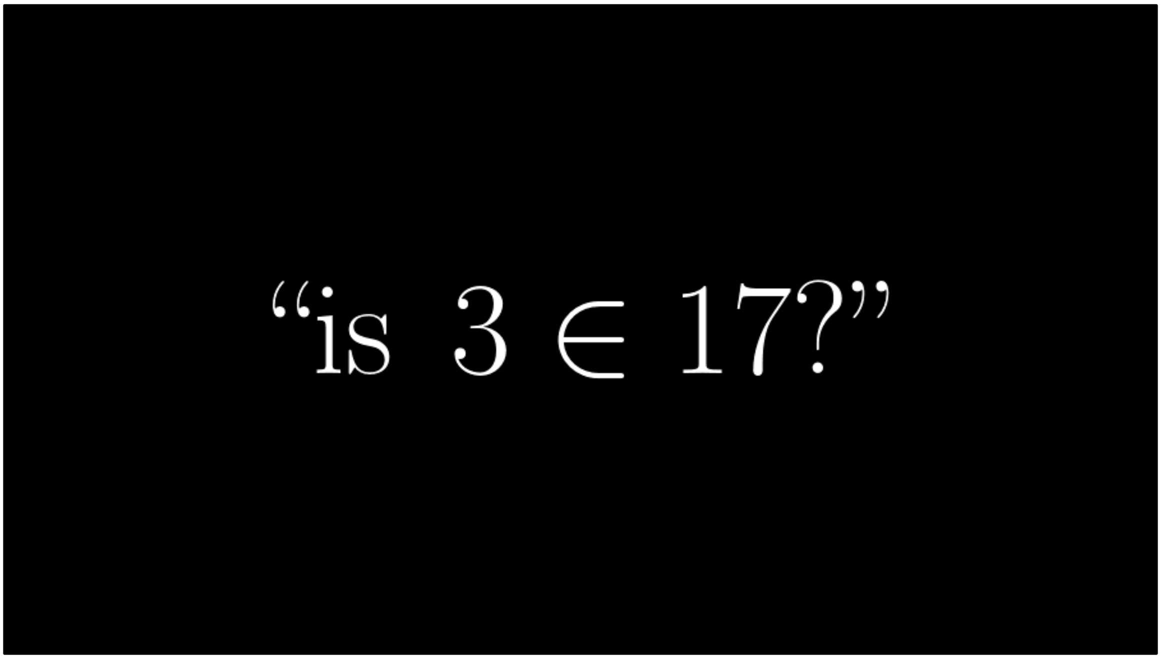
In the same way, ZFC allowing us to ask weird questions that don't have meaningful answers doesn't stop it from being useful as a foundation.



And this is very reasonable: ZFC is, of course, a perfectly functional foundational system. It's the most common choice of foundations for a reason.

However, can we find a way to resolve these problems that is more mathematically pleasing than just ignoring them?

One enlightening exercise, as posed by Benacerraf is to consider the question,



is 3 in 17?

Benacerraf describes two children, Johnny and Ernie – named in reference to John von Neumann and Ernst Zermelo – who have learnt about the natural numbers from axiomatic foundations, as opposed to the more commonly preferred method of starting from “counting”, which he calls the “vulgar way”.

However, there is some choice here.

$$N$$

The (set of) *natural numbers*

Johnny is taught that there is a set, N, that contains what ordinary people call the natural numbers.

$$0$$

The natural number *zero*

Furthermore, this set contains an element that ordinary people call the natural number zero.

$$s : N \to N$$

The *successor function*

And this set is equipped with a function called the successor, defined like this:

$$s(n) := n \cup \{n\}$$

The *successor function*

This is the standard construction of the naturals from von Neumann.

The normal properties of natural numbers can then be concretely proved as theorems for Johnny. For instance, the vulgar idea of counting is just the formal notion of cardinality.

The "vulgar" way	Axiomatic foundations
counting	cardinality
addition	simple recursion
less-than	well-ordering

The normal properties of natural numbers can then be concretely proved as theorems for Johnny. For instance, the vulgar idea of counting is just the formal notion of cardinality.

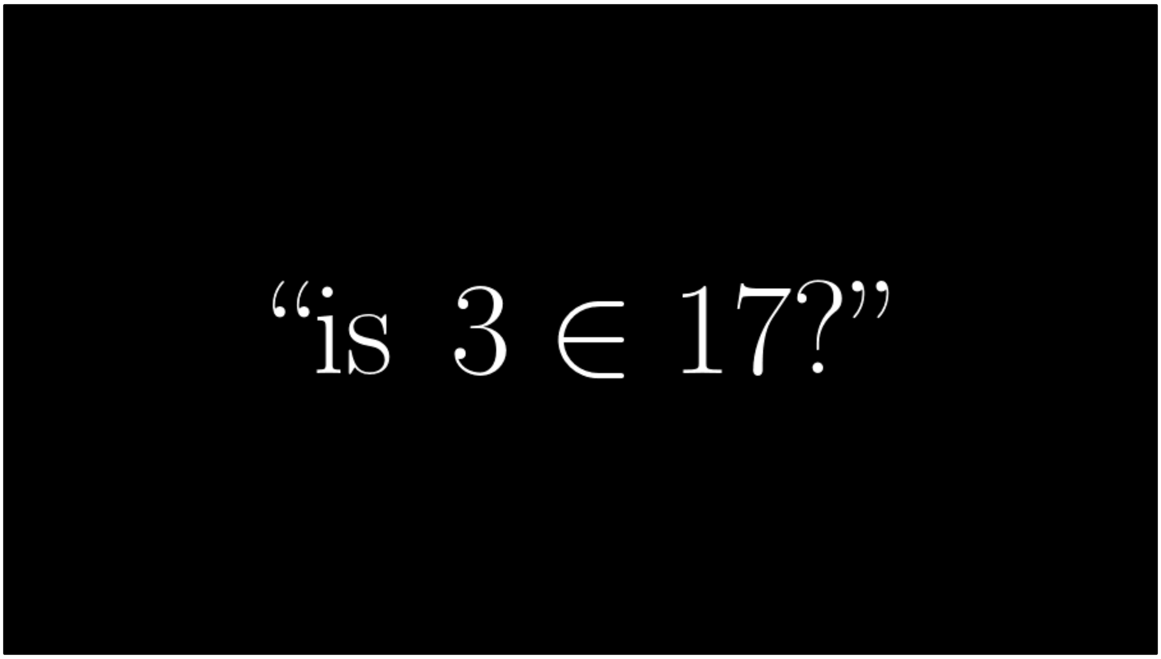
Having constructed all of these things in his first order theory, it is clear that Johnny can now communicate with the vulgar, with all common extramathematical uses of numbers being easily defined in terms of these set theoretic constructions.

$$N, \quad 0, \quad s$$

Note that all we have done is specify these three things and everything else is derivable from there, so this information is both necessary and sufficient to characterise the natural numbers for communicating with the vulgar.

Now, this story could also have been told about Ernie. Like Johnny, Ernie is provided with a set N, a distinguished element zero, and a successor function s. So, the two are equally knowledgeable about the natural numbers, and in conversations with the ordinary people, they are in complete agreement.

The problems first arise when they ask



is 3 in 17? Johnny argues that the statement is true, while Ernie disagrees. Attempts to resolve this by consulting with ordinary people are met with nothing but confusion. After all, to ordinary people, 3 and 17 are just numbers, and not sets.

Examining their given information reveals the origin of this discrepancy: by Johnny's definition of a successor function,

$$s_{\text{Johnny}}(n) := n \cup \{n\}$$

$$17 = \{0, 1, 2, \ldots, 16\}$$

By Johnny's definition of a successor function, 17 is the set of numbers less than it, so 3 is in 17.

$$s_{\text{Johnny}}(n) := n \cup \{n\}$$

$$17 = \{0, 1, 2, \ldots, 16\}$$

$$\text{so } 3 \in 17$$

But if we now look at what Ernie was given, this implies that 17 contains only 16. Clearly, 3 isn't in 17.

$$s_{\text{Ernie}}(n) := \{n\}$$

$$17 = \{16\}$$

But if we now look at what Ernie was given, this implies that 17 contains only 16. Clearly, 3 isn't in 17.

This isn't the only disagreement between the two systems either.

$$s_{\text{Ernie}}(n) := \{n\}$$

$$17 = \{16\}$$

$$\text{so } 3 \notin 17$$

But by Ernie's definition, 17 contains only 16. Clearly, 3 isn't in 17.

This isn't the only disagreement between the two systems either.

Proposition. *A set X has n elements if and only if there is a bijection between X and the natural number n.*

Johnny claims that a set X has n elements if and only if there is a bijection between X and the natural number n. And for Johnny, yes, this is true.

Proposition. *A set X has n elements if and only if there is a bijection between X and the natural number n.*

$$s_{\text{Johnny}}(n) := n \cup \{n\} \qquad s_{\text{Ernie}}(n) := \{n\}$$

$$n = \{0, 1, 2, \ldots, n-1\} \qquad n = \{n-1\}$$

Johnny claims that a set X has n elements if and only if there is a bijection between X and the natural number n. And for Johnny, yes, this is true.

Proposition. *A set X has n elements if and only if there is a bijection between X and the natural number n.*

$$s_{\text{Johnny}}(n) := n \cup \{n\} \qquad s_{\text{Ernie}}(n) := \{n\}$$

$$n = \{0, 1, 2, \ldots, n-1\} \qquad n = \{n-1\}$$

True for Johnny...

Johnny claims that a set X has n elements if and only if there is a bijection between X and the natural number n. And for Johnny, yes, this is true.

But for Ernie, every number contains only a single element (apart from zero, which is empty), so their notions of cardinality also disagree.

Proposition. *A set X has n elements if and only if there is a bijection between X and the natural number n.*

$$s_{\text{Johnny}}(n) := n \cup \{n\} \qquad s_{\text{Ernie}}(n) := \{n\}$$

$$n = \{0, 1, 2, \ldots, n-1\} \qquad n = \{n-1\}$$

True for Johnny... but not for Ernie.

But for Ernie, every number contains only a single element (apart from zero, which is empty), so their notions of cardinality also disagree.

The source of these disagreements is obvious:

$$s_{\text{Johnny}}(n) := n \cup \{n\}$$
$$s_{\text{Ernie}}(n) := \{n\}$$

The source of these disagreements is obvious: the difference between their successor functions, and by extension, the set N. But what is *not* obvious is how these disagreements should be reconciled.

Each account of the naturals is equally valid and correct in isolation, with neither one to be preferred over the other. That is, both constructions yield valid models of the Peano axioms, or the resulting semirings are isomorphic.

So, if we accept Johnny's account of the naturals, there is no good reason why we shouldn't also accept Ernie's.

Of course, we could choose to accept both accounts and agree that

$$\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\} = 3 = \{\{\{\emptyset\}\}\}$$

these sets are in fact equal, but this is clearly absurd.

The alternative is that at least one

$$s_{\text{Johnny}}(n) := n \cup \{n\} \qquad s_{\text{Ernie}}(n) := \{n\}$$

At least one of these must be "wrong"...

The alternative is that at least one of the two accounts is false.

The belief that there *is* a true account is called

Set-theoretic Platonism:

There is a "true" account; there is a particular set that is the "real" set of natural numbers.

The belief that there *is* a true account is called set-theoretic Platonism. This is the idea that there is a particular set N that is *really* the natural numbers, regardless of whether there exists an argument to prove this or not, or even if we can ever find it.

Set-theoretic Platonism:

There is a "true" account; there is a particular set that is the "real" set of natural numbers.

That is, there is a "correct" assignment of sets to

The belief that there *is* a true account is called set-theoretic Platonism. This is the idea that there is a particular set N that is *really* the natural numbers, regardless of whether there exists an argument to prove this or not, or even if we can ever find it.

Set-theoretic Platonism:

There is a "true" account; there is a particular set that is the "real" set of natural numbers.

That is, there is a "correct" assignment of sets to

$$N, \ 0, \ s$$

and all other assignments are wrong.

The belief that there *is* a true account is called set-theoretic Platonism. This is the idea that there is a particular set N that is *really* the natural numbers, regardless of whether there exists an argument to prove this or not, or even if we can ever find it.

Benacerraf rejects this as ridiculous.

"...if the number 3 is really one set rather than another, it must be possible to give some cogent reason for thinking so; for the position that this is an unknowable truth is hardly tenable.

If a number really is one specific set, then there should very well be a good reason for it.

But on the other hand, all these accounts

But there seems to be little to choose among the accounts. Relative to our purposes in giving an account of these matters, one will do as well as another, stylistic preferences aside."

But on the other hand, all these accounts agree on everything we care about the natural numbers.

For the purposes of describing the natural numbers, and the natural numbers only, any of these accounts would do, because they only differ in aspects that aren't inherent to natural numbers.

This is the idea behind *structuralism.*

Structuralism

Mathematics, as mathematicians actually use it, does not demand of the natural numbers that they exist as some specific object, but only that they have the structure we require. When we work with the natural numbers, we don't are about the specific construction used, only that they have semiring structure, that they support induction, etc.

In fact, this is how we usually describe and use objects in mathematics. For instance, vectors and tensors. A vector space is anything that satisfies the vector space axioms

A *vector space* over a field, K, is a set, V, along with two maps, $+ : V^2 \to V$ and $\cdot : K \times V \to V$, called *vector addition* and *scalar multiplication*, respectively, that satisfies the following axioms for all $\mathbf{u},\mathbf{v},\mathbf{w} \in V$ and $a,b \in K$:

(V1) $(V,+)$ is an abelian group.

- (A1) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ (commutativity of vector addition);
- (A2) $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$ (associative of vector addition);
- (A3) $\exists \mathbf{0}_V$ such that $\mathbf{v} + \mathbf{0}_V = \mathbf{0}_V + \mathbf{v} = \mathbf{v}$ (existence of vector additive identity);
- (A4) $\exists(-\mathbf{v}) \in V$ such that $\mathbf{v} + (-\mathbf{v}) = (-\mathbf{v}) + \mathbf{v} = \mathbf{0}_V$ (existence of vector addition inverses);
- (A5) $\mathbf{u} + \mathbf{v} \in V$ (closure of vector addition).

(V2) $a \cdot (\mathbf{u} + \mathbf{v}) = a \cdot \mathbf{u} + b \cdot \mathbf{v}$ (distributivity of scalar multiplication over vector addition);

(V3) $(a + b) \cdot \mathbf{v} = a \cdot \mathbf{v} + b \cdot \mathbf{v}$ (distributivity of scalar multiplication over field addition);

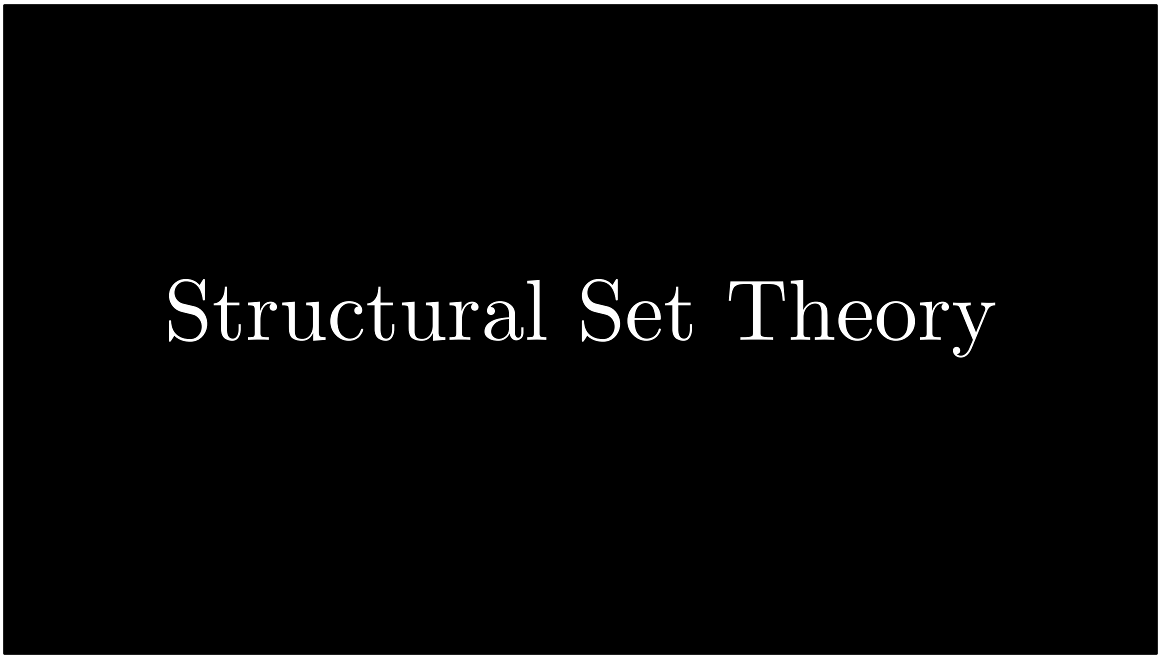
(V4) $(ab) \cdot \mathbf{v} = a \cdot (b\mathbf{v})$ (compatibility of scalar multiplication with field multiplication);

(V5) $1_K \cdot \mathbf{v} = \mathbf{v}$ (existence of scalar multiplicative identity).

