Background and context for my 2017 papers on underdetermination

Arne Hole, February 1st 2017

The story of this project began around the year 2000, when I read Paul Ernest's book *Social Constructivism as a Philosophy of Mathematics*. While I found the chapters on social constructivism very interesting, I disagreed with some of the conclusions drawn in the first chapter of the book. This chapter is entitled *A Critique of Absolutism in the Philosophy of Mathematics*. For me, coming from a background in pure mathematics, the "relativism" argued for here came through as unjustified and based on misunderstandings. I discussed this with Paul, who visited Oslo at the time, and I argued that while mathematical objects admittedly have a degree of "relativity" attached to them, due to the existence of non-standard models, our inability to describe the concept "finite" precisely etc., on the other hand the modern mathematical proof standard has a sense of absoluteness in it. A modern proof corresponds, I maintained, to something which can be formalized and checked by a computer, and there is no real element of subjectivity in it. I remember I compared this to famous chess matches, such as Fischer-Spassky in Reykjavik. Some modern mathematical proofs have been checked by so many people that the probability of someone someday discovering an error in them, is on the level of the probability that someone will someday discover an illegal move in one of the Fischer-Spassky chess games. It simply will not happen. The fact that there are no illegal moves in these chess games, is an absolute truth, I argued. As I remember it, Paul replied that he understood my point, but that he thought I was stretching my "absolute" aspect of mathematics into a very thin thread.

After this, I read more on the philosophy of science and mathematics, and not surprisingly I found numerous discourses across different fields which I labeled as "misunderstood relativism". While I knew from before that many influential modern philosophers were rather relativist in their thinking, I did not know to what extent some people in philosophy had tried implementing relativist ideas even to "exact" subjects such as mathematics and physics. As a reaction, I wrote a draft for a little book called "The absolute and the relative - mathematics and physics in the 20th century". In my typical style, I did not write for any particular audience or with publication in mind, I mostly wanted to clarify for myself the distinction between "the relative" and "the absolute" in these subjects. So I finished the book draft and was quite happy with it for a while.

But then, after a couple of years, came a surprising turnaround. I suddenly realized that the very backbone in my view on the philosophy on mathematics as expressed in this book, namely the division line between the absolute (checkability of modern proofs) and the relative (mathematical objects), was actually not well defined. What I suddenly saw, was that due to the simple fact that the admissible length of a proof is supposed to finite, the relative concept of finiteness creeps in at the "absolute" side of my division line as well! In a way, this simple observation showed me that there is circularity in the definition of truth in
the structure of natural numbers. The structure itself determines the admissible lengths of proofs one can use to infer knowledge about it. Thus sits on both sides of the table.

My initial gut reaction was that this must imply inconsistency of theories such as ZF; theories which prove that any sentence in the language L(PA) of Peano Arithmetic is true or false in N. However, after a period of failed ZF inconsistency proof attempts, I came to the conclusion that it does not. Although the finiteness condition on proofs breaks the formal symmetry between N and non-standard models of number theory, there is no reason to believe that it will not be compatible with what can be proved about it formally. The problem is the other way round: I found that when I was thinking about the concept of a proof intuitively in the traditional way, I tacitly assumed that we have a degree of control over this concept which formally speaking, we actually do not have. As a result, it may not be possible to infer as much "knowledge" about infinite structures such as N as we may consistently assume is possible in a formal theory such as ZF. Thus we may be dealing with a massive underdetermination of truth in all infinite mathematical structures. In particular, one may expect there to be an infinite "sea" of statements in languages such as L(PA) whose truth values in N cannot be decided, not even in principle.

So then my search for a mathematical statement without a constructible truth value started. Such statements could naturally be called absolutely undecidable. From the background thinking described above, it was obvious to me that considering sentences of L(PA) and truth values in N, should be sufficient. The underdetermination effect should be fully visible already in that case; passing to a more powerful language should not be necessary. But things were moving very slowly. It was not until 2005 that I got an idea which I thought, for a while, would work. This idea was to use Gödel numbering for constructing an infinite chain of L(PA) statements such that the first statement says something about the second, the second says something about the third, and so on. In this way one could, I thought, make the truth value of the first statement escape through an infinite regression. The concrete construction I came up with was such that the first sentence stated consistence of PA extended with the second sentence. The second sentence then stated consistence of PA extended with the third, and so on. I was so happy with this that I actually submitted a little paper describing it to a journal. However, the competent referee was able to readily decide the truth value of my sentences, which really surprised me. (I was not a logician, I approached this from a background in other fields of mathematics.) The editor was very patient, and I think there were at least three exchanges back and forth with more and more complicated constructions on my part, constructions whose truth values were all decided by the helpful referee, until I realized that the whole infinite regression idea, at least in the form I had tried it so far, probably would not work.

