Abstract Assertions

by Sven Nilsen, 2018

Sometimes it is desirable to communicate ideas that can pass proof boundaries of knowledge. This means the idea is expressed in a sentence without that sentence being an expression in a formal language, but the sentence still has meaning that expands to a broader, still unknown, context. Indeed, such sentences can have a deep meaning and be extremely useful in general. I call such sentences for "abstract assertions". In this paper I discuss some mathematical properties of these sentences to show that they are different from arbitrary sentences. I also give an example of a universal abstract assertion that govern all of mathematics, without being expressible directly within mathematical language itself.

Imagine that you are poor and average intelligent man, but has a rich and eccentric father who told you his fortune will be yours tomorrow. The fortune changes hands only if you deliver a paper describing specifically how it will be used at 12 o'clock, only two hours away from the moment you first get to know this. A large teams of security guards will assure that the fortune is spent exactly how you described, or else you will get arrested and lose control over the fortune. No cheating is allowed.

Feeling bewildered and confused, you start panicking. In one way, the fortune will be yours tomorrow, but in another way, it only belongs to the version of yourself that exists the next two hours.

For example, if your two-hour-self likes mountain trips, you can add a program of visiting all the world's big mountains to the list. However, will your future self desire mountain trips? What happens to your goals when you update your beliefs in hindsight and after the first thousand mountain trips, you figure this will get pretty boring and a waste of your time? In the worst case, you can just stop going on the trips you planned, the money will be taken away, and you get back to poor as normal.

Still, the idea that you could use the money as you would desire in the future is a reality, as long you can figure it out within two hours! You believe you should make an attempt.

Luckily, there is a smart woman living down the street, a logical assumptionist, with a strong reputation for making objective judgements and a deep understanding of what people desire. You run out the door, down the road and knock on her door. She opens, you tell her the situation, and she answers:

"I will help you for a million dollars. If you are unsatisfied with the solution, you will get the money back in the fraction of your remaining expected life time divided by your current remaining expected life time. You can pay me tomorrow."

Of course, considering your father is a billionaire, you agree to her terms. She helps you writing down a list, containing some surprises to you, but on closer look you see that this is stuff you expect to desire more as you grow older. The list is not perfect, has not a tight program, contains a lot of charity donations for satisfying your need for making social contributions, has some pilgrim journeys that are not meant only as vacations for pleasures, and a significant amount of activities that might help you grow deep friendships and relationships with other people of high quality content. Everything is written in a language that is accurate enough to be specific, in order to pass your father's test, but loose enough to allow great enough freedom in practice to not feel overly burdened.

The list is reasonable when you look at it in hindsight, but not things you could figure out in two hours. You deliver the list to your father, who nods approvingly to himself while reading it, and next day you wake up as a rich man. The logical assumptionist gets her million dollars, and you keep living by the program that was set up on the list. Surprisingly, you find out that the choices she helped you with are pretty good. She is not a perfect predictor, so there are occasions where you get bored and wish there could be something else you could do, feeling the plan is a bit too much constrained. However, you are not sure that it would be actually better to quit the program in hindsight, so you continue living as a rich person straight up to the point where you pass the expected life time and owe her zero dollars.

This story is designed to introduce background context for two concepts:

1. The concept of a proof boundary
2. The concept of an abstract assertion

Some times and some places it is necessary to make choices long time before the consequences are visible. Even if you could always become smarter and more clever given enough time and resources, it might simply not be enough time and resources! So, things you can prove, things that you can reach knowledge about within that slice of spacetime, is limited. It is like imagining a high dimensional surface surrounding the possible proofs that is constrained when the inference algorithm is “running out of gas”. Proof boundaries are very real, because at any given moment there is only such and such amount of things we can do. We do not have unlimited choices. Worse, the proof boundary changes when we make certain actions in a particular order. For example, if we do A before B, we might get to C, but not if we do B before A etc. If you close a door you have to open it again to go through it. The same laws for navigating physical spaces is valid for thoughts inside your brain, because thoughts are just as real particles as the ones making up the body that moves around!

An abstract assertion is a technique to work around proof boundaries. This might sound impossible at first, but actually it is not.

Think of all possible moves you can do within some limited time, where each step is the best you could do given the knowledge of the previous step, and an abstract assertion can beat that! How does that NOT sound completely impossible!!?? The key here is in the phrase “you could do”.

Instead of formulating a sentence that can be used for reasoning in a formal language, you are formulating a sentence in an abstract way, and then delivering to somebody else that makes a better use of that knowledge than you are able to do yourself.

In the example of the fortune, you simply go and ask somebody smarter than yourself to help you out. That person does it because she knows that if you help you out, she will get payed. She even suggests a way such that you both can agree on the deal! It is actually much easier to have smart people staying around yourself and just be able to speak “please help me” than it is to try figuring out hard problems with long lasting consequences on your own. Better yet, there is no downside! You pay them with the money that they help you to earn, so in a way, the help is completely free.