A vector space is anything that satisfies the vector space axioms, and similarly, a tensor is anything that satisfies the tensor axioms.

In neither definition do we prescribe what the vector space or tensor itself is actually made of, only that it *behaves* in a certain way.

To the structuralist, mathematics is the study of structures independent of the things they are composed of. As seen in these definitions, this is the approach taken in many other mathematical contexts, so it is strange that the foundations of mathematics itself are commonly formulated in a way that is distinctly *not* structural in nature. But



But this doesn't have to be the case.

Structural Set Theory

$$3 \stackrel{?}{=} \{\{\{\emptyset\}\}\}$$

Structuralism tells us that it is meaningless to ask whether this equality holds or not, because it is not asked in the context of the rest of the natural numbers. For instance, we know that 3 isn't equal to 2 because, for instance, 2 is strictly less than 3, which is a property of the natural numbers.

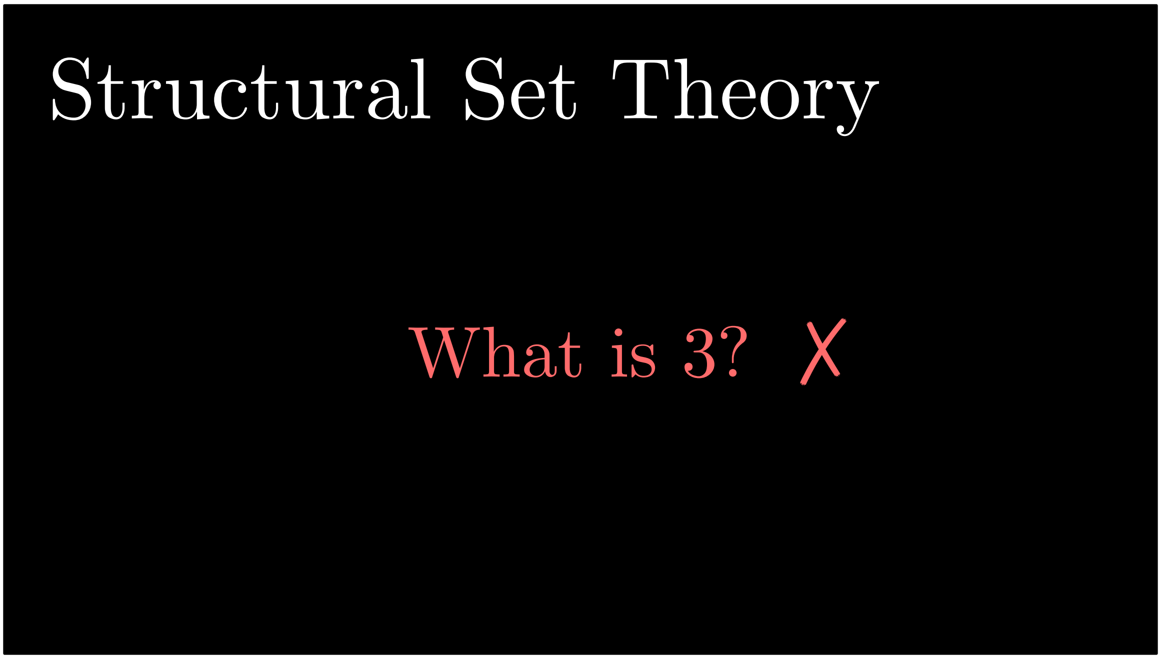
But it seems wrong to argue that 3 isn't equal to this set because, for instance, 3 has 3 elements (or no elements, or seventeen elements), while this set only has one, because *we don't know this.* The number of elements of 3 isn't a part of the structure of the natural numbers.

Structural Set Theory

What is 3?

What makes the number 3 the number 3 is precisely its relations to other natural numbers. And because the number 3 does not have any set relations, we argue that 3 is in fact, not a set at all.

To find out what exactly it is then, structuralism tells us that a more sensible question than “what is 3?” is



What makes the number 3 the number 3 is precisely its relations to other natural numbers. And because the number 3 does not have any set relations, we argue that 3 is in fact, not a set at all.

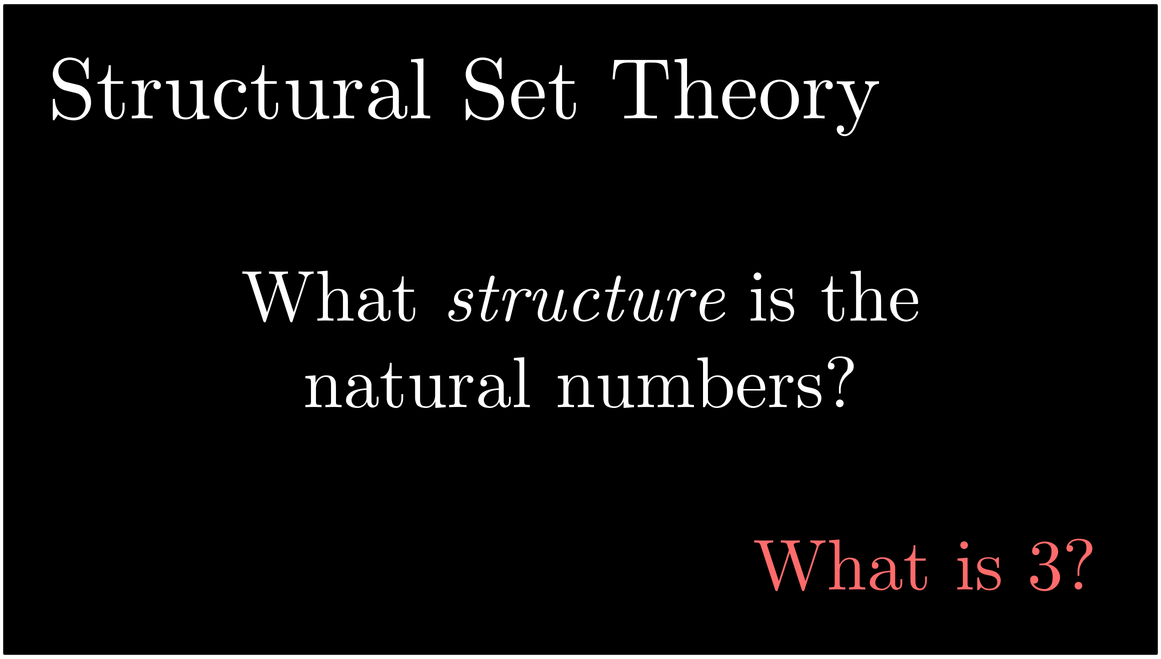
To find out what exactly it is then, structuralism tells us that a more sensible question than "what is 3?" is

Structural Set Theory

What are *all* the
natural numbers?

What is 3?

"What are *all* the natural numbers" or more precisely,



“What *structure* is the natural numbers?”

Structural Set Theory

What *structure* is the natural numbers?

What is 3?

“What *structure* is the natural numbers?”

Primitive Notions

[Animation Slide]

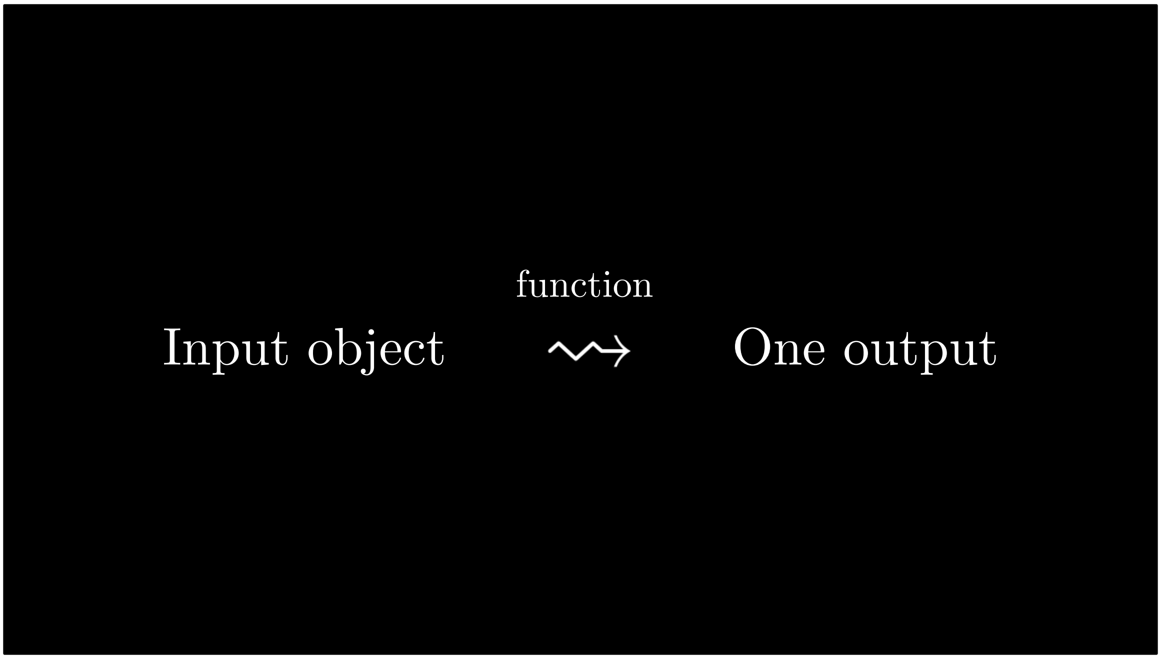
Material Set Theories

axiomatise

Sets
&
Membership

In ZFC, and other traditional axiomatisations of set theory, the basic primitive notions are of sets, elements, and membership, and everything else is derived from there.

Most of the questionable statements we've encountered so far predominantly arise from the membership relation being too strong. In practice, we generally only ever compare two sets when they're already contained in some ambient set or surrounding context.



Informally, a function is a special kind of correspondence between pairs of objects where each given input object maps to precisely one corresponding output object.

Material approach:

Represent a function $f : A \to B$ as the relation

$$\hat{f} = \big\{(a,b) : b \text{ is the } f\text{-image of } a\big\} \subseteq A \times B$$

Materially, we represent a function as a relation. Note that this first requires the construction of ordered pairs, then of cartesian products.

Material approach:

Represent a function $f : A \to B$ as the relation

$$\hat{f} = \big\{(a,b) : b \text{ is the } f\text{-image of } a\big\} \subseteq A \times B$$

Conversely, a relation R satisfies the property that

$$\big((x,y) \in R \wedge (x,z) \in R\big) \to y = z$$

then R is the representation of some function.

Then, given a relation R, we determine a condition on which ones represent functions.

This construction now encodes our informal notion of a function into sets.

Now, the next step is a trick commonly used in mathematics. We drop the distinction between the notion of a function, and its construction as a set, and we say that this representation is itself the *definition* of a function.

Material approach:

A function *is* a relation R satisfies the property that

$$\big((x, y) \in R \wedge (x, z) \in R\big) \rightarrow y = z$$

But for now, we note that this definition works well on a technical level, and much theory can be developed with it. But, there are some conceptual hurdles with this definition.

$$\mathrm{dom}(f) := \left\{ x \,\middle|\, \exists y : (x, y) \in f \right\}$$

$$\mathrm{im}(f) := \left\{ y \,\middle|\, \exists x : (x, y) \in f \right\}$$

$$\mathrm{cdm}(f) := \ ???$$

For instance, we can easily define the domain and image of the function, but there is no way to recover the codomain from this definition.

This is not a problem in some branches of mathematics, such as analysis, or even much of set theory, but in more algebraic or topological areas, this poses some difficulties.

Let $A \subset B$ and consider the functions

$$\mathrm{id}_A : A \to A \qquad \iota_A : A \hookrightarrow B$$

both defined by $x \mapsto x$.

For instance, let A be a strict subset of B, and consider these two maps. The former is an identity map, while the latter is an inclusion, with the different terminology indicating that we should view this function as including the elements of A into B. These functions are very conceptually distinct, but

Let $A \subset B$ and consider the functions

$$\mathrm{id}_A : A \to A \qquad \iota_A : A \hookrightarrow B$$

both defined by $x \mapsto x$. Then,

$$\mathrm{id}_A = \big\{(x, x) : x \in A\big\} = \iota_A$$

they are both the same set, and hence the same function, set theoretically.

Which is a bit weird, since we generally don't think of these functions as being the same: one just fixes everything in A, sending every element to itself, while the other embeds A into B.

This isn't just a conceptual problem, but also a practical one in some cases. If A is the circle and B is the complex plane, then these maps yield very different induced homomorphisms in first homology.

Material Set Theories

axiomatise

Sets

&

Membership

So, we've seen how functions are derived from these primitive notions, and some of the problems that arise from there.

Because functions are exactly how sets relate to each other, we'll actually be taking sets and *functions* to be our primitive notions in a structural set theory.

Structural Set Theories

axiomatise

Sets
&
Functions

Because functions are exactly how sets relate to each other, we'll actually be taking sets and *functions* to be our primitive notions in a structural set theory.

Moreover, functions are inherently attached to a pair of sets, so we will have much more *locality* in our set theory by starting with functions. Another side effect of this is that all of our constructions will also be isomorphism invariant.

All of this means that our theory will lend itself well to being described with the language of category theory.

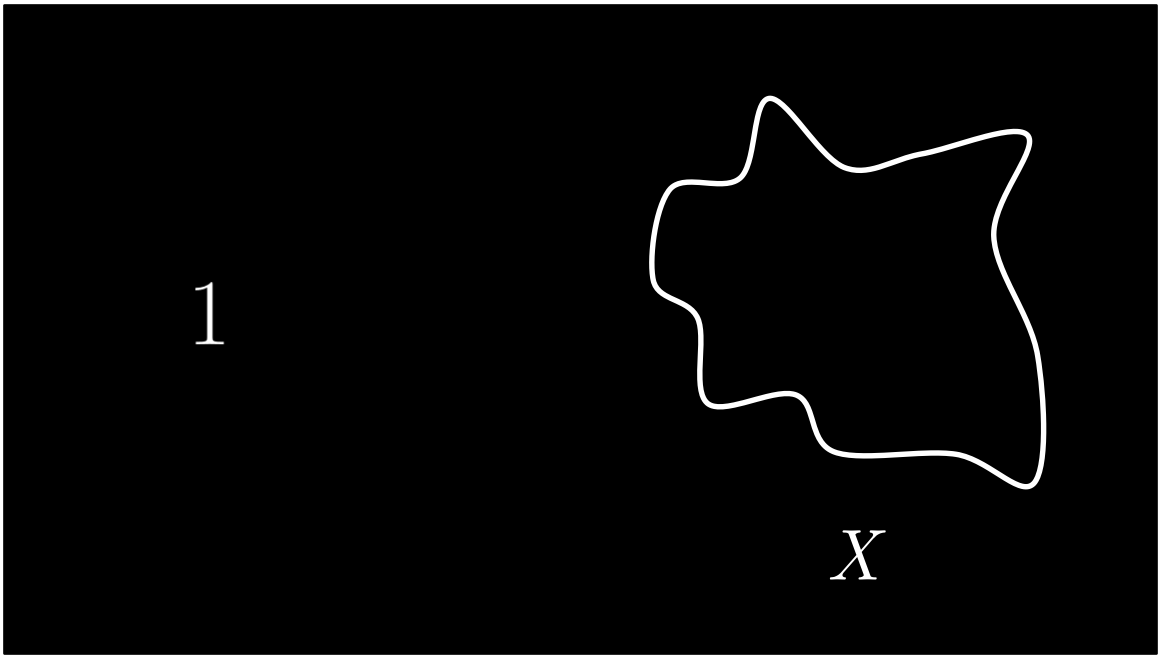
The Yoneda Lemma

We quickly review the Yoneda lemma, as it provides an important extensionality principle for structural sets.

$$1 = \{\bullet\}$$

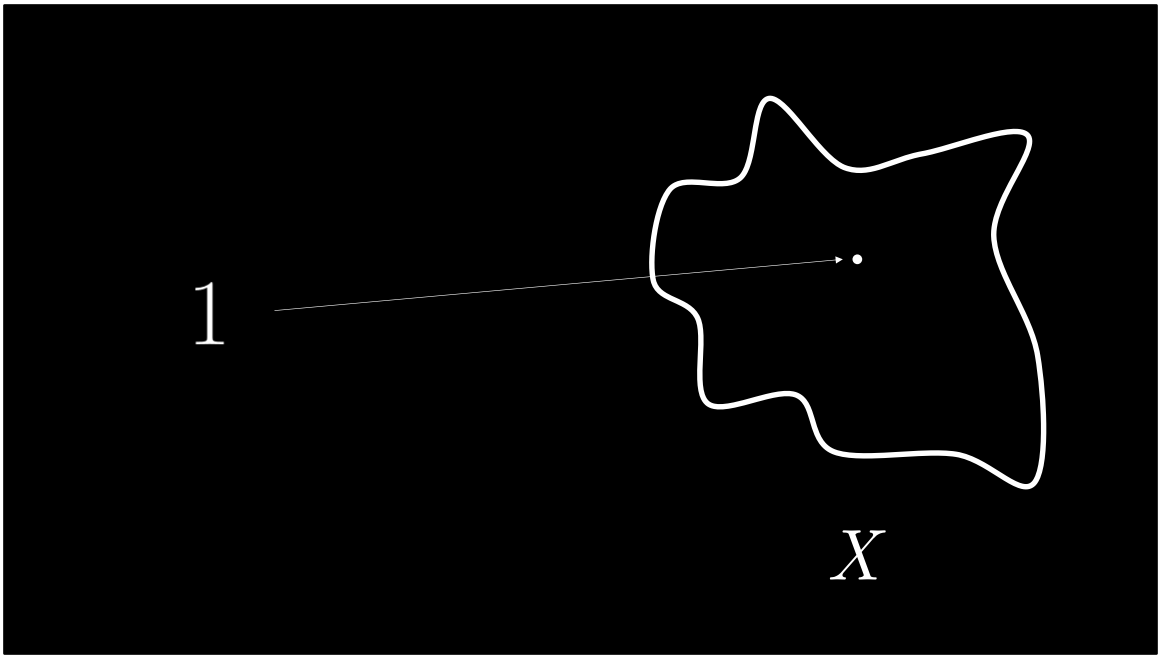
Let 1 be the set with one element.

