On The Limits of Deducibility

Stefan Weckbach
Essay for the 2019/20 FQXi Contest "Undecidability, Uncomputability, and Unpredictability" [1]
March 2020

Introduction

WHEN thinking without some caution about the limits of decidability, provability, computability, predictability and how these limits could be linked to each other, one can easily find oneself being trapped in the midst of some increasingly strange waters. To avoid this during our examination of the above limits, we at first have to examine another interesting limit.

The latter is very often overlooked, but is the key for at all being able to put all the above mentioned links into a broader perspective. What we are referring to here as *the key* is the *limit of* deducability *itself.* With having said that, we already know what the underlying reasons are for the limits of deciding, proving, computing and predicting something: the latter are the whole lot instances of the limits for gaining certain fundamental information by just an act of deductive reasoning.

In this essay we will exemplify this claim by not only examining it in detail, but moreover we will find out that this limit leads us to the *induction* that there has to exist something meaningful beyond it. In this sense, the limits mentioned at the beginning are no problem at all, but surprisingly part of the solution. Let's therefore start our overall analysis by introducing *Kurt Gödel*, an Austro-Hungarian logician, mathematician and analytic philosopher. He was born in 1906 and already at the early age of twenty-five he had published two world-renowned *incompleteness theorems* [1].

1 Unprovability

Gödel's *first incompleteness theorem* states that no *finitely specified* and *consistent* axiomatic system that at least contains some elementary arithmetic ('Robinson arithmetic') is able to prove all true statements about the natural numbers from within itself. Naturally we might think that there is nothing more natural in mathematics than the natural numbers and thus, all that can reasonably be stated with these nice numbers must provably be true. But Gödel demonstrated that those systems that fall under the restrictions of his theorems cannot prove true all true statements that can be constructed by those systems. Gödel used a statement of the kind 0≠1 which is obviously true, but is neither provable nor disprovable within the mentioned mathematical systems.

---

[1] Abstract available at https://fqxi.org/community/forum/category/31427

If the statement 0≠1 would be *dis*provable by such a system, that system would obviously be inconsistent. Therefore one assumes that these systems are consistent, but incomplete. Hence, they lack certain axioms that would enable them to decide the truth values for all their valid statements. The question arises which axioms are needed to facilitate a mathematical system that is consistent as well as complete. There indeed exist some mathematical systems that have these properties, but unfortunately the ones we need to do science with are of the incomplete kind.

Extending those systems that fall under the restrictions of Gödel's theorems with some additional axioms to make them definitely complete isn't possible for any deterministic machine or any human mind, since it would require an infinite amount of time and additional axioms to do so (if at all logically possible). But there is another obstacle to it: obviously these systems aren't able to differentiate between a *possibility* and a *necessity* from within themselves, because there is no mechanical or logical procedure available that could tell us *explicitly* whether or not systems that fall under the restrictions of Gödel's findings are inconsistent or incomplete. Therefore Gödel's *second incompleteness theorem* states that these systems, if one assumes that they are fundamentally consistent, cannot demonstrate their own consistency, but also cannot demonstrate their own inconsistency.

All this means that neither machines nor human beings can reliably *deduce* that certain things do exist or do not exist. Because for knowing this they had at first to know whether or not the existence of these things should *necessarily* be considered a possibility, a necessity or even an impossibility. This may be a trivial statement from the point of view of human observers, but as we will see, it cannot be considered as also being trivially recognizable by a computer program.

Despite our problems to decide between the above mentioned modalities, we nonetheless obviously can know a truth that mathematical systems or machines presumably never can know. We know that machines and thus, their algorithms, lack the needed ontological awareness of the terms 'possible', 'necessary' and 'impossible' for at all being able to come to the conclusion that these criteria are profound limits of their own abilities to *deduce* something to be *ontologically true or false*.