It might sound like “please help me” is something that an automatic theorem prover would accept as input. Yes, that would actually work, given the theorem prover is sufficiently advanced. However, it would not work the way we usually think about theorem proving. The sentence “please help me” is

something easily understandable for humans. You can use it basically wherever people understand english. It is not understandable for e.g. a SAT solver or a first order logic theorem prover.
A formal language must be programmable using building blocks that can be agreed on universally. When you do this, the sentences in that language can be accepted by theorem provers, who can be constructed universally, by transmitting a relative small program from one place to another. With other words, we are talking about relative small proof boundaries.

The sentence “please help me” is not something you can use within a small proof boundary. If you were imprisoned and isolated, the sentence “please help me” would not work. However, if you can transmit the sentence to somebody outside, that are also capable of freeing you, it would be extremely useful.

One key insight about abstract assertions, is that you first realize that proof boundaries are very real. Next, you learn that proof boundaries often have holes. An abstract assertion is something that can pass through one of these holes. An algorithm who only uses formal sentences to put thoughts into words, can not take advantage of the holes, because it is constructively limited to the proof boundary!

Therefore, the very definition of an abstract assertion, is that is NOT expressible as a sentence in a formal language! If it was a formal sentence, then it would be useful within the proof boundary, but then it would not possible to use it to beat the proof boundary, because the boundary would surround the result that follows from the use of that sentence.

Can you define what “please help me” mean? No. This is because it uses a language that is too complex to describe, that depends in a complex way on the surroundings you live inside. You are only able to define sentences that are relatively easy to express in a formal language. Only then the idea can be communicated to someplace else with completely different norms and context.

Still, the sentence “please help me” can be translated to many human languages and be used basically wherever you are on earth. Only because it is not definable in a universal sense, does not mean that the sentence is useless or arbitrary. This tells us something about abstract assertions: They can be generally useful even if they do not originate from the use of formal languages.

The question is: Since abstract assertions can be generally useful, are there any abstract assertions that are universally useful?

A such sentence is not expressible in any formal language, that is not possible to understand within the theory constructed by universal building blocks, but it will be perceived as universally useful for all beings of some level of complexity. While this might sound impossible, it is not. Here is an example:

Identity of two objects means that everything that can be said about them is the same.

This sentence is identical to the single axiom of path semantics, and used to create mathematic concepts! Since mathematics is universally useful, so is this sentence.

Now, can I prove informally that this sentence can not be described within any formal mathematical language? Yes, I can!

Most mathematical ideas are based on either set theory or constructive mathematics.

In set theory, you can not create a set which contains all but only sets that do not contain themselves. This is true because if the set contains itself, it should not contain itself, but if it does not contain itself, it should contain itself. This is known as Russell's paradox. Yet, you can speak about a such set, so you can refer to it, without that description being a formal set object. It can be referred to as an impossible set. It is possible to express this object in the language of set theory, but it is impossible to have that set being an object.

Likewise, it should be possible to do the same about sentences themselves in formal languages. A such sentence describes how sentences are described and then excluding itself from it. For example, "the shortest sentence that is not describable within the language of set theory". A such sentence, if it could be described in set theory, would not be that sentence, but it seems perfectly describable within natural language! Therefore, the sentence described by "the shortest sentence that is not describable within the language of set theory" is not describable within the language of set theory, but since something is said about something, it is included in everything describable by "everything that is describable about something", even if what we mean is that a such thing is impossible.

Therefore, when we say "identity of two objects means that everything that can be said about them is the same", we are not talking about sets, but about something larger, indescribable within set theory, that might be an impossible things by itself, but even if something was possible, it would be included! With other words, the meaning of the sentence takes on the greatest universality that is possible.

In the same way, one can say "the shortest sentence that is not possible to construct within the language of constructive mathematics". The meaning of this is quite harder than set theory. One can write a program that outputs "the shortest sentence that is not possible to construct within the language of constructive mathematics" and thereby proving that this sentence can be constructed. However, the sentence does not mean itself. It points to another sentence, so you would have to construct that other sentence in order to disprove it. With other words, if we knew that sentence, we could construct it, but then it would not be that sentence, therefore we do not know that sentence. It does not mean that the other sentence is necessarily not in the set of arbitrary sentences. It means that we do not know which sentence in particular, because if we could prove that there was no such sentence in general, then we would not work within the language of constructive mathematics, because we must give a proof the sentence that can not be constructed, which again means the sentence must be constructible. Therefore, a such proof can not exist, which is perfectly expressible in natural language, but not possible to express within the language of constructive mathematics, because the proof is non-expressible within.

For every branch of mathematics that takes various approaches, if the approach is X, one can just use "the shortest sentence that is not possible with X" and still talk about it. Since this is something that can be said, it is part of the definition that we use in natural language about identity.

Therefore, the sentence "identity of two objects means that everything that can be said about them is the same" is an abstract assertion, because it not expressible in any formal language, but it is universally useful because you can translate it into any analogue definition. It can be narrowly interpreted, e.g. for atomic functions and used to demonstrate computation and type theory. It allows us to intuitively understand Russell's paradox. It is in the spirit of all mathematics, appearing with many faces and shapes, but it does not mean any of those in particular. Just like asking a smart person for help with promise of payment later, one can assume all consequences of this idea, as if it was "borrowed confidence" from the future belief in the idea, and thereby it passes every proof boundary.