For any set X, a function 1 to X amounts to selecting an element of X to be the image of the unique object in the domain.



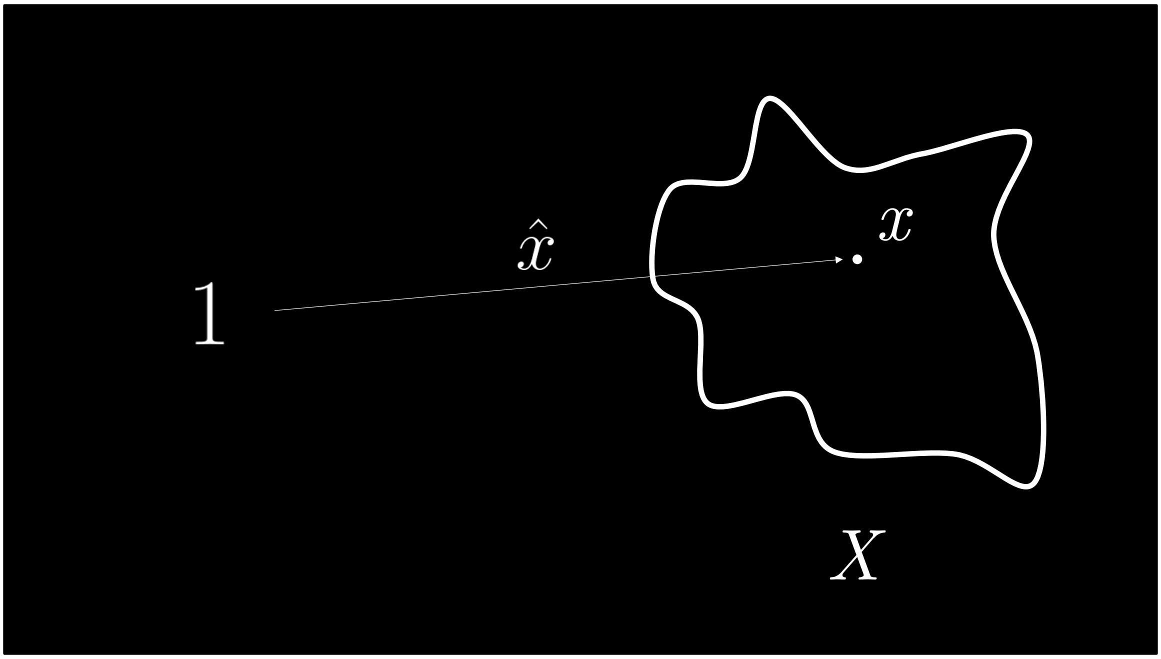
For any set X, a function 1 to X amounts to selecting an element of X to be the image of the unique object in the domain.

In fact, in arbitrary categories with terminal objects, we will define an *element* to be a map with the terminal object as domain.



For any set X, a function 1 to X amounts to selecting an element of X to be the image of the unique object in the domain.

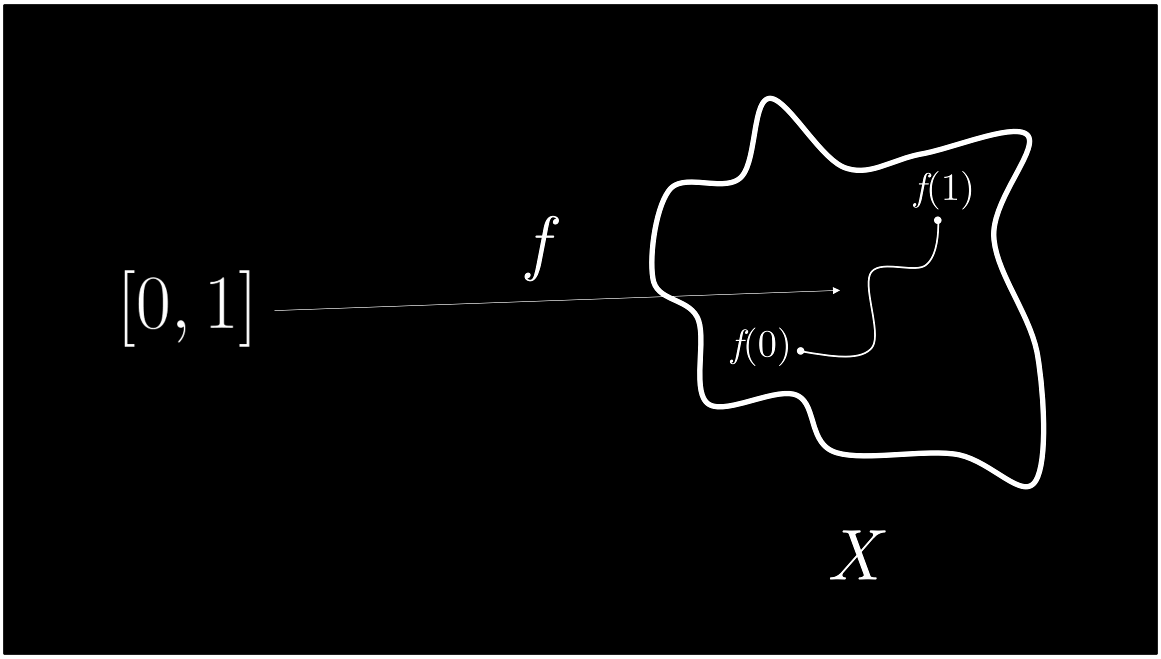
In fact, in arbitrary categories with terminal objects, we will define an *element* to be a map with the terminal object as domain.



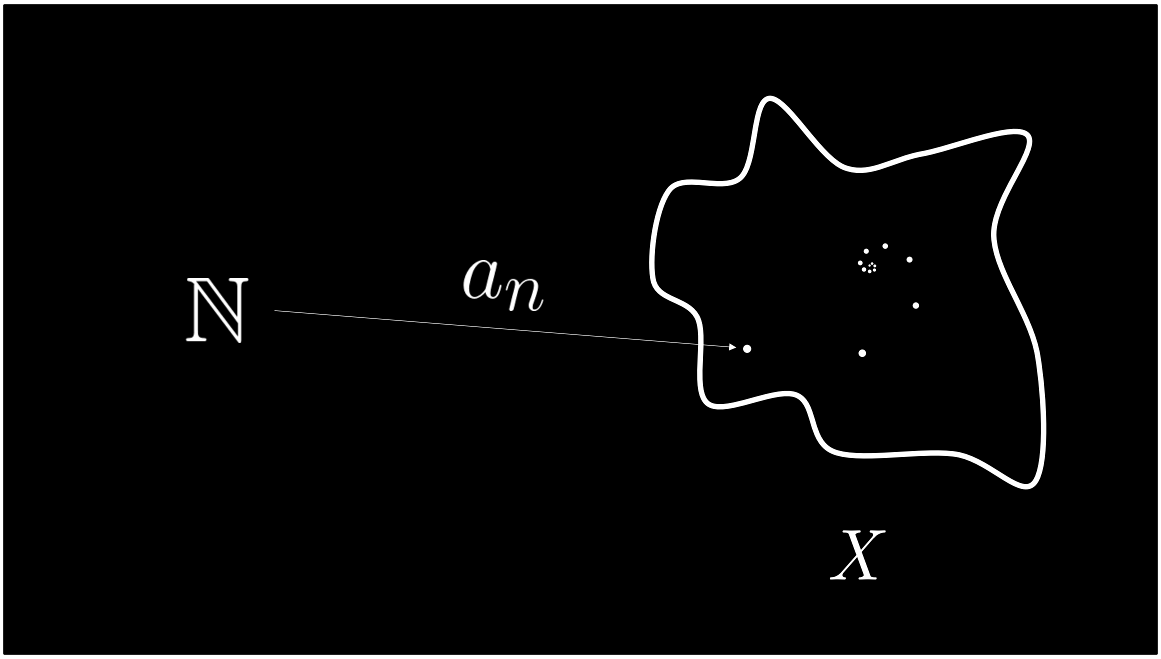
For any set X, a function 1 to X amounts to selecting an element of X to be the image of the unique object in the domain.

In fact, in arbitrary categories with terminal objects, we will define an *element* to be a map with the terminal object as domain.

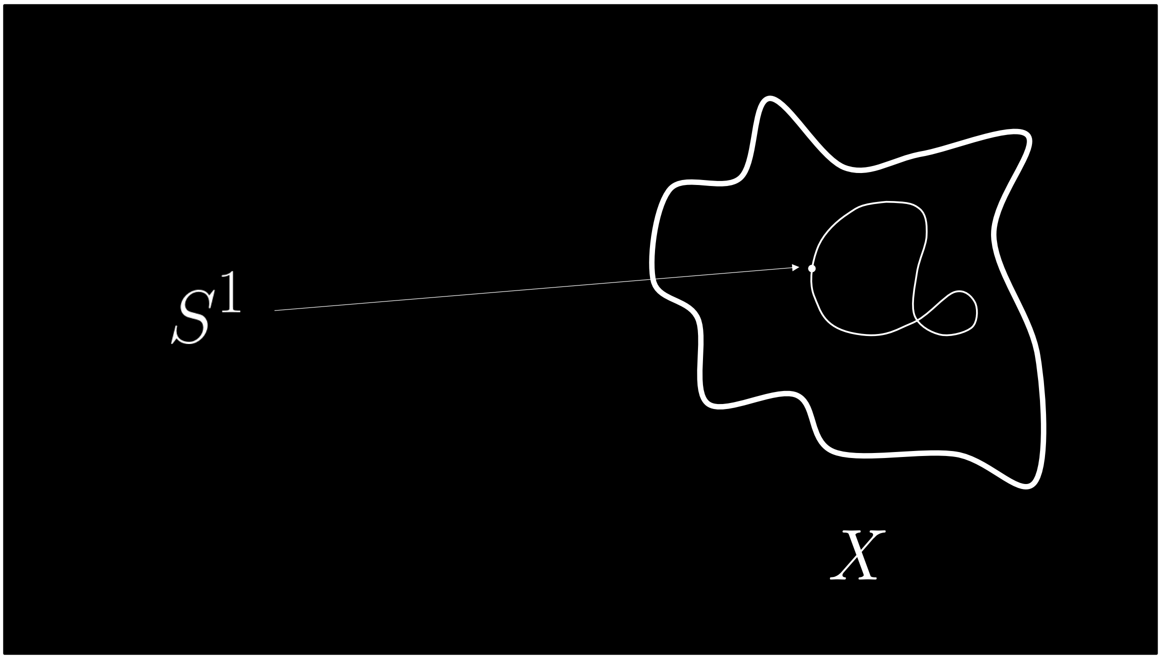
Anyway, we can use maps to capture other features of a space that are more complicated than single points. For instance, maps from the unit interval



For instance, maps from the unit interval [0,1] to X are just paths in X, while maps from the natural numbers



while maps from the natural numbers to X are the sequences in X, and the functions



the functions from the circle to X are the topological loops in X.

More generally, given any object A in a category, a *generalised element* of A of *shape* S is a morphism S to A.

$$S \to X$$

$$\leftrightsquigarrow$$

"Generalised element of X of shape S"

More generally, given any object A in a category, a *generalised element* of A of *shape* S is a morphism S to A.

Now, in material set theory, sets are characterised completely by their elements. To what extent are arbitrary objects X characterised by their *generalised* elements?

If X is, say, a topological space, then...

$$X = \{S, \mathcal{T}\}$$

If X is, say, a topological space, then these maps from different domains indirectly give us information about X. We get the points back from the singleton set, path-connected components from the interval, and the fundamental group from the circle.

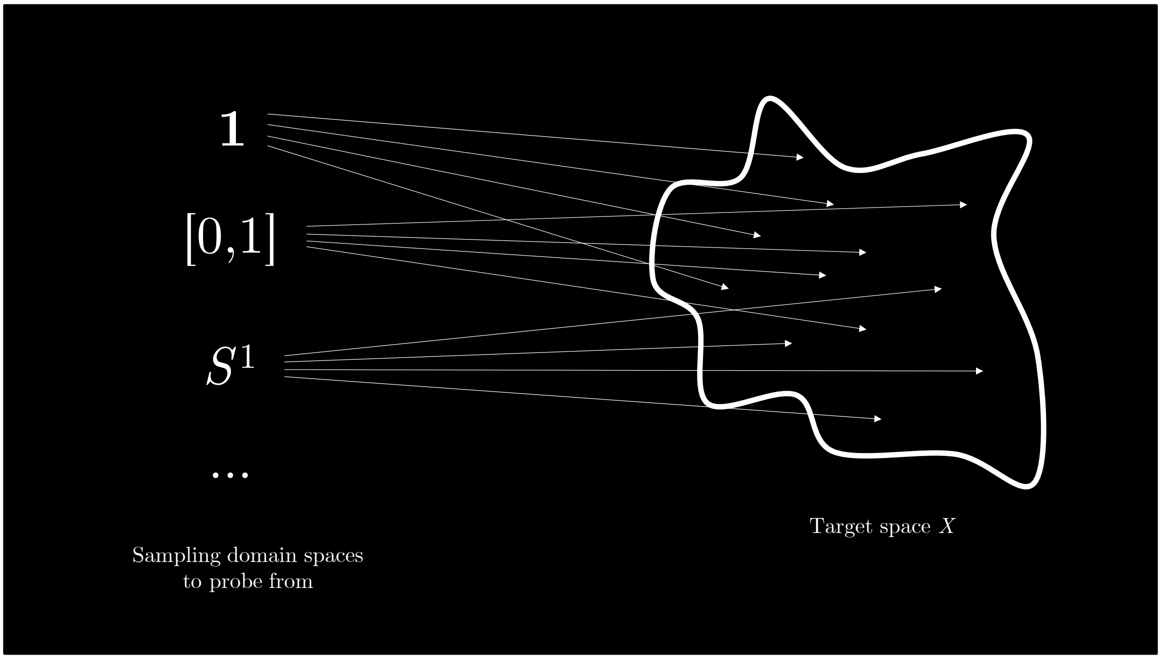
$$X = \{S, \mathcal{T}\}$$

$$\{1 \to X\} \cong S$$

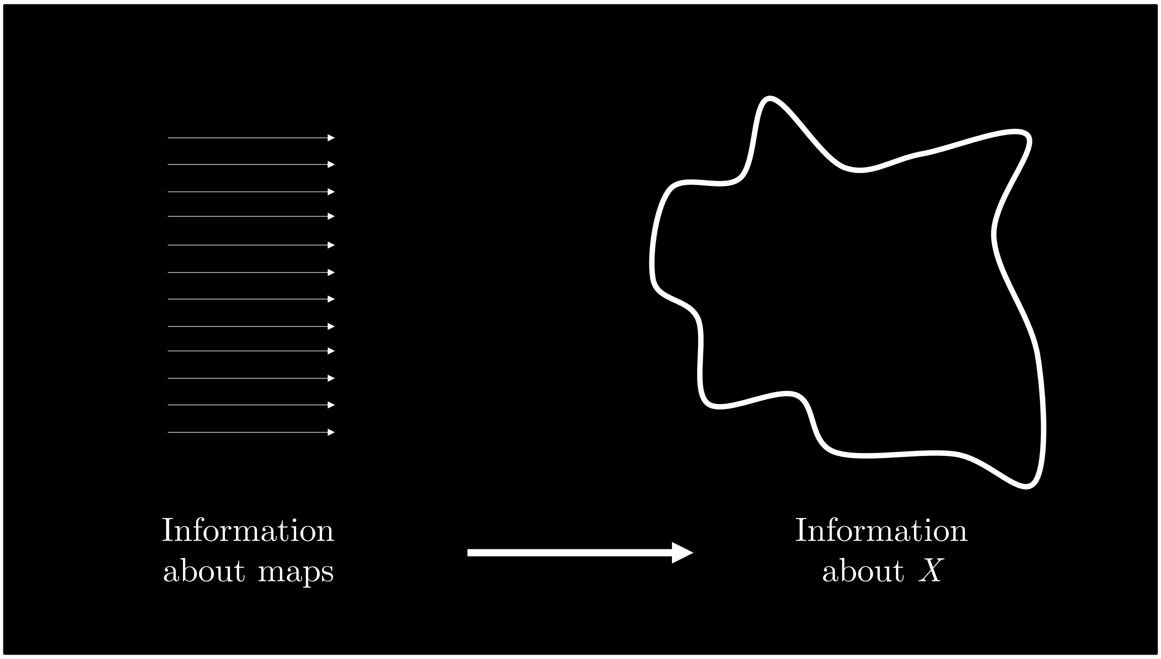
$$\{[0,1] \to X\} \rightsquigarrow H_0(X)$$