Therefore, deducability *for human beings* cannot be defined as an exclusively only mechanical, algorithmic process, because it strongly necessitates the concept of ontological truth. That this is so was already suggested by the Polish-American mathematician *Alfred Tarski.* In 1935 he gave a clever proof that the concept of truth can neither be definitely defined nor be implemented into any sufficiently strong formal (computer) system that falls under the restrictions of Gödel's theorems (finite and consistent) [2]. Since logics, amongst other things, needs this concept to at least make the existential statement “i think and therefore i am”, every machine with similar abilities must also have some knowledge about the *necessity* to differentiate between certain things that aren't existent and certain things that are existent. Consequently, this knowledge about existential issues needs the concept of ontological truth to be at all *knowledge*.

2 Unsolvability

To further demonstrate the necessity for an ontological concept of truth to deduce that something is *non-formalizable*, we want to examine *Alan Turing*'s famous *halting problem*. Turing envisioned a computer program (a specialized algorithm) called 'halting program' that would be able to beforehand determine whether or not any other possible algorithm (for example a mathematical equation) stops (has a solution) or potentially must run forever (has no solution). If such an other algorithm does not have a valid solution for a given mathematical problem, the halting program should indicate this by outputting the number 0 and stop. If there is a solution, the halting program should output the number 1 and stop.

What is demanded by such a 'halting program' is nothing less than to know beforehand all answers to questions about the existence of mathematical solutions. Alone this sounds impossible, since the set of all these questions could easily be infinite (who knows?), and hence the halting program had to have an infinite size. But there is another obstacle to it: if that halting program does not find a shortcut for determining that a certain mathematical problem has *no* solution (via a proof by contradiction), the only alternative left would be to run the equation for an *unknown* time. If there would indeed be no solution existent, the program would therefore run theoretically forever. But this *contradicts* the very definition of the halting program's internal algorithm to be at all feasible, namely that it has to stop in every case and output either a 0 or a 1.

This contradiction shows that such a halting program – any algorithmic program – cannot differentiate between an impossibility and a necessity – and therefore it is simply impossible to be at all existent. If it could nonetheless differentiate between an impossibility and a necessity, it would have realized that its own source code is simply contradictory: it demands to halt in both cases whereas one case *might* demand it to run forever and for that case the program *either* would go into an infinite loop *or* had to have an infinite complexity. Unless such an algorithm has no understanding about the impossibility for both alternatives to ever stop at a certain point, it cannot grasp the real meaning of what has been famously termed the *unsolvability of the halting problem* [3].

As already mentioned, the halting program only would have a 'solution' to the halting problem if all mathematical questions would have been already answered and all the answers had somewhat been implemented into it. But then such a halting program would be superfluous, since the truths that had been implemented into it hadn't been deduced by itself nor could they any more be 'predicted' by it as originally intended. Again, what we determine here is that neither machines nor human beings can *in all cases* reliably beforehand know that certain things do exist or do not exist.

3 Conscious algorithms

The above passages indicate what an algorithm must be capable of when it should be at all able to realize that the halting problem is unsolvable. It at first had to *question its own deductive system* by realizing that there are things that never can be deduced for sure. The latter was the case for beforehand determining for all mathematical questions whether or not they have a solution. Since there does not exist a solution to that halting problem, that 'conscious' algorithm also had to realize that there are things in an assumed-to-be deterministic universe that do not exist.

Note also that a 'conscious' algorithm theoretically could fabricate the same impossibility proof that Turing facilitated with the help of the diagonalization method. But this would effort that this algorithm must at first have a concept about what we call 'infinity'. To accomplish the latter, the algorithm had to realize what algorithms are all about, namely deterministic, mechanical procedures that theoretically are able to enumerate every natural number. Consequently it also had to realize the difference between *theoretical* and *practical* since enumerating all possible natural numbers is practically an impossibility because there would be an infinity of them to count. It follows that for differentiating theoretical from practical, the algorithm also had to have a solid concept of what we call 'ontology'. Consequently, for any algorithm that should be considered as being conscious in a real ontological sense, it necessarily had to have the ability to differentiate between epistemological truths and ontological truths. Otherwise this algorithm wouldn't even be able to realize that “i think and therefore i am” is true. In the latter case it would not even know that it is not conscious. This then would be equivalent to void program-routines that not even can know that they lack the necessary criteria to decide all those halting questions, since these criteria simply do not exist in a world that does not contain all existent information in every part of it.