In 2009 I got another, much simpler idea which made me lose interest in my infinite regression construction attempts altogether. This marked the first of the two main breakthroughs in the project. For me, my new approach settled the question about absolute
undecidability in mathematics once and for all. It completely changed my view on a lot of things, because now I also saw the relevance for physics. It was a total revolution, I never thought I would experience anything like that. My new candidates for absolutely undecidable sentences of L(PA) were very simple in spirit. They were based on ordering the set of L(PA)-theories on the form PA+A, where A is a sentence of L(PA), by the Gödel numbers of A. I then considered sentences of L(PA) expressing statements similar to this format: There is an infinite set of integers n such that if PA+A is the nth consistent theory in the ordering, then <property p> holds for the Gödel number of A. Property p could be, for instance, existence of a sequence of zeros of some length in the base n expansion of the Gödel number in question. To use the terminology of my 2017 papers, I then argued that no reachable theory could possibly decide these sentences. The reasoning was that by Gödel's incompleteness theorem, the given reachable theory can identify the ordering number n only for a finite number of the consistent theories in the ordering. To me, given the background thinking I had been through, this was obvious. However, referees were not convinced. The reactions I got frustrated me, because I felt that now, unlike before, the problem was that people simply did not understand.

As a result, I started looking in different directions for attempting to sharpen my examples. For a period, I tried using truth predicates to actually derive contradictions, formally in theories such as ZF, based on my ideas. The results of this were only embarrassing technical errors on my part, once again pointed out by referees. I have really made a fool of myself on numerous occasions when working on this project. Finally, I understood that the existence of truth predicates do not in any way alter the basic underlying conclusion: When discussing absolute undecidability, we are dealing only with underdetermination, not overdetermination.

The second main breakthrough of the project came around 2014, when I discovered the relevance of infinite convergent products to the construction of absolutely undecidable sentences. This gave me much sharper results than I had been able to derive previously. Generalizing my first 2014 example and expressing it in the terminology of my 2017 paper, this example roughly corresponds to letting the NIN of each formula A in $E^+$ be the Gödel number of the shortest PA-proof of inconsistency of A if such a proof exists, and 0 otherwise. Then contribution sequences may be defined as before if the NIN is different from zero. If the NIN is zero, we may let all contribution sequences consist of zeros only. Then we let the nth group of bits in $\tau$ be the length n contribution sequence of the nth inconsistent theory in $E^+$. With these stipulations, the main theorem of my 2017 paper goes through as stated. In other words, for a reachable theory T to prove D in this case, either (i) T must show the existence of an infinite number of winners with respect to $\tau$, or else (ii) there is a natural number n such that T proves that if it is inconsistent, then there is a zero in its length n contribution sequence, which is based on its own (nonexistent!) inconsistency proof.
However, I did not formulate things this way in 2014. Also, I considered theories in L(PA) only, stating without explanation that the results could be generalized to other languages. Maybe that is part of the reason why once again, when I wrote a little paper describing these results, I got a response which to me seemed like being based on lack of understanding.

This brings us to my 2017 paper and its companion draft paper on underdetermination in physics. Once again, I have hopes that people will agree with me, or to put it as it looks from my point of view: finally understand my point. However, taking into account my complete failure in trying to communicate this over the past six years, I realize that it may still take a long time before any widespread acceptance is reached. I think part of the reason why many logicians may instinctively close their eyes on this, is because they consider the idea of being convinced by an argument to mean the same thing as being convinced that the argument is formalizable. Which is something such underdetermination arguments by nature cannot be. As I have written, the existence of convincing arguments of this sort shows that mathematics is, after all, not exclusively about finding reasonable axioms and establishing logical consequences of such axioms. This contradicts what has been the standard view for more than a century.

The main result in my 2017 paper on underdetermination of truth in mathematics may be generalized in a number of directions. For instance, it is obviously not necessary to use an infinite bit sequence divided into groups with \( n \) bits in the \( n \)th group. More generally, it is not necessary to formulate things in terms of infinite bit sequences. However, I presently do not know about any way in which results of similar strength can be obtained without involving, directly or indirectly, an infinite product of real numbers in the interval \((0,1)\) converging to a number in this interval. The fact that such a product in a way corresponds to a probability experiment with an infinite number of independent steps, and yet with two different possible outcomes at the end of the infinite process, in some way seems to be related to the undecidability of the formulas \( D \) which one may construct. At least this is how it looks for me, at the moment.