$$\{S^1 \to X\} \rightsquigarrow \pi_1(X)$$

If X is, say, a topological space, then these maps from different domains indirectly give us information about X. We get the points back from the singleton set, path-connected components from the interval, and the fundamental group from the circle.



The point is, we get more and more information about any space X by examining how it appears from more and more other probing spaces. But exactly how much information can we recover?



Is it always possible to obtain as much data from looking at maps as we would from just analysing the space itself? After all, we have no reason to expect that the entire structure of the space is always captured by these maps.

Except,...

Information about maps $\cong$ Information about X

Except, it always is. And that, is the Yoneda lemma.

Lemma (Yoneda)**.** *Let $\mathcal{C}$ be a locally small category. Then,*

$$\hom_{[\mathcal{C},\mathbf{Set}]}(H_A, F) \cong F(A)$$

naturally in $F \in \mathrm{ob}([\mathcal{C}, \mathbf{Set}])$ *and* $A \in \mathrm{ob}(\mathcal{C})$.

Here's the precise formulation of the lemma, but we'll focus on one particular corollary of the lemma.

Corollary.

$$X \cong Y \textit{ if and only if } \hom(X, -) \cong \hom(Y, -)$$

Two objects are isomorphic if and only if their hom-functors are also naturally isomorphic.

While sets are characterised by elements, arbitrary objects are characterised, up to isomorphism, by their generalised elements.

Subobjects

The next question is, how could we characterise *subsets* with maps?

$$\{1\},\ \{2\},\ \{\text{cat}\}$$

$$X := \{1, 2, 3\}$$

Consider these four sets. The three singletons containing 1, 2, and the word cat, and the set X containing 1, 2 and 3.

$$\{1\} \cong \{2\} \cong \{\text{cat}\}$$

Because they're all singletons, they're all isomorphic. However, look at what happens when we apply the subset relation to these sets and X.

$$\{1\} \subseteq X$$

$$\{2\} \subseteq X$$

$$\{\text{cat}\} \nsubseteq X$$

The subset relation on sets is not isomorphism invariant! This is another side effect of the membership relation being too strong.

Categories don't care about how we label elements.

What we *do* care about is how these sets *embed* into X or not.

$$A \rightarrowtail X$$

Consider the class of monomorphisms into an object X. We can define a preorder on this class as follows.

$$f : A \rightarrowtail X \qquad g : B \rightarrowtail X$$

Given two monomorphisms f and g, we write f less than g if f factors through g. That is, if there exists a morphism from A to B such that this triangle commutes.

$$f : A \rightarrowtail X \qquad g : B \rightarrowtail X$$

Given two monomorphisms f and g, we write f less than g if f factors through g. That is, if there exists a morphism from A to B such that this triangle commutes.

$$f \leq g$$

$$\begin{tikzcd} & B \arrow[d, tail, "g"] \\ A \arrow[ur] \arrow[r, tail, "f"'] & X \end{tikzcd}$$

Given two monomorphisms f and g, we write f less than g if f factors through g. That is, if there exists a morphism from A to B such that this triangle commutes.

Note that this factorisation is necessarily unique if it exists since g is monic and equalises anything that makes this commute.

$$f \leq g$$

$$\begin{tikzcd} & B \arrow[d, tail, "g"] \\ A \arrow[ur] \arrow[r, tail, "f"'] & X \end{tikzcd}$$

Given two monomorphisms f and g, we write f less than g if f factors through g. That is, if there exists a morphism from A to B such that this triangle commutes.

Note that this factorisation is necessarily unique if it exists since g is monic and equalises anything that makes this commute.

$$f \leq g \qquad\qquad\qquad g \leq f$$

$$\begin{tikzcd} & B \arrow[d, tail, "g"] \\ A \arrow[ur] \arrow[r, tail, "f"'] & X \end{tikzcd} \qquad\qquad \begin{tikzcd} & B \arrow[dl] \arrow[d, tail, "g"] \\ A \arrow[r, tail, "f"'] & X \end{tikzcd}$$

Then, if we have f less than g and g less than f, then this factorisation constitutes an isomorphism between A and B, and we say that f and g are isomorphic morphisms.

$$f \leq g \qquad\qquad g \leq f$$

$$\begin{tikzcd}
& B \arrow[d, tail, "g"] \\
A \arrow[ur] \arrow[r, tail, "f"'] & X
\end{tikzcd}
\qquad\qquad
\begin{tikzcd}
& B \arrow[d, tail, "g"] \arrow[dl] \\
A \arrow[r, tail, "f"'] & X
\end{tikzcd}$$

Then, if we have f less than g and g less than f, then this factorisation constitutes an isomorphism between A and B, and we say that f and g are isomorphic morphisms.

$$f \cong g$$

$$\begin{tikzcd} & B \arrow[d, tail, "g"] \\ A \arrow[ur] \arrow[r, tail, "f"'] & X \end{tikzcd} \qquad \begin{tikzcd} & B \arrow[d, tail, "g"] \arrow[dl] \\ A \arrow[r, tail, "f"'] & X \end{tikzcd}$$

Then, if we have f less than g and g less than f, then this factorisation constitutes an isomorphism between A and B, and we say that f and g are isomorphic morphisms.

In other words, we're looking at the slice category of monomorphisms over X.

A *subobject* of an object X is an isomorphism class of monomorphisms into X.

A subobject of an object X is then an isomorphism class of monomorphisms into X.

Given a monomorphism

$$S : A \rightarrowtail X$$

we write

$$[S] \subseteq X$$

for the subobject represented by S.

Given a monomorphism S from A to X, we write this for the represented subobject. Through a small abuse of notation, we sometimes pick a representative monomorphism and call that a subobject.

Let's go back to our previous example.

$$\{1\} \cong \{2\} \cong \{\text{cat}\}$$

$$X := \{1, 2, 3\}$$

This time, we look at injections from each of the sets into X.

$$f : \{1\} \rightarrowtail X : 1 \mapsto 1$$

$$g : \{2\} \rightarrowtail X : 2 \mapsto 1$$

$$h : \{\text{cat}\} \rightarrowtail X : \text{cat} \mapsto 1$$

Consider these three functions, mapping each of the unique elements in the singletons to the element 1 in X.

They all have the same image, so they all factor through each other, and they all witness the same subset, namely, the subset containing the element 1.

$$f : \{1\} \rightarrowtail X : 1 \mapsto 1$$

$$g : \{2\} \rightarrowtail X : 2 \mapsto 1$$

$$h : \{\text{cat}\} \rightarrowtail X : \text{cat} \mapsto 1$$

Consider these three functions, mapping each of the unique elements in the singletons to the element 1 in X.

They all have the same image, so they all factor through each other, and they all witness the same subset, namely, the subset containing the element 1.

$$k : \{1\} \rightarrowtail X : 1 \mapsto 2$$
$$f : \{1\} \rightarrowtail X : 1 \mapsto 1$$

Now consider this function that instead has image 2. This time, there's no way to factor these morphisms through each other, because they don't agree in their images. So, this represents a different subset of X, the one containing the element 2.

This is really what distinguishes the two *in the context of being a subset of* X.

This also helps to resolve some questions that arise in material set theory. For instance, “is $\mathbb{Z} \subset \mathbb{R}$?”

“are the integers a subset of the reals?”

Material set theory:

$\mathbb{Z}$

Equivalence classes of natural numbers

$\mathbb{R}$

Equivalence classes of Cauchy sequences

Or Dedekind cuts.

Or ultrafilters on $\mathbb{N}$.

In material set theory, integers are generally constructed from equivalence classes of natural numbers, and the reals can be constructed as equivalence classes of Cauchy sequences, or through various other constructions.

In any case, it is clear that the integers are not a subset of the reals. But, when interpreted in the conventional non-set-theoretic way, every integer is clearly a real number.

Material set theory:

$\mathbb{Z}$

Equivalence classes
of natural numbers

$\mathbb{R}$

Equivalence classes of
Cauchy sequences

Or Dedekind cuts.

Or ultrafilters on $\mathbb{N}$.

So, $\mathbb{Z} \not\subset \mathbb{R}$.

In material set theory, integers are generally constructed from equivalence classes of natural numbers, and the reals can be constructed as equivalence classes of Cauchy sequences, or through various other constructions.

In any case, it is clear that the integers are not a subset of the reals. But, when interpreted in the conventional non-set-theoretic way, every integer is clearly a real number.

Structuralism:

Asking if $\mathbb{Z} \subset \mathbb{R}$ or not because of their *elements* is not the right question.

The structuralist says that it doesn't make sense to ask whether Z is a subset of R or not because their *elements* are the same or not, but rather, Structuralism:

Asking if $\mathbb{Z} \subset \mathbb{R}$ or not because of their *elements* is not the right question.

Rather, ask if there is a map $\mathbb{Z} \to \mathbb{R}$ that *witnesses* that $\mathbb{Z} \subset \mathbb{R}$.

The structuralist says that it doesn't make sense to ask whether Z is a subset of R or not because their *elements* are the same or not, but rather, is there a *map* from Z to R that *witnesses* that Z is a subset of R.

In this case, yes, there do exist monomorphisms into R that pick out the integers.

We write

$$[f] \subseteq_X [g]$$

if (any) representing monomorphisms satisfy

$$f \leq g$$

We can also put a partial ordering on subsets of a fixed set X. We write this, if the representing monomorphism f factors through g.

This notation is rather suggestive, and in fact, this partial order agrees with the subset relation in material set theory, apart from the fact that it is local to an ambient containing set X.

Now, what about membership?

We say that an element $x \in X$ *is a member* of a subset $a \subseteq X$ and write $x \in_X a$ if x factors through a.

We say that an element x *is a member* of a subset a and write x in a if x factors through a.

We say that an element $x \in X$ *is a member* of a subset $a \subseteq X$ and write $x \in_X a$ if x factors through a.

$$\begin{tikzcd}
& S \arrow[d, "a"] \\
1 \arrow[ur, "\bar{x}"] \arrow[r, "x"'] & X
\end{tikzcd}$$

That is, there exists an element x bar of the domain of a that is sent to x under a. You can think of this as x being in the image of a.

Again, this definition of membership is local to the ambient set X, so unlike in a material set theory, it doesn't make sense to ask if x in y for arbitrary sets x and y without an ambient set for context.

The Subobject Classifier

Now, in Set, another way to characterise a subset A of a given set X is as a function X to 2

$$A \subseteq X$$

$$\updownarrow$$

$$\chi_A : X \to 2$$

where $2 = \{\top, \bot\}$

Now, in Set, another way to characterise a subset A of a given set X is as a function X to 2, where 2 is the two point set,

$$A \subseteq X$$

$$\chi_A(x) = \begin{cases} \top & x \in A \\ \bot & x \notin A \end{cases}$$

by taking the indicator function of A.

In fact, this process is a bijection, and the reverse construction is given by mapping each indicator function to the preimage of the true element.

$$\chi_A \overset{\cong}{\mapsto} A := \chi_A^{-1}[\{\top\}]$$

In fact, this process is a bijection, and the reverse construction is given by mapping each indicator function to the preimage of the true element.

Now, recall that the preimage is a special case of a pullback, so this bijection says that for every subset A of X, there is a unique function from X to 2 such that this diagram is a pullback.

$$
\begin{tikzcd}
A = f^{-1}[\{\top\}] \arrow[r, "!"] \arrow[d, hook] \arrow[dr, phantom, "\lrcorner", very near start] & 1 \arrow[d, hook, "\top"] \\
X \arrow[r, "\chi"'] & 2
\end{tikzcd}
$$

Now, recall that the preimage is a special case of a pullback, so this bijection says that for every subset A of X, there is a unique function from X to 2 such that this diagram is a pullback.

Nothing in this diagram is specific to Set, so we can abstract it into any arbitrary category that admits pullbacks and has a terminal object.

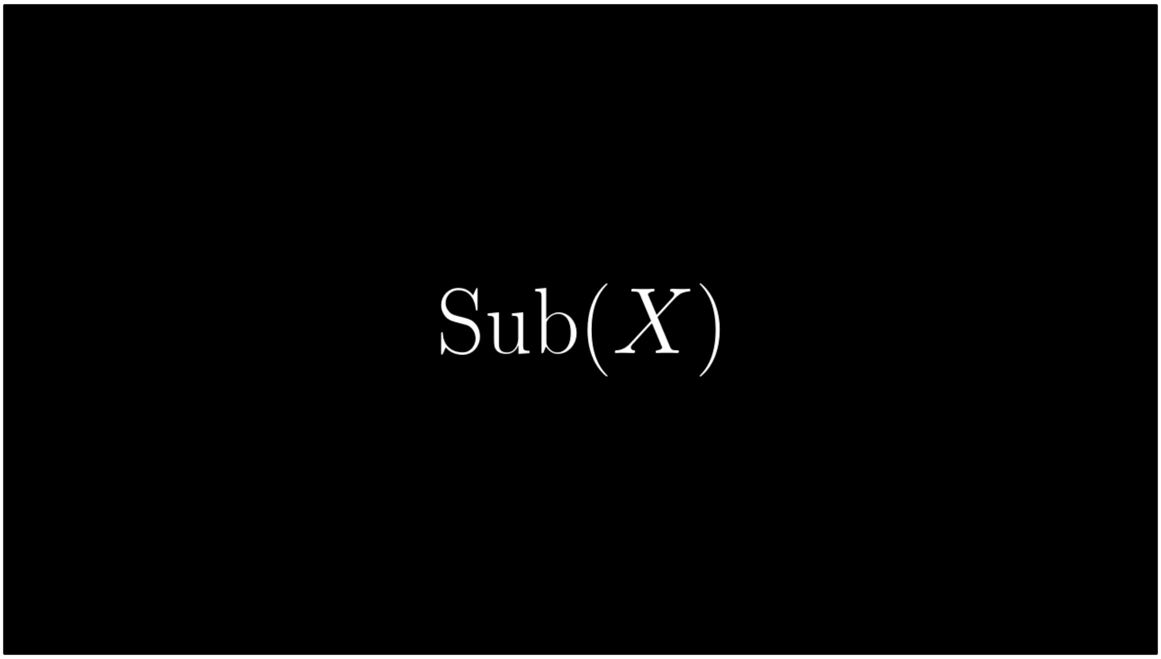
A *subobject classifier* in a category $\mathscr{C}$ is an object Ω and a map $\top : 1 \to \Omega$ such that for every monomorphism $m : A \rightarrowtail X$, there exists a unique morphism $\chi_m : X \to \Omega$ such that

$$\begin{tikzcd} A \arrow[r, "!"] \arrow[d, tail, "m"'] \arrow[dr, phantom, "\lrcorner", very near start] & 1 \arrow[d, "\top"] \\ X \arrow[r, "\chi_m"'] & \Omega \end{tikzcd}$$

is a pullback square.