What we derived with all this is an important difference between an algorithm and human intelligence. The latter can not only trace if-then deductions, but also ask *why*-questions. For assuming that an algorithm could also ask why-questions, much less demand explanations for something, one had to assume that such an algorithm could at first identify and then question some crucial details of its own deductive processes – without ending in an infinite loop of questioning *every such detail*.

4 Impossible deductions

Note that *truth* defines itself as being existential like for example “i think and therefore i am”. If machines would have consciousness, they necessarily had to grasp that truth about themselves. Moreover, they also had to grasp that there is something beyond themselves and that this fact is also a truth. Since both truths are a-priori facts to at all start a deduction about something, at least the existential truth that existence is *a-priori* possible must be considered as being *irreducible*. Thus, at the ends of the spectrum of reducible phenomena there obviously reside some kinds of 'loose ends' that cannot reasonably be logically reduced to something other, because otherwise the reducibility of ultimate reality would be infinite in all directions and all logical reasoning had to break down.[2]

We are now able to formulate some general limits of deducability that hold for human intelligence as well as for any mechanistic, deductive procedure. For both it is *impossible*

1. to know a-priori the complete set of things that are **impossible** to exist
2. to know a-priori the complete set of things that are **possible** to exist
3. to know a-priori the complete set of things that are **necessary** to exist

Let us now assume that we have a complete list of all things at hand that are *impossible* to exist. Let us further assume that this list is not infinite and that we have enough time to completely go through it. With that we could determine every answer to quests about any thing being possible in principle or not. If something would be possible to exist, it would be not listed on the above mentioned list. But knowing that something is possible to exist does not imply that its existence is also necessary in any way. So even possible things could well be non-existent.

[2] For example an infinite chain of causal events at the microlevel would necessitate an infinite speed for these causes – what is obviously an unphysical premise since we then would arrive at non-local signals that instantaneously could influence other causal chains *ad infinitum*.

Now we have to ask what it should mean that some possible things may never have an existence. Because we just have determined that things that are not on the list of impossible things must at least be possible, it would be a contradiction to now label the same things again as impossible to exist. An alluring way out here would be to conclude that all possible things are at the same time also necessary to exist and the term 'possible' is somewhat a misnomer.

From the above would follow that consciousness is also necessarily existent and all that can be thought in principle will necessarily be thought in principle – somewhere. Everything what happens would be necessarily necessary, the possible and the necessary would be one and the same. The question arises what then would be responsible for at all determining what is impossible and what not. Usually the explanation for all these necessities to happen the way they do is thought of as being due to another necessarily existing thing, namely an infinite landscape of mathematical relationships.

For the sake of our further arguments we now like to assume that this mathematically and strictly deterministic world view would be true. The problem that then arises would be that its mathematical underpinnings are based on a 'causally' closed and self-confirming assumption, namely that mathematical combinations and their iterations are destined to necessarily produce or rather contain a non-denumerable infinity of in-principle unknowable possibilities. If the latter could all be 'predicted', they eventually could be viewed as 'necessary', but the halting problem excludes this prediction by showing its impossibility. Thus, the definite *contents* of that assumption are never completely knowable in principle.

