The Boundary between Mind and Machine

Dingzhou Fei

Abstract

The mind-body problem is one of the important topics in philosophy of mind and cognitive science. Following the analytical tradition of linguistic and logical analysis, we focus on two aspects of the mind-body problem: one is around Gödel's incompleteness theorem, and the other is on cognitive logic, especially on the question of whether Epistemological Arithmetic and machines are private. In the former case, in response to the popular view that the Gödel Incompleteness Theorem supports dualism in the mind-body problem, H. Putnam constructs a counter-argument that defends the computationalist agenda: the equivalence of mind and Turing machine. In this regard, the paper discusses Putnam's line of analysis and then argues that it is unsound. In the latter case, the paper examines in detail the cognitive form of the expression of privateness and concludes, together with recent results in EA, that the existence of privateness in Turing machines is logically untenable and, therefore, we argue that the mind is not