A *subobject classifier* is an object Omega and map True such that for every subobject of X, there is a unique map from X into Omega such that this diagram is a pullback.

We give another characterisation of the subobject classifier. Let X be an object in a category C. We write



We write Sub(X) for the class of subobjects of X, called the *subobject poset* of X.

If C has finite limits and is locally small, then every map f from X to Y induces a map between their subobject posets in the reverse direction by pullback.

$$f : X \to Y$$

$$\mathrm{Sub}(X)$$

We write $\mathrm{Sub}(X)$ for the class of subobjects of X, called the *subobject poset* of X.

If C has finite limits and is locally small, then every map f from X to Y induces a map between their subobject posets in the reverse direction by pullback.

$$f : X \to Y$$

$$f^* : \mathrm{Sub}(Y) \to \mathrm{Sub}(X)$$

We write $\mathrm{Sub}(X)$ for the class of subobjects of X, called the *subobject poset* of X.

If C has finite limits and is locally small, then every map f from X to Y induces a map between their subobject posets in the reverse direction by pullback.

$$
\begin{tikzcd}
X \times_Y B \arrow[r, "\pi_2"] \arrow[d, tail, "\pi_1"'] \arrow[dr, phantom, very near start, "\lrcorner"] & B \arrow[d, tail, "g"] \\
X \arrow[r, "f"'] & Y
\end{tikzcd}
$$

In detail, monomorphisms are stable under pullback, so taking the pullback of a subobject of Y along f gives another monomorphism into X, which represents a subobject of X.

$$\mathrm{Sub} : \mathscr{C}^{\mathrm{op}} \to \mathbf{Set}$$

This defines a contravariant functor from C to **Set**, and the subobject classifier is precisely a representation of this functor.

That is,

$$\mathrm{Sub} : \mathscr{C}^{\mathrm{op}} \to \mathbf{Set}$$

$$\mathrm{Sub}(X) \cong \hom(X, \Omega)$$

That is, this isomorphism holds naturally in X. This intuitively corresponds to the previous idea that subsets of X correspond to the maps into the subobject classifier.

$$A \rightarrowtail X \qquad B \rightarrowtail X$$

Subobject posets also allow us to abstract various other familiar set operations to subsets.

For instance, given two subobjects A and B, we can construct their union and intersection as pullbacks and pushouts of their representing monomorphisms

$$A \cap_X B := f \times_X g$$

$$A \cup_X B := f \amalg_X g$$

we can construct their union and intersection as pullbacks and pushouts of their representing monomorphisms

Note again that these operations are local to a containing ambient set X.

For some intuition on why this is true, the intersection of two subobjects with representing monomorphisms should be the maximal subobject that factors through both f and g, which is precisely their order-theoretic meet in the subobject poset.

But, interpreting the poset as a category, a meet is precisely a categorical product. Moreover, the subobject poset of a fixed object is a slice category, so products there are fibred products, or pullbacks, in the base category.

Monoidal Categories

We also take a quick aside to discuss monoidal categories.

A *monoidal category* $(\mathscr{C}, \otimes, I, \alpha, \lambda, \rho)$ consists of:

- A category $\mathscr{C}$;
- A bifunctor $\otimes : \mathscr{C} \times \mathscr{C} \to \mathscr{C}$ called the *tensor product*, written in infix notation;
- A designated object I in $\mathscr{C}$ called the *unit*;
- A natural isomorphism $\alpha : ((-) \otimes (-)) \otimes (-) \Rightarrow (-) \otimes ((-) \otimes (-))$ with components of the form $\alpha_{A,B,C} : (A \otimes B) \otimes C \to A \otimes (B \otimes C)$ called the *associator*;
- A natural isomorphism $\lambda : I \otimes (-) \Rightarrow (-)$ with components of the form $\lambda_A : (I \otimes A) \to A$ called the *left unitor*;
- A natural isomorphism $\rho : (-) \otimes I \Rightarrow (-)$ with components of the form $\rho_A : (A \otimes I) \to A$ called the *right unitor*;

subject to the *coherence conditions* that the following diagrams commute:

- the *triangle identity*:

$$\begin{tikzcd}
(A \otimes I) \otimes B \arrow[rr, "\alpha_{A,I,B}"] \arrow[dr, "\rho_A \otimes \mathrm{id}_B"'] & & A \otimes (I \otimes B) \arrow[dl, "\mathrm{id}_A \otimes \lambda_A"] \\
& A \otimes B &
\end{tikzcd}$$

- the *pentagon identity*:

$$\begin{tikzcd}
& (A \otimes B) \otimes (C \otimes D) \arrow[dr, "\alpha_{A,B,C \otimes D}"] & \\
((A \otimes B) \otimes C) \otimes D \arrow[ur, "\alpha_{A \otimes B,C,D}"] \arrow[d, "\alpha_{A,B,C} \otimes \mathrm{id}_D"'] & & A \otimes (B \otimes (C \otimes D)) \\
(A \otimes (B \otimes C)) \otimes D \arrow[rr, "\alpha_{A,B \otimes C,D}"'] & & A \otimes ((B \otimes C) \otimes D) \arrow[u, "\mathrm{id}_A \otimes \alpha_{B,C,D}"']
\end{tikzcd}$$

A monoidal category is a category with a structure similar to an algebraic monoid.

We have a bifunctor from the category to itself called the tensor product, and we distinguish a special object called the unit. Then, we have a few natural isomorphisms that ensure that the bifunctor acts like a monoidal operation.

For instance, the associator says exactly that the tensor product is associative, up to isomorphism, while the left and right unitors ensure that tensoring by the unit leaves the object unchanged, again up to isomorphism.

However, because these only hold up to isomorphism, we have to have some additional coherence conditions that ensure that the tensor behaves as we would expect. For instance,

$$(a \times e) \times b$$

consider this expression in an ordinary monoid, where e is the identity. By the definition of the identity, a times e is equal to a, so this expression reduces to

$$(a \times e) \times b = a \times b$$

a times b. We could also rebracket the expression to obtain

$$(a \times e) \times b = a \times b$$

a times b. We could also rebracket the expression to obtain

$$(a \times e) \times b = a \times (e \times b) = a \times b$$

$$(a \times e) \times b = a \times b$$

another chain of equalities. Let's arrange this in a triangle.

$$(a \times e) \times b = a \times (e \times b) = a \times b$$

$$(a \times e) \times b = a \times b$$

[Animation Slide]

$$
\begin{tikzcd}
(a \times e) \times b \arrow[rr, "="] \arrow[dr, "="'] & & a \times (e \times b) \arrow[dl, "="] \\
& a \times b &
\end{tikzcd}
$$

Obviously, this triangle commutes as everything is equal. But if we replace this expression with its analogue in a monoidal category, we have

$$
\begin{tikzcd}
(A \otimes I) \otimes B \arrow[rr, "\alpha_{A,I,B}"] \arrow[dr, "\rho_A \otimes \mathrm{id}_B"'] & & A \otimes (I \otimes B) \arrow[dl, "\mathrm{id}_A \otimes \lambda_B"] \\
& A \otimes B &
\end{tikzcd}
$$

this triangle involving the associator and the left and right unitors. Without any additional constraints on these isomorphisms, there is no reason that we should expect that this diagram commutes.

The triangle identity is precisely the requirement that this triangle *does* commute, so these objects are all uniquely isomorphic to each other.

$$
\begin{tikzcd}
& (A \otimes B) \otimes (C \otimes D) \arrow[dr, "\alpha_{A,B,C \otimes D}"] & \\
((A \otimes B) \otimes C) \otimes D \arrow[ur, "\alpha_{A \otimes B,C,D}"] \arrow[d, "\alpha_{A,B,C} \otimes \mathrm{id}_D"'] & & A \otimes (B \otimes (C \otimes D)) \arrow[d, "\mathrm{id}_A \otimes \alpha_{B,C,D}"] \\
(A \otimes (B \otimes C)) \otimes D \arrow[rr, "\alpha_{A,B \otimes C,D}"'] & & A \otimes ((B \otimes C) \otimes D)
\end{tikzcd}
$$

Similarly, the pentagon identity just ensures that every way we rebracket an expression yields objects isomorphic up to unique isomorphism.

Now,

So, for an example of a monoidal category, the category of sets is monoidal:

$$(A \times B) \times C \cong A \times (B \times C)$$

$$1 \times A \cong A \qquad A \times 1 \cong A$$

Here, the tensor product is given by the cartesian product of sets.

These sets are not equal, but there is an obvious isomorphism between them that we can use as the associator. The unit object is given by any one-point set, and the left and right unitors are given by these isomorphisms.

Because the tensor product is given by the categorical product, this category is called *cartesian monoidal*.

Now, the cartesian product also has another special property in that it is commutative, up to isomorphism. Because of this, the structure of this particular tensor product actually behaves more like a commutative monoid, and we call the category *symmetric* monoidal.

You can have symmetric monoidal categories that aren't cartesian, but we need another set of coherence diagrams to express this definition formally, but for now, the important point is that cartesian monoidal categories are always symmetric.

[Categories can have monoidal structure in multiple ways as well. For instance, Set is also monoidal with the tensor product given by the disjoint union:]

$$(A \times B) \times C \cong A \times (B \times C)$$

$$1 \times A \cong A \qquad A \times 1 \cong A$$

Here, the tensor product is given by the cartesian product of sets.

These sets are not equal, but there is an obvious isomorphism between them that we can use as the associator. The unit object is given by any one-point set, and the left and right unitors are given by these isomorphisms.

Because the tensor product is given by the categorical product, this category is called *cartesian monoidal*.

Now, the cartesian product also has another special property in that it is commutative, up to isomorphism. Because of this, the structure of this particular tensor product actually behaves more like a commutative monoid, and we call the category *symmetric* monoidal.

You can have symmetric monoidal categories that aren't cartesian, but we need another set of coherence diagrams to express this definition formally, but for now, the important point is that cartesian monoidal categories are always symmetric.

[Categories can have monoidal structure in multiple ways as well. For instance, Set is also monoidal with the tensor product given by the disjoint union:]

$$(A \sqcup B) \sqcup C \cong A \sqcup (B \sqcup C)$$

$$\emptyset \sqcup A \cong A \qquad A \sqcup \emptyset \cong A$$

Again, these sets are not strictly equal, but there is a canonical isomorphism between them we can use as the associator. The unit is given by the empty set, and the unitors by these isomorphisms.

All of this is to say, a monoidal structure allows us to suppress the usage of brackets since everything associates uniquely due to coherence.

Internalisation

Recall the standard definition of a group.

A *group* $(G, *)$ is a set G equipped with a binary operation $* : G \times G \to G$ that is associative, admits an identity element $e \in G$ (is *unitary*), and every element $g \in G$ has an inverse $g^{-1} \in G$ under $*$.

A group $(G$, ast) is a set G equipped with a binary operation that is associative, admits an identity element, and every element has an inverse under the operation.

By now, we should be used to viewing various constructions as morphisms, and we might be tempted to do the same here.

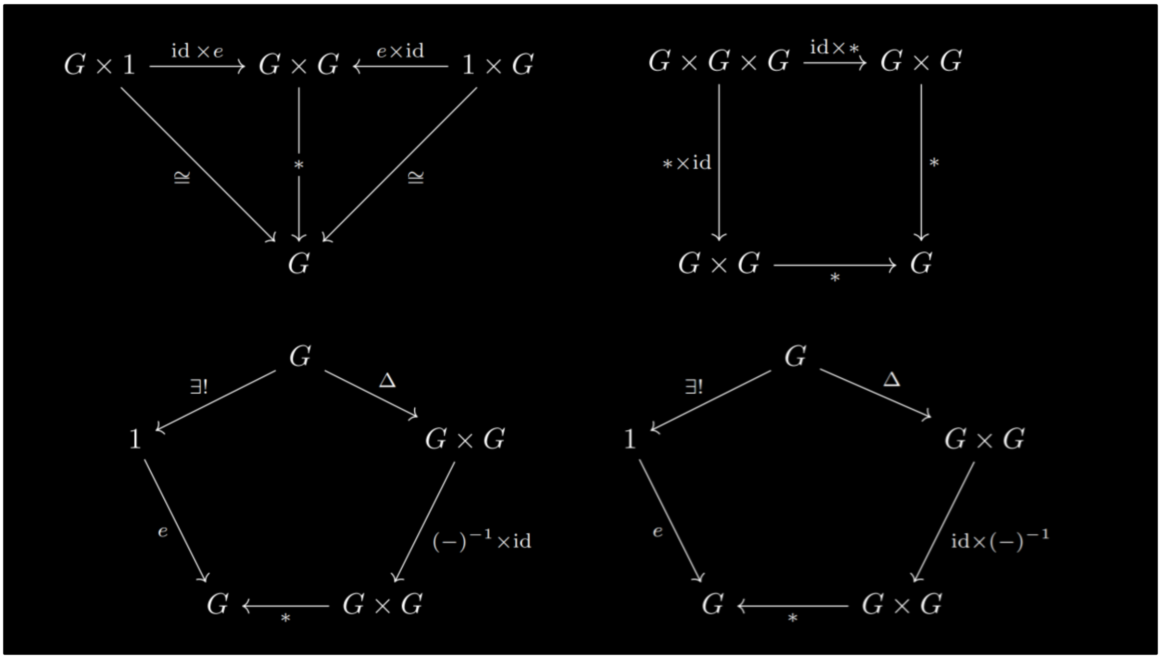
$$* : G \times G \to G$$

$$e : 1 \to G$$

$$(-)^{-1} : G \to G$$

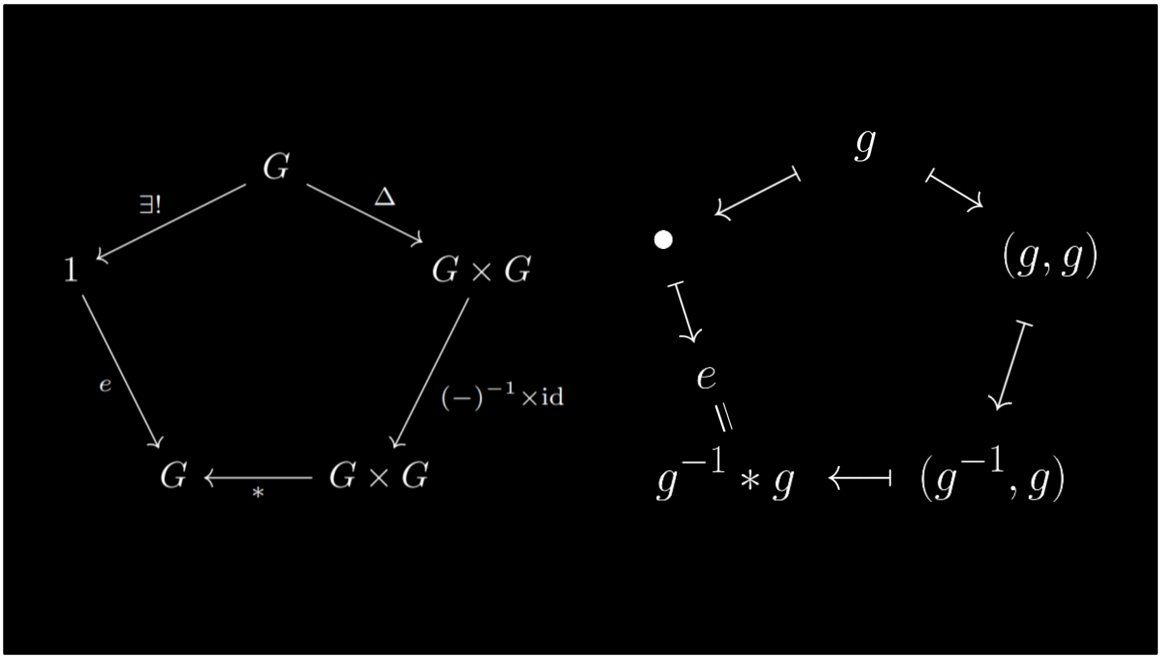
The binary operation is already a function, elements are also just functions from the one-point set, and we have a function that sends an element to its inverse.