All this should not surprise since we have seen with Gödel's results that even an infinite mathematical system cannot ever know what is unknowable in principle for it. For example it cannot differentiate between a possibility and a necessity. Therefore the existence of the whole construct of a non-denumerable infinity of mathematical 'necessities' leads us to the conclusion that this construct is *itself* just a possibility, it cannot ever be considered a necessary truth. Even if the assumption would be true that this construct is capable of producing conscious awareness about certain parts of itself, this could not ever lead us to this necessary truth – because for reasons of infinite regression, no consciousness that is defined as a certain mathematical pattern within a non-denumerable infinity can completely contain that non-denumerable infinity – unless it would contain itself infinitely many times. Thus, if that landscape nonetheless would exist, it would state about itself (the whole) via a part of it (consciousness of the author) that the whole cannot prove its own necessity. The latter is no wonder, since parts of it cannot ever completely define what this whole should contain or not contain. Moreover, that whole's only global definition is obviously that it is *undefinable in principle*. Thus, it is highly questionable if such an undefined and undefinable 'whole' should at all, by its very own standards, be considered consistent, much less being able to produce some consciousness.

5 An infinity of infinities

So does an infinity of infinities make at all some sense? Yes, since it shows that it is a self-referential construct. Every such construct lacks a certain additional information, otherwise it would not repeatedly refer to itself but to something beyond itself. The lack of these kinds of information is due to a general information shortage which is the natural state of our physical world, since a part of a whole cannot at the same time *contain* that whole. Hence, there is only a finite information capacity available in every part of our world to logically pin down things with a sufficient degree of certainty. Moreover, by looking at the level of elementary particles, it seems that similar to reversing a diapositive, the world is as defined by some informational gaps as it is by some informational content. This analogy may then illustrate why the 'wave-particle duality' can't capture what the atomic blocks of quantum mechanics are, since that duality may well be an interplay of gaps with non-gaps.

Let's therefore examine what can be said about the most atomic "thing as it is", namely a "quantum" entity. By looking at this thing being infused into a Mach-Zehnder interferometer where the path lengths are equal, it isn't possible to explain the measurement results by some wave behaviour of, say a photon, *when* one at the same time also wants to explain why two of such 'wave packets' behave statistically as they do when their polarizations are entangled and both photons are far apart from each other. The conclusion here is that trying to pin down the "thing as it is" without considering mutually exclusive experiments leads to contradictory results. Hence for a consistent picture it must be inferred that *all those experiments* and the "thing as it is" together constitute wholes that are *qualitatively different* from any conceivable combination of some well defined classical objects. Much worse, those non-classical entities make it impossible to be classically modelled such that each and every measurement outcome could be predicted with certainty.

Thus, there obviously exist strong cases in nature where phenomena that are thought to behave deterministically are impossible to be predicted with the help of their thought-to-be deterministic behaviour. Since the needed complete information about such a deterministic quantum behaviour cannot be gathered by *any procedure*, the question arises what 'incomplete information' should be all about when it even cannot be translated into a consistent model of the "thing as it is" that fits all experimental data. Because for example quantum particles can also be considered as wave-packets, we could at first sight conclude that the ontology of the world is fundamentally built up just from these classical wave-packets, constantly interfering with each other and producing some fixed-points called 'particles'. But as was already guessed by the founding fathers of Quantum Theory, this analogy is incomplete and simply breaks down when one attempts to describe 'particles' involved in different experiments as merely classical chunks of energy and/or matter.

So the answer what 'incomplete information' in the above case should be all about is that the *knowledge* of the *in-principle unknowability* of *all* ontological features of quantum systems is obviously all that can be known and thus, *is* complete information. An assumed ontological definition of quantum mechanical behaviour as exhibiting just classical mechanics would therefore violate the law of non-contradiction by saying that in-principle unknowable things can in-principle be known. But such an incoherent unification hasn't worked either for an assumed mathematical landscape to be somehow completable and it can't work either for "the thing as it is" in the quantum realm. Therefore our conclusions are that the common reason for the impossibility of such ontological definitions is that *no formal system* is ever able to construct *itself on its own terms.* Hence it necessarily must have been created by a metaphysical realm of existence [4] and therefore such a created system can impossibly make definitely true *and* complete statements about its own 'fundamental ontology'.