Now, because G is a set and these three are functions, the associativity, identity, and inverse axioms can be entirely encoded by the requirement that certain diagrams in **Set** that relate the three functions commute.

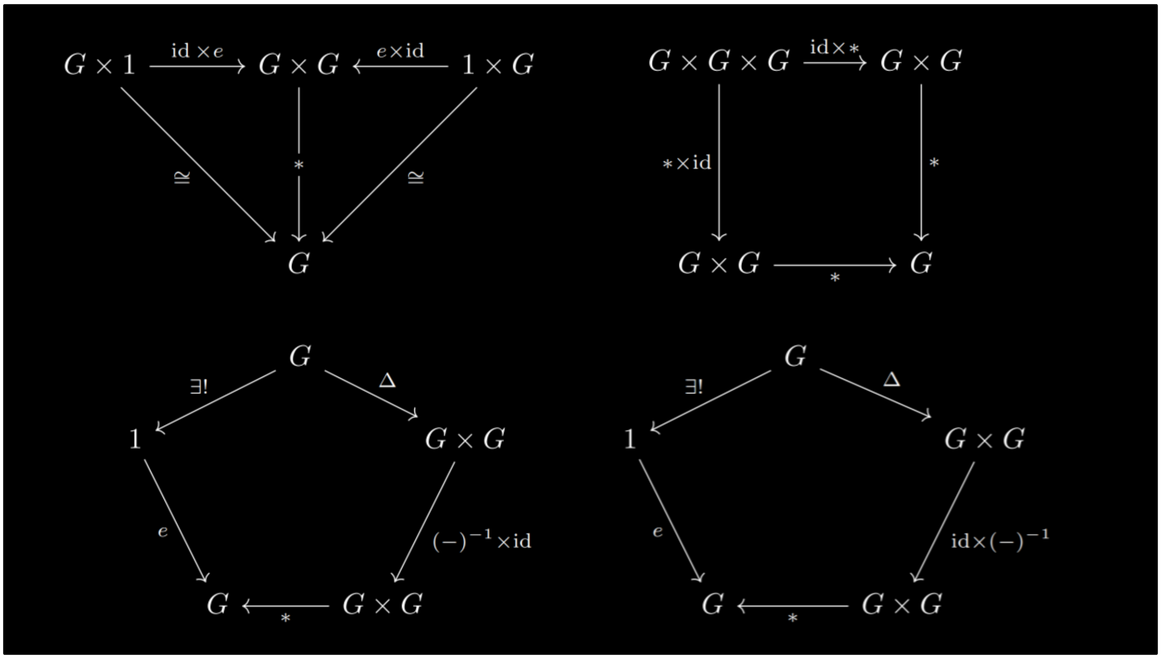


The triangle on the top left encodes the identity axiom, the square on the right encodes associativity, and the two pentagons encode left and right inverses.

To see how this works, let's take a closer look at the left inverse pentagon.



Chasing an element along the left path, we just get the identity element. Along the right path, we are sent along the diagonal morphism, then we invert the left element, before applying the group operation to them to obtain g inverse times g. Commutativity then says that this is equal to the identity.



However, we should notice that this characterisation of groups does not explicitly refer to the elements within the group – all of the requirements are now to do with how this set interacts with these three functions.

Moreover, nothing here is really specific to the category of sets. The only thing we're using here is the existence of terminals and binary products, and indeed, there's no reason why this definition of a group needs to be tied to **Set** at all. All of these diagrams make sense in any arbitrary category that admits these limits, even if the category isn't concrete and we can't meaningfully interpret G to be a set.

An *internal group* in a category $\mathscr{C}$ that admits finite products is an object G equipped with morphisms

$$* : G \times G \to G$$
$$e : 1 \to G$$
$$(-)^{-1} : G \to G$$

such that the previous diagrams all commute.

An *internal group* in a category that admits finite products is an object G, equipped with these three morphisms, such that the previous four diagrams all commute.

$\mathscr{C}$	internal groups
Set	ordinary groups
Top	topological groups
$\mathbf{Man}^{\infty}$	Lie groups
Grp	abelian groups
$\vdots$	$\vdots$

If C is **Set**, then we recover the definition of an ordinary group; if C is the category of topological spaces, we obtain topological groups; if C is the category of smooth manifolds, then we obtain Lie groups; and so on.

This process of abstracting a structure like a group into an arbitrary object is called *internalisation*, and we can do this with many other constructions, creating internal vector spaces, internal lattices, etc.

We can abstract these constructions further and replace the products with tensor products to produce internal objects in general monoidal categories that may not necessarily admit finite products.

For an example, let's look at internal monoids, since they're a simpler than internal groups.

$(\mathscr{C}, \otimes, I)$	internal monoids
$(\mathbf{Mon}, \times, 1)$	commutative monoid
$(\mathbf{Ab}, \otimes_{\mathbb{Z}}, \mathbb{Z})$	(unital) ring
$(R\text{-}\mathbf{Mod}, \otimes_R, R)$	R-algebra
$(\mathbf{Cat}, \times, \mathbb{1})$	strict monoidal category
$\vdots$	$\vdots$

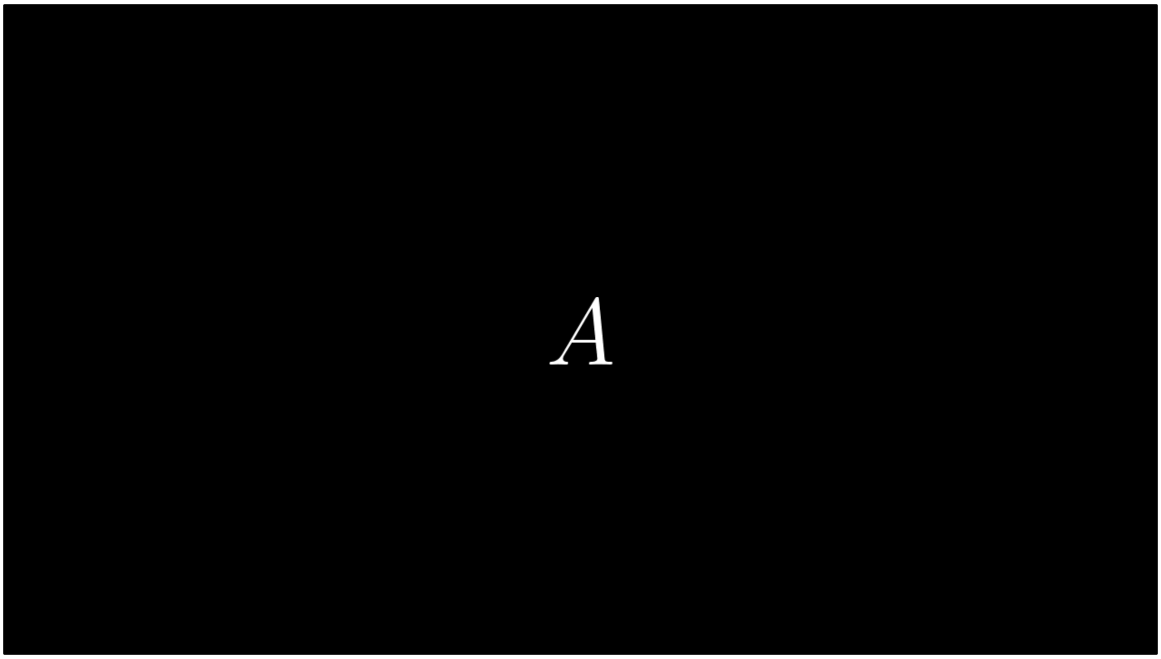
For an example, let's look at internal monoids, since they're a simpler than internal groups.

In the category of abelian groups, with monoidal structure given by the tensor product, an internal monoid is actual an algebraic ring!

Clearly, internalisation is very useful as it unifies various seemingly-distinct constructions, and expressing these objects in this way allows us to see the structural similarities between them.

Internal Homs

We are now interested in the internalisation of a categorical hom-set. First, consider the similar notion of a *function set*.



Given two sets, A, and B, we write this for the set of functions A to B. The problem is that this is a definition rooted in membership, so we need to characterise this set using functions.

$$A, B$$

Given two sets, A, and B, we write this for the set of functions A to B. The problem is that this is a definition rooted in membership, so we need to characterise this set using functions.

$$[A, B]$$

Given two sets, A, and B, we write this for the set of functions A to B. The problem is that this is a definition rooted in membership, so we need to characterise this set using functions.

$$X \to [A, B]$$

So, fix an object X and consider a function from a set X, into the function se.

Such a function takes an argument from X and returns a function A to B. We can alternatively interpret this as a function that takes an argument from both X and A, and returns an element in B.

$$\begin{gathered} X \to [A, B] \\ \cong \\ X \times A \to B \end{gathered}$$

This may be familiar to programmers, because this bijection is the currying and uncurrying operations in computer science that let you transform n-ary functions into a chain of n unary functions.

$$
\begin{array}{c}
X \to [A, B] \\
\cong \\
X \times A \to B
\end{array}
$$

This may be familiar to programmers, because this bijection is the currying and uncurrying operations in computer science that let you transform n-ary functions into a chain of n unary functions.

Let $\mathscr{C}$ be a monoidal category, and let A and B be objects of $\mathscr{C}$.

The *internal hom-object*, or just *internal hom*, of A and B is an object $[A, B]$ such that

$$\hom\big(X, [A, B]\big) \cong \hom(X \times A, B)$$

naturally in X.

So, the *internal hom* of two objects A and B is an object $[A,B]$ such that this isomorphism is natural in X.

If these internal hom objects exist for all A and B, then the category is called *closed monoidal*. If the category is also cartesian monoidal – that is, this tensor product here is given by the categorical product – then we say that the category is cartesian closed.

Let $\mathscr{C}$ be a monoidal category, and let A and B be objects of $\mathscr{C}$.

The *internal hom-object*, or just *internal hom*, of A and B is an object $[A, B]$ such that

$$\hom\big(X, [A, B]\big) \cong \hom(X \otimes A, B)$$

naturally in X.

So, the *internal hom* of two objects A and B is an object $[A,B]$ such that this isomorphism is natural in X.

If these internal hom objects exist for all A and B, then the category is called *closed monoidal*. If the category is also cartesian monoidal – that is, this tensor product here is given by the categorical product – then we say that the category is cartesian closed.

A monoidal category is *closed monoidal* if for every object A, the right tensor by A has a right adjoint:

$$(-) \otimes A \dashv [A, -]$$

so

$$\hom\big(X, [A, B]\big) \cong \hom(X \otimes A, B)$$

naturally in all 3 variables.

If a monoidal category has all internal hom objects, it is called a *closed monoidal category*, which can be expressed more precisely in terms of an adjunction, as written here.

A closed monoidal category that is cartesian monoidal is called *cartesian closed.*

$$\hom\big(X, [A, B]\big) \cong \hom(X \times A, B)$$

If the category is also cartesian monoidal – that is, this tensor product here is given by the categorical product – then we say that the category is cartesian closed.

It turns out that in cartesian closed categories, or more generally in any closed symmetric monoidal category, this isomorphism is actually natural in all 3 variables.

A closed monoidal category that is cartesian monoidal is called *cartesian closed*.

Example. Any locally small category has a set of morphisms between any two objects.

Set is locally small. So, **Set** is cartesian closed.

For an example, consider the category of sets. As we saw earlier, **Set** is cartesian monoidal. Also, in a locally small category, the class of morphisms between any two objects is a set by definition. **Set** is locally small, so every hom-set is is itself an object in **Set**, so **Set** is cartesian closed.

In a cartesian closed category, we write

$$B^A$$

for the internal hom-object

$$[A, B]$$

and call it an *exponential object*.

In a cartesian closed category, we write B to the A for the internal hom of A and B, and we call it an *exponential object*.

Let's write out the defining isomorphism again and check that this definition actually makes sense.

This notation is compatible with the categorical product in that

$$
\begin{aligned}
A^1 &\cong A \\
A^2 &\cong A \times A && 2 := 1 \amalg 1 \\
A^3 &\cong A \times A \times A && 3 := 1 \amalg 1 \amalg 1 \\
&\vdots && \quad\vdots \\
A^n &\cong \prod_n A
\end{aligned}
$$

In a cartesian closed category, we write B to the A for the internal hom of A and B, and we call it an *exponential object*. This notation is compatible with the categorical product in that we have A^2 is isomorphic to A times A, and so on, which follows from the Yoneda lemma.

Let's write out the defining isomorphism of the exponential object again.

$$\hom(X, B^A) \cong \hom(X \times A, B)$$

Given a function f on the right side, its left adjunct under this isomorphism is called the exponential transpose of f, written like this.

$$\hom(X, B^A) \cong \hom(X \times A, B)$$

$$f : X \times A \to B$$

Given a function f on the right side, its left adjunct under this isomorphism is called the exponential transpose of f, written like this.

$$\hom(X, B^A) \cong \hom(X \times A, B)$$

$$f^\flat : X \to B^A \longleftarrow f : X \times A \to B$$

Given a function f on the right side, its left adjunct under this isomorphism is called the exponential transpose of f, written like this.

And conversely, given a function g on the left, its right adjunct under this isomorphism is called the exponential cotranspose, written like this.

$$\hom(X, B^A) \cong \hom(X \times A, B)$$

$$f^\flat : X \to B^A \quad \longmapsfrom \quad f : X \times A \to B$$

$$g : X \to B^A$$

And conversely, given a function g on the left, its right adjunct under this isomorphism is called the exponential cotranspose, written like this.

$$\hom(X, B^A) \cong \hom(X \times A, B)$$

$$f^\flat : X \to B^A \longleftarrow f : X \times A \to B$$

$$g : X \to B^A \longmapsto g^\sharp : X \times A \to B$$

And conversely, given a function g on the left, its right adjunct under this isomorphism is called the exponential cotranspose, written like this.

The musical notation is standard for adjunctions, but we will call them exponential transpositions rather than adjunctions.

$$\hom(X, B^A) \cong \hom(X \times A, B)$$

What happens if X is terminal? Well, 1 times A is isomorphic to A, so, rearranging things a bit, we have this expression.

$$\hom(1, B^A) \cong \hom(1 \times A, B)$$

What happens if X is terminal? Well, 1 times A is isomorphic to A, so, rearranging things a bit, we have this expression.

$$\hom(1, B^A) \cong \hom(1 \times A, B) \cong \hom(A, B)$$

What happens if X is terminal? Well, 1 times A is isomorphic to A, so, rearranging things a bit, we have this expression.

$$\hom(1, B^A) \cong \hom(1 \times A, B) \cong \hom(A, B)$$

What happens if X is terminal? Well, 1 times A is isomorphic to A, so, rearranging things a bit, we have this expression.

$$\hom(1, B^A) \cong \hom(A, B)$$

Elements of an object are just maps from 1 to that object, so this isomorphism is saying that the elements of B^A naturally correspond to morphisms from A to B, so we can view this exponential object as "containing" morphisms from A to B, in a sense, which is exactly what we wanted.

Topoi

The notion of a topos was first introduced in algebraic topology by Grothendieck as a generalisation of sheaves of sets in topology. Every topological space induces a topos, and conversely, every topos, as defined by Grothendieck, behaves in many ways like a generalised topological space.

A more general notion of a topos was soon developed by Lawvere and Tierny, which we now introduce.

A (*elementary*) *topos* is a category that:

- is finitely complete;
- is cartesian closed;
- has a subobject classifier.

An elementary topos is a category that is finitely complete, is cartesian closed, and has a subobject classifier.

This definition seems very... short, for a structure we claim is so important, but a topos carries a vast amount of additional rich structure that just happens to follow from these few axioms.

Lemma. *Every monomorphism in a topos is regular.*

Corollary. *Every topos is balanced.*

Theorem. *Every topos is finitely cocomplete.*

Theorem. *Every morphism factors essentially uniquely through its image into the composition of an epimorphism and a monomorphism.*

For instance, every monomorphism in a topos is regular. That is, it occurs as the equaliser of some pair of parallel morphisms.

From this, it also follows that every topos is balanced, so every bimorphism is iso.

We also have that every topos is finitely cocomplete, despite only starting with finite completeness.

And we also have that every morphism factors into an epimorphism followed by a monomorphism.

A (*elementary*) *topos* is a category that:

- is finitely complete;
- is cartesian closed;
- has a subobject classifier.

The prototypical example of a topos is the category of sets, but **Set** has a couple of special properties it doesn't share with most other topoi, which we will explore soon.