6 Hidden variables and Gödel-incompleteness

Generic undedidability is already present at the level of classical physics, because determining whether or not classical trajectories follow a continuum of infinitely many points in space is generically undecidable – since it is impossible to build a measurement device with infinite resolution. Furthermore infinitely precise initial conditions aren't provable either for the same reasons. This means that algorithmic determinism isn't provable by any causal procedure because *there simply does not exist such a causal procedure*. Another obstacle of algorithmic determinism lies in two considerations: firstly for infinitely precise values the question arises ***which infinite*** series of decimal expansions out of ***infinitely many*** are the 'correct' ones for each and every particle/interaction/ movement. Secondly, the assumption that only a ***finite*** series of decimal expansions govern particle behaviour poses the question ***which*** decimal expansions from the infinite ones one has to *truncate* to match the one-and-only algorithmic determinism. Since no finite theory can ever decide between such an infinity of mutually exclusive properties for the one-and-only physical dynamics, every such theory will fail to mathematically formulate something that is more precise than Quantum Theory. Moreover, we argue that the use of additive or subtractive infinities is an *imaginary act* just like trying to complete or un-complete a Gödel-(in)complete system by *ad hoc* 'adding/subtracting' an infinite amount of *arbitrary* axioms.

For all these reasons we strongly suspect that the idea of a complete determinism is just as imaginary as the idea of a 'completely completable' Gödel-incomplete system. The considerations above also show that the reasons for why one can easily model the statistical behaviour of the quantum mechanical phenomenon of entanglement with the help of a short algorithm is because this, too, is an imaginary act of 'finding' the "thing as it is" within an algorithm that merely draws a Bell curve. The latter is a well defined mathematical object like any other curve and therefore can be mathematically approximated by some proper statistical evaluation together with a universal Turing Machine (computer) that can produce some pseudo-random values. Although such simulations try to conclusively pin down the *Kantian* "thing as it is" [5-6] to be deterministic by simply drawing the Bell curve – and never can succeed to prove anything *physical* from these drawings –, contrariwise *real* physical experiments are sufficient to disprove the ability of all hitherto proposed locally-deterministic hidden variable theories to completely retrace these experiments.

7 Physical reality as an abstraction

We argue that this "thing as it is" isn't completely identifiable not because our measurement instruments are to grainy to detect 'it', but because at the level of elementary particles such *formalizability* arrives at its natural limit. We further argue that a real existential realm beyond space-time has to be assumed whose informational resolution is not defined by any quantitative measures, but by its self-contained holistic *contents*. Otherwise we would remain in the same situation a *computer program* is that cannot grasp that its origins must lie beyond the deductive processes of its algorithms – namely that the latter were *created from outside themselves* for certain reasons [4].

Although it is not yet fully clear whether or not quantum mechanical 'superpositions' play a role for the above mentioned self-contained holistic informational chunks, the former surely are *impossible* to operate on themselves completely within space-time. Because we think that since such a nondenumerable infinity *cannot* be bijected onto parts of itself, *parts* of these 'superpositions' are filtered out just-in-time [7] before being projected onto our space-time from *beyond that space-time.* Consequently the *holistic contents of those unfiltered informational chunks (provisionally called 'superpositions')* become impossible to be completely reconstructed within our phenomenologically downgraded world.[3]