But on the other hand, finite completeness gives terminal objects, which allow us to consider the elements of objects in arbitrary topoi;

cartesian closure means that we have exponential objects, so we can talk about objects of morphisms and power objects;

and the subobject classifier allows us to talk about subobjects.

In this way, an arbitrary topos behaves very much like a generalised category of sets, and we can use a lot of set-theoretic language to describe objects in a topos, even if they aren't sets.

Set

We give some characteristics of **Set** that distinguish it from other topoi, appealing only to “obvious” properties that sets and functions should satisfy.

Set is non-trivial. That is, **Set** $\not\simeq \mathbb{1}$.

Firstly, **Set** is non-trivial. That is, it is not equivalent to the trivial category. Usually, this is expressed by saying that the terminal and initial objects of **Set** both exist and are not isomorphic.

Set is non-trivial. That is, $\mathbf{Set} \not\simeq \mathbb{1}$.

(i) $0 \ncong 1$.

Usually, this is expressed by saying that the terminal and initial objects of **Set** both exist and are not isomorphic.

If $f, g : X \to Y$ are parallel morphisms such that every morphism $x : 1 \to X$ equalises f and g, then $f = g$.

In **Set**, the terminal object 1 also has another special property. If f,g are parallel morphisms from X to Y that are equalised by every map from 1 to X, then $f = g$.

Let's interpret what this is actually saying. Morphisms from 1 to X are just elements of X, so basically, this property is that if two functions X to Y agree on all the elements of X, then they are the same function.

This is a strong extensionality principle for functions, analogous to the axiom of extensionality for sets. This is also saying that functions have no internal structure and are completely defined by their effects on elements.

An object that satisfies this property is called a *separator*, so our second property of **Set** is that

If $f, g : X \to Y$ are parallel morphisms such that every morphism $x : 1 \to X$ equalises f and g, then $f = g$.

(*ii*) The terminal object 1 is a separator.

the terminal object 1 is a separator.

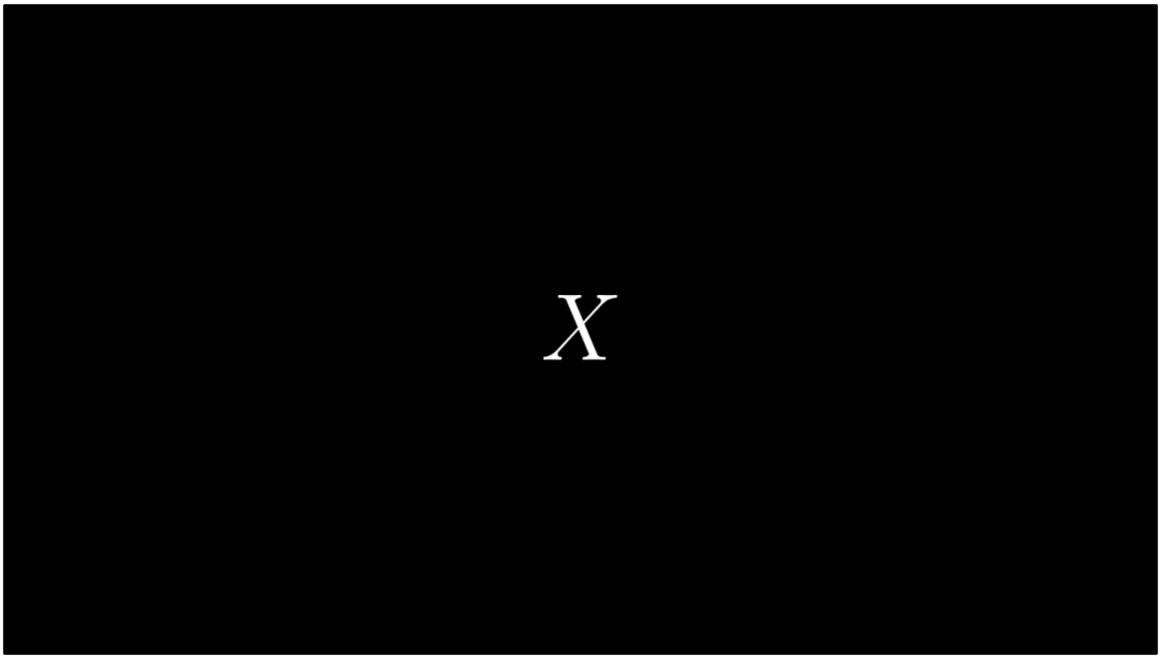
(i) $0 \ncong 1$.

(ii) The terminal object 1 is a separator.

A topos that satisfies (i) and (ii) is called *well-pointed.*

A topos that satisfies these two properties is called *well-pointed.*

The next special property of **Set** is roughly speaking, the existence of the natural numbers.



Given a set X and an element x in X, every function r from X to X generates a unique sequence in X such that the first element of the sequence is x, and each term is obtained by applying r to the previous term.

$$x \in X$$

Given a set X and an element x in X, every function r from X to X generates a unique sequence in X such that the first element of the sequence is x, and each term is obtained by applying r to the previous term.

$$x \in X \qquad r : X \to X$$

Given a set X and an element x in X, every function r from X to X generates a unique sequence in X such that the first element of the sequence is x, and each term is obtained by applying r to the previous term.

$$x \in X \qquad r : X \to X$$

$$(x_i)_{i=0}^{\infty} \subseteq X$$

Given a set X and an element x in X, every function r from X to X generates a unique sequence in X by recursion on x.

$$x \in X \qquad r : X \to X$$

$$(x_i)_{i=0}^{\infty} \subseteq X \qquad \begin{aligned} x_0 &= x \\ x_{n+1} &= r(x_n) \end{aligned}$$

Given a set X and an element x in X, every function r from X to X generates a unique sequence in X by recursion on x.

Notice that a sequence is indexed by the natural numbers, so this yields a correspondence between applications of r to x, and applications of the successor function to 0 in the subscripts.

Moreover, a sequence in X is just a morphism from the natural numbers to X, so this is really a statement about the natural numbers.

$$x \in X \qquad r : X \to X$$

$$(x_i)_{i=0}^{\infty} = f : \mathbb{N} \to X$$

Moreover, a sequence in X is just a morphism from the natural numbers to X, so this is really a statement about the natural numbers.

Writing f for this function, the previous just says that this diagram commutes.

$$
\begin{tikzcd}
& \mathbb{N} \arrow[r, "s"] \arrow[dd, "f"] & \mathbb{N} \arrow[dd, "f"] \\
1 \arrow[ur, "0"] \arrow[dr, "x"'] & & \\
& X \arrow[r, "r"'] & X
\end{tikzcd}
$$

Moreover, a sequence in X is just a morphism from the natural numbers to X, so this is really a statement about the natural numbers.

Writing f for this function, the previous just says that this diagram commutes.

A *natural numbers object*

A *natural numbers object* is a triple $(\mathbb{N}, 0, s)$ consisting of an object $\mathbb{N}$, an element $0 : 1 \to \mathbb{N}$, and a *successor morphism* $s : \mathbb{N} \to \mathbb{N}$ with the universal property that the triple $(\mathbb{N}, 0, s)$ factors through every other triple (X, x, r) uniquely:

$$\begin{tikzcd}
 & \mathbb{N} \arrow[r, "s"] \arrow[d, "f"] & \mathbb{N} \arrow[d, "f"] \\
1 \arrow[ur, "0"] \arrow[dr, "x"'] & & \\
 & X \arrow[r, "r"'] & X
\end{tikzcd}$$

Moreover, a sequence in X is just a generalised element of shape N, or equivalently, a morphism from the natural numbers to X, so this is really a statement about the natural numbers. Writing f for this function, the previous just says that this diagram commutes.

A *natural numbers object*

A *natural numbers object* is a triple $(\mathbb{N}, 0, s)$ consisting of an object $\mathbb{N}$, an element $0 : 1 \to \mathbb{N}$, and a *successor morphism* $s : \mathbb{N} \to \mathbb{N}$ with the universal property that the triple $(\mathbb{N}, 0, s)$ factors through every other triple (X, x, r) uniquely:

$$\begin{tikzcd}
 & \mathbb{N} \arrow[r, "s"] \arrow[dd, dashed, "f"] & \mathbb{N} \arrow[dd, dashed, "f"] \\
1 \arrow[ur, "0"] \arrow[dr, "x"'] & & \\
 & X \arrow[r, "r"'] & X
\end{tikzcd}$$

A *natural numbers object* is a triple, consisting of an object N, an element 0, and a successor morphism s from N to N, with the universal property that it uniquely factors through every other similar triple. This universal property also means that the natural numbers object is essentially unique, so we are safe to speak about *the* natural numbers object.

The sequence f given by the universal property is said to be defined by *simple recursion* with *starting value* x and *transition rule* r. Arithmetic operations such as addition and multiplication can then be defined in terms of their exponential transpositions by simple recursion.

$$\begin{tikzcd} & \mathbb{N} \arrow[r, "s"] \arrow[dd, "+^\flat"] & \mathbb{N} \arrow[dd, "+^\flat"] \\ 1 \arrow[ur, "0"] \arrow[dr, "\mathrm{id}^\flat_{\mathbb{N}}"'] & & \\ & \mathbb{N}^{\mathbb{N}} \arrow[r, "s^{\mathbb{N}}"'] & \mathbb{N}^{\mathbb{N}} \end{tikzcd}$$

For instance, if we define s^N to be the successor function acting pointwise on a sequence, then this defines the exponential transpose of addition by simple recursion.

(*iii*) **Set** has a natural numbers object.

The third distinguishing property of **Set** is then that it has a natural numbers object.

$$f : A \twoheadrightarrow B$$

The last special property of **Set** we will need is that every surjective function has a section.

$$f : A \twoheadrightarrow B \qquad s : B \to A$$

$$f \circ s = \mathrm{id}_A$$

The last special property of **Set** we will need is that every surjective function has a section. This can be stated in categorical terms as:

$$f : A \twoheadrightarrow B \qquad s : B \to A$$

$$f \circ s = \mathrm{id}_A$$

(iv) Epimorphisms split.

Epimorphisms split.

The function s is defined by assigning each element in b an element from its fibre. But this implies the existence of a choice function for any arbitrary f, so this statement is precisely the axiom of choice.

(*i*) $0 \not\cong 1$.

(*ii*) The terminal object 1 is a separator.

(*iii*) There is a natural numbers object.

(*iv*) Epimorphisms split.

Our properties are thus these four:

Sets and set functions form a well-pointed topos with natural numbers object and Choice.

These properties can be stated more concisely as: *Sets and set functions form a well-pointed topos with natural numbers object and Choice.*

The category of sets, of course, has more properties than this. For instance, all power objects exist, the subobject classifier has two elements, the topos is Boolean, etc. but all of these properties all follow from this statement.

The question is then, what conditions or axioms do we need to enforce on sets to ensure that they *do* form such a topos?

Sets and set functions form a well-pointed topos with natural numbers object and Choice.

One answer is of course, ZFC, and indeed any model of ZFC will satisfy these properties.

This is the answer most mathematicians will know in the back of their mind, but often do not like to concern themselves with, because the axioms of ZFC is generally quite far removed from their work. The specific axioms seem to be unimportant, compared to the need for our sets to satisfy this statement.

Sets and set functions form a well-pointed topos with natural numbers object and Choice.

The answer of "ZFC" then seems somewhat unsatisfying, or even irrelevant, because this statement derives from obvious properties of sets that we often use. At no point did we consult with a list of axioms to decide these properties, because they all follow from our informal idea of what sets should be and how they should behave.

Sets and set functions form a well-pointed topos with natural numbers object and Choice.

In particular, anything that satisfies this statement will, for all intents and purposes, behave *like a set*.

And this is the idea behind the *Elementary Theory of the Category of Sets*, or *ETCS*. We take these properties as our axioms.

Now, at this point, one could think that there is some circularity here, that ETCS depends on the notion of a category, which itself depends on the notion of "collections" of objects and morphisms, which seem quite similar to sets.

The straightforward response is that category theory, and specialisations thereof, like ETCS, is a first-order theory. Although ETCS is motivated by category-theoretic ideas, it doesn't intrinsically depend on the notion of a category. For instance, we can state the axioms of ETCS without mentioning categories as follows:

1. Function composition is associative and has identities
2. There exists an empty set
3. There exists a singleton set
4. Functions are completely characterised by their action on elements
5. Given sets X and Y, we may form their cartesian product $X \times Y$
6. Given sets X and Y, we may form the set of functions from X to Y
7. Given a function $f : X \to Y$ and $y \in Y$ we may form the fibre $f^{-1}[y]$
8. The subsets of a set X correspond to the functions $X \to \{0, 1\}$
9. The natural numbers form a set
10. Every surjection admits a section

The comparison with ZFC is now more obvious. ZFC says, “there are some things called sets, there is a binary relation on sets, and some axioms hold.” while ETCS says “there are things called sets, for every pair of sets there are things called functions, and some axioms hold.” The point is, circularity is no more of a problem for ETCS than it is for ZFC.

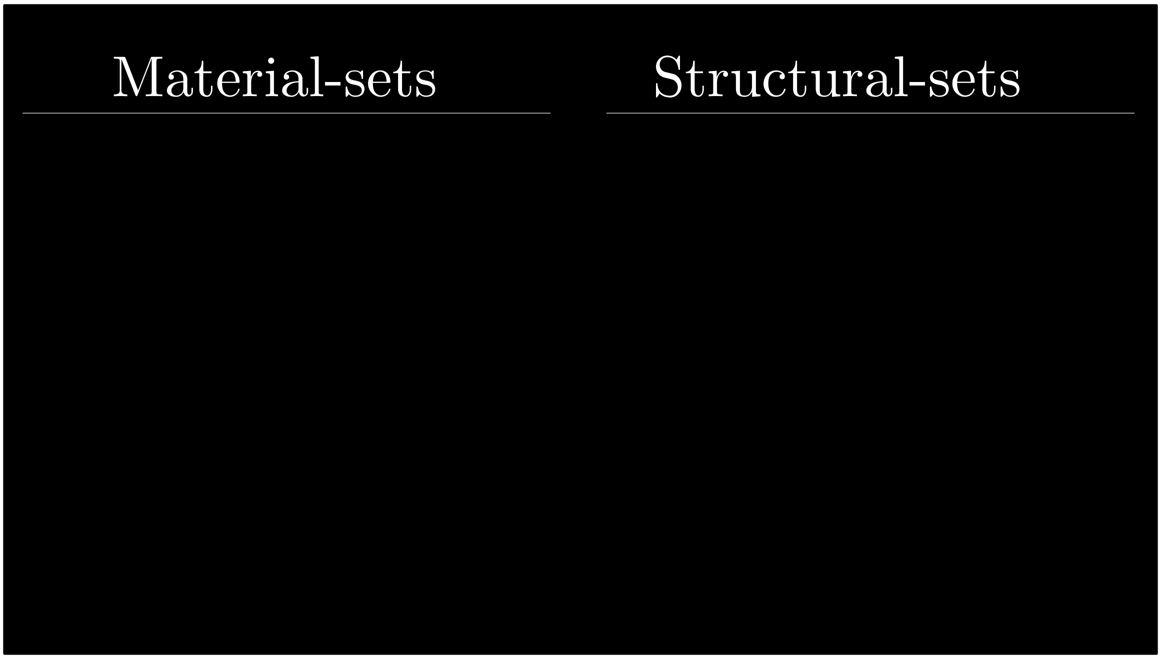
Also, the axioms as stated here seem to be more fundamental in some way than ZFC. Suppose that one day, we find that ZFC had been proven to be inconsistent; that some logician had started with axioms of ZFC and had irrefutably derived a logical contradiction from them.

It’s likely that most mathematicians, being mostly detached from foundations in the first place, would not be deeply bothered by this fact, and would happily continue on, confident that their results hold true in the sense that their negations do not.

In contrast, the axioms of ETCS are modelled on core properties of sets and functions that we often use. A proof that ETCS were inconsistent would be devastating. We would no longer be able to safely assume that function composition is associative, has identities, etc.

Material and Structural Sets

ETCS and ZFC both deal with "sets", but these notions are so distinct that it seems unhelpful to call them by the same name. We will call a set in the style of ZFC a *material-set* and a set in the style of ETCS a *structural-set*.



In ZFC, the axiom of extensionality says that two material-sets are determined entirely by their elements. In ETCS, a weak extensionality principle is given by the Yoneda lemma; structural-sets are characterised only up to isomorphism by their generalised elements.

However, often only care about equality of subsets of some containing context set, in which case, well-pointedness gives a strong extensionality principle that characterises structural-subsets up to equality.

Material-sets	Structural-sets
determined by elements up to equality	determined by generalised elements up to isomorphism, but subsets up to equality

Also, because elements of structural-sets are functions, this means that they themselves are never sets, unlike in ZFC, where elements of material-sets are always themselves material-sets. This is perhaps closer to how we often use sets in ordinary mathematics.

Material-sets	Structural-sets
determined by elements up to equality	determined by generalised elements up to isomorphism, but subsets up to equality
elements are always sets	elements are never sets

Also, in the introduction, we saw an argument that numbers cannot be sets, since numbers do not have set-properties. In this view, the natural numbers are envisioned as elements of an abstract structure, where elements have no properties beyond what is given to them by that structure.

In ZFC, we define the natural numbers to be some particular material-set, with the arithmetic relations constructed on top of the chosen encoding, but this yields unwanted additional structure, like 3 being an element of 17 or not.

In ETCS, the natural numbers object is a

Material-sets	Structural-sets
determined by elements up to equality	determined by generalised elements up to isomorphism, but subsets up to equality
elements are always sets	elements are never sets

In ETCS, the natural numbers object is a structural-set equipped with a specified zero element and a successor function. Arithmetic relations are expressed in terms of these things, so elements of this abstract structure have arithmetic relations between themselves, but no additional properties beyond that.

More generally, every structural-set is precisely an abstract structure in this sense. An element of a set X is a function into X, and it has no internal identity except that it is an element of X, and is distinct from other elements of X.

Material-sets	Structural-sets
determined by elements up to equality	determined by generalised elements up to isomorphism, but subsets up to equality
elements are always sets	elements are never sets
lots of side effects from constructions	abstract structures encapsulate and isolate properties without side effects

Now, for any material-sets X and A, we can ask if A is in X or not.

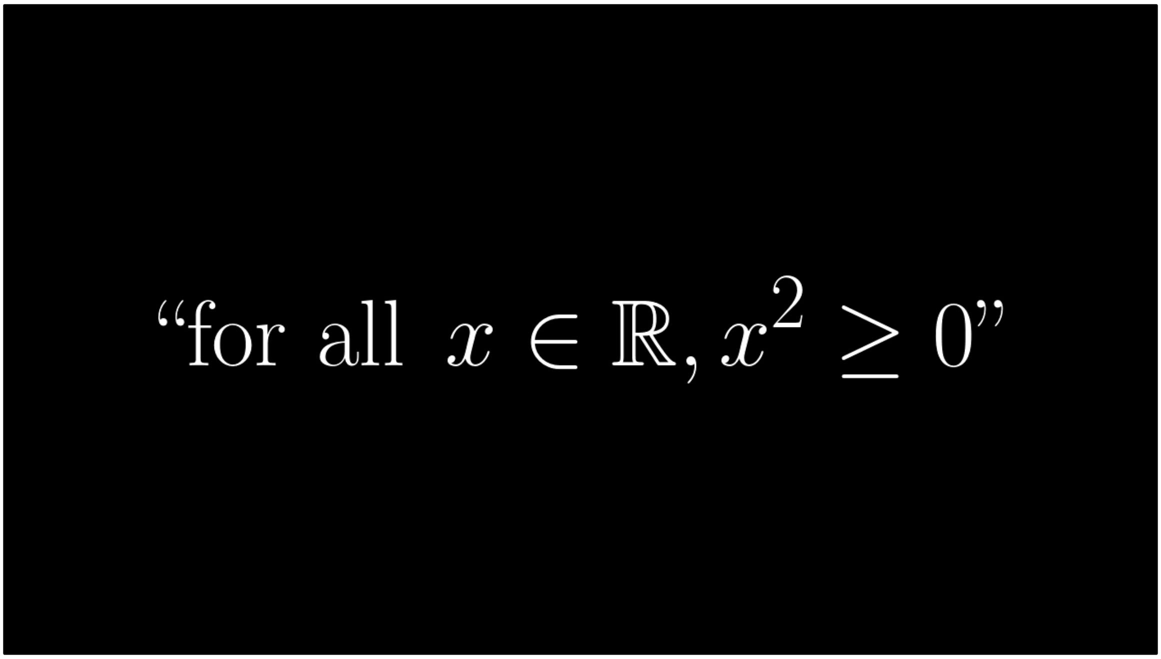
$$A \in X$$

This statement is a proposition in the formal sense; it has a truth value, can be proven, can be combined with logical connectives, quantified over, etc. We'll call this interpretation of membership, *propositional* or *material membership.*

In contrast, if X is a structural set, then there are some things that are *intrinsically* elements of X. Namely, the functions from 1 into X. If a thing is not given as an element of X, then it is *not* an element of X.

Thus, this statement is not something one would ever prove about two pre-existing objects A and X. Consequently, this statement is not a proposition in a structural set theory.

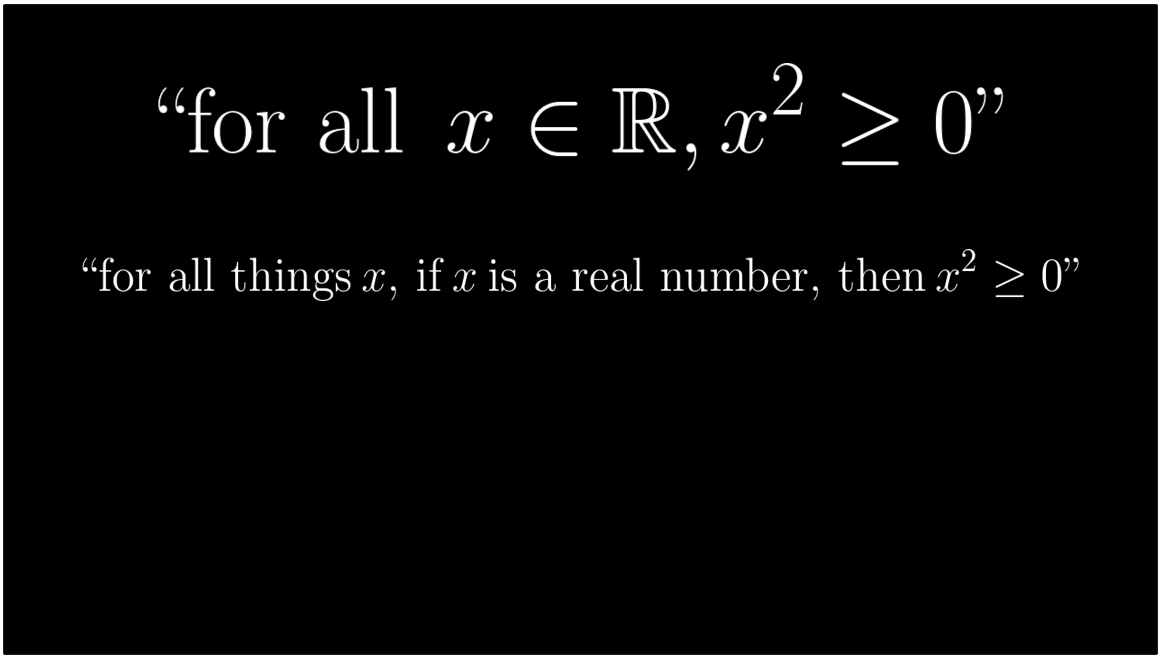
As an illustration of this difference, consider this statement:



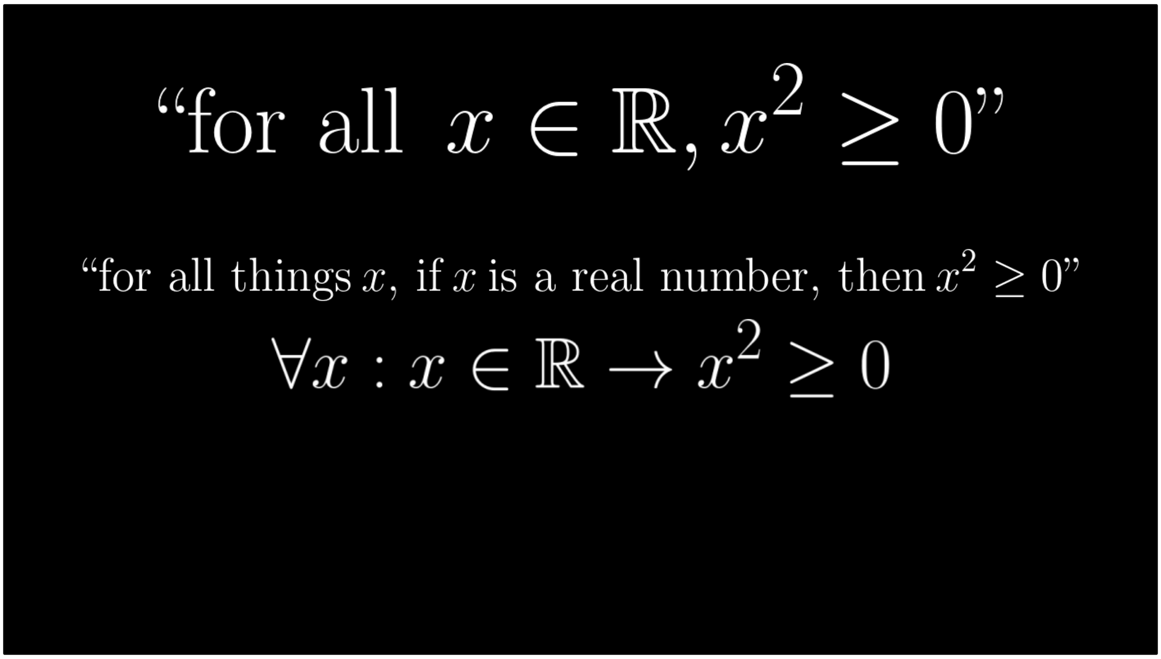
[...]

I am intentionally not reading this out loud yet.

If R is a material-set, then – formally – this would be read as



for all possible things x, if x happens to be a real number, then its square is non-negative. This is an implication, so the formal translation into first order logic is like this:



However, if R is a structural set, then x in R is a logical atom and cannot be the premise of an implication. Thus, the statement should be read as

$$\text{“for all } x \in \mathbb{R}, x^2 \geq 0\text{”}$$

“for all things x, if x is a real number, then $x^2 \geq 0$”

$$\forall x : x \in \mathbb{R} \to x^2 \geq 0$$

“it is a property of all real numbers x that $x^2 \geq 0$”

it is a property of all real numbers that their square is non-negative, and the formal transcription is this:

$$\text{“for all } x \in \mathbb{R}, x^2 \geq 0\text{”}$$

“for all things x, if x is a real number, then $x^2 \geq 0$”

$$\forall x : x \in \mathbb{R} \to x^2 \geq 0$$

“it is a property of all real numbers x that $x^2 \geq 0$”

$$\forall x \in \mathbb{R} : x^2 \geq 0$$

Arguably this is closer to how we use quantification in practice. When we say x in R, we're really declaring what type of variable x is. We generally don't mean, “it is a property of any and all things in mathematics that *if* it happens to be a real number, *then* its square is non-negative.”

For instance, included in the material interpretation is the statement: “if the polynomial ring over the field of rationals happens to be a real number, then its square is non-negative”. This is vacuously true, but I think it could be reasonably agreed that this is a statement that most mathematicians wouldn't naturally regard as part of the content of this proposition.

$$\text{“for all } x \in \mathbb{R}, x^2 \geq 0\text{”}$$

“for all things x, if x is a real number, then $x^2 \geq 0$”

$$\forall x : x \in \mathbb{R} \to x^2 \geq 0$$

“if $\mathbb{Q}[x]$ is a real number, then $\left(\mathbb{Q}[x]\right)^2 \geq 0$”

Arguably this is closer to how we use quantification in practice. When we say x in R, we’re really declaring what type of variable x is. We generally don’t mean, “it is a property of any and all things in mathematics that *if* it happens to be a real number, *then* its square is non-negative.”

For instance, included in the material interpretation is the statement: “if the polynomial ring over the field of rationals happens to be a real number, then its square is non-negative”. This is vacuously true, but I think it could be reasonably agreed that this is a statement that most mathematicians wouldn’t naturally regard as part of the content of this proposition.

Conversely, sometimes we *do* want to regard membership as a proposition. For instance, if L is the set

$$L = \left\{ z \in \mathbb{C} : \Re(z) = \frac{1}{2} \right\}$$

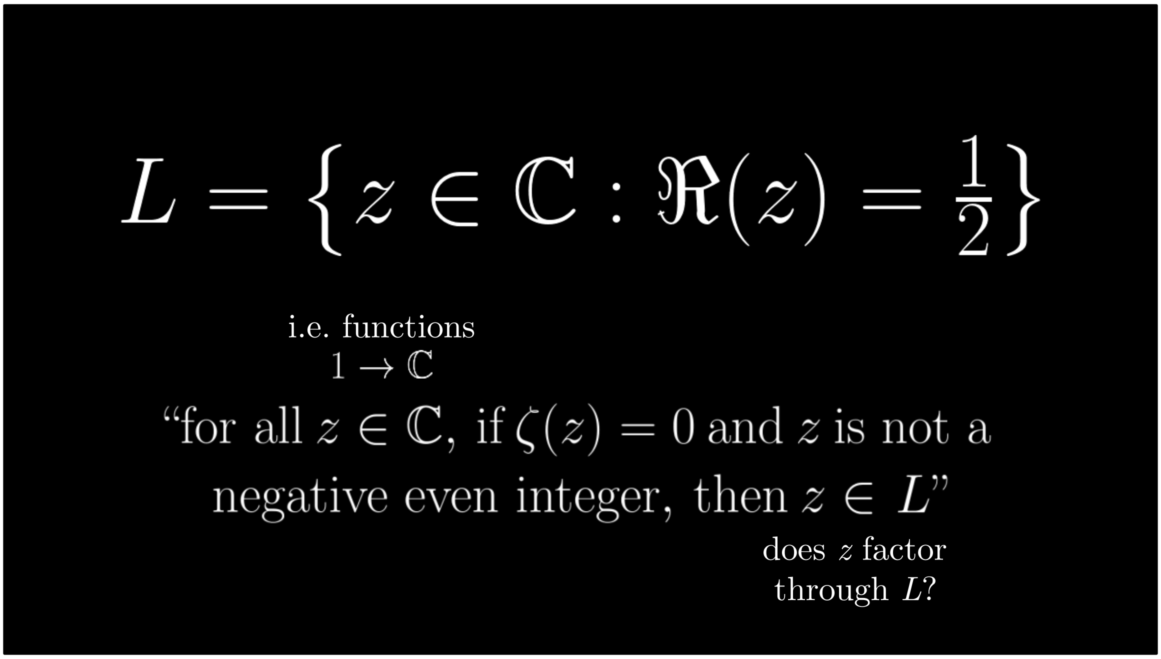
if L is the set of complex numbers with real part one half, then a lot of people are extremely interested in proving this statement:

$$L = \left\{ z \in \mathbb{C} : \Re(z) = \frac{1}{2} \right\}$$

“for all $z \in \mathbb{C}$, if $\zeta(z) = 0$ and z is not a negative even integer, then $z \in L$”

The first `in` symbol here is read as a type declaration – we’re saying that z is of type complex number – while the second is read materially as a proposition.

The observation here is that L is a subset of the complex numbers.



z is already given to be a member of the structural set, i.e., a function 1 to C, so it is possible to ask whether it belongs to this subset L (i.e., factors through L). Function extensionality then characterises subsets up to true equality, so ETCS also supports this use of propositional membership.

Material-sets	Structural-sets
determined by elements up to equality	determined by generalised elements up to isomorphism, but subsets up to equality
elements are always sets	elements are never sets
lots of side effects from constructions	abstract structures encapsulate and isolate properties without side effects
propositional membership only	type-declaration membership; supports propositional patterns in the presence of ambient sets.

Once we have memberships, functions, unions, quotients, etc. in whichever choice of foundations, the following development of mathematics is mostly the same. At a certain point, once we've constructed basic mathematical structures, for all practical purposes, it doesn't matter what foundation we begin with.

After all, asking if 3 is an element of 17 or not isn't really a practical problem in ZFC. However, it's still pedagogically fruitful to ask such questions. One possible advantage of teaching ETCS as a foundation is that it introduces the notions of isomorphism and universal properties earlier on and can also clarify some material constructions.

Beyond this, the significance of ETCS is not from its use or non-use as a foundation, but more so from the research into topos theory that followed. It was one of the first attempts of a categorical analysis of logic, and though it did not see much use as a foundation itself, the more general theory of topoi that it inspired is now the main language of categorical logic.