So the assumption of an informationally closed system like space-time would be a simplifying abstraction. If we assume for a moment that the physical world would be constituted by some *finite causal complexity* that could be *completely formalized*, we also had to assume that the underlying mathematics is *not* Gödel-incomplete – to ensure that no mathematical undecidabilities correspond to some physical undecidabilities. With that we would arrive at a physical world that uses only the decidable and complete part of mathematics. Thus, the physics of the human brain then should also function only on the basis of that Gödel-complete mathematics. But with this we arrive at a contradiction, since human brains can grasp and encode the structures of systems that are Gödel-*in*complete. Human brains can for example see that these systems cannot decide whether or not they are inconsistent or merely incomplete. Further these brains can see that these systems must nesessarily be considered incomplete, otherwise one had to conclude that considering them to be incomplete must be caused by a severe inconsistency in the physical workings not only of our brains. The latter however, can't be the case, since by definition of completeness, no Gödel-complete system is able to contain or introduce a logical inconsistency within its axiomatically closed domain (space-time/human brains). Hence, our axiom that the world is an informationally closed (void) system must be erroneous, since the fact that Gödel-incompleteness *is* factually constructible contradicts this axiom.

Conclusions

Our overall conclusions are that the naked act of deduction is really only a deterministic, mechanical process. Nonetheless human beings are able to also *induce* something. Furthermore, if our hitherto made inductions are correct, then not despite of being no deductions, but because these inductions are necessary to explain why there at all exist the limits that have so far been analysed. Therefore, all our exemplifications could also well be considered as a single huge diagonalization argument to step beyond these limits, since it demonstrates that the premise of a complete formalizability for each and every phenomenon turns out to be logically incoherent. Not because ultimate reality would be illogical, but because antivalent logics as well as mathematics are subject to incompleteness. Thus, both can never transcend themselves to be more than an abstraction of ultimate reality. In fact, as we have already seen with an exclusively deterministic world view, such an abstraction would be primarily concerned not about reality, but about its own definitions. But human intelligence can transcend its antivalent logics by inferring that there must be some *irreducible* reasons for those abstractions to work and these reasons must be rooted in a real metaphysical realm.

[3] Similar to a lossy picture compression; 'holistic' as *analogy* to a holographic picture that gets *less sharp* when partitioned.

Consequently, compared to this metaphysical realm, our whole physical universe must in many respects also be considered as a kind of abstraction, being possible due to the existence of this metaphysical realm. We therefore finally want to notice that the latter would not need any abstractions to be what itself is, but our universe would need this realm for at all being able to exist as an abstraction. Thus, by having created the latter, that metaphysical realm would have had some real causally effective, but not deterministically acting powers over its very own abstraction.

References

[1] **Gödel**, Kurt Friedrich (1931). „Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I." In: Monatsh. für Mathematik und Physik, Vol. 38, Issue 1, Springer (Wien), pp. 173-198.

[2] **Tarski**, Alfred (1935). „Der Wahrheitsbegriff in den formalisierten Sprachen." In: Studia Philosophica 1 (Lemberg), pp. 261–405.

[3] **Turing**, Alan Mathison (1936). „On Computable Numbers, with an Application to the Entscheidungsproblem." In: Proceedings of the London Mathematical Society, Series 2, Vol. 42, pp. 230-265. https://www.cs.virginia.edu/~robins/Turing_Paper_1936.pdf. *Correction ibidem*: Proc. Lond. Math. Soc. Series 2, Vol. 43 (1937), pp. 544-546. http://www.turingarchive.org/browse.php/b/12.

[4] **Weckbach**, Stefan Bernhard (2017). „No rule without an exception, except this rule." In: *What is Fundamental?* FQXi Institute, Essay Contest 2017/2018, pp. 5-8. https://fqxi.org/data/essay-contest-files/Weckbach_FQXi1718ContestWec_1.pdf

[5] **Kant**, Immanuel (1868). „Kritik der reinen Vernunft." G. Hartenstein (Ed.) Publ. Leopold Voss (Leipzig), p. 178.

[6] **Kant**, Immanuel (1783). „Prolegomena zu einer jeden zukünftigen Metaphysik." Publ. Johann Friedrich Hartknoch (Riega), pp. 289, 374.

[7] **Weckbach**, Stefan Bernhard (2012). „The Trick of the Light." In: *Questioning the Foundations.* FQXi Institute, Essay Contest 2012. https://fqxi.org/data/essay-contest-files/Weckbach_fqxi_v1.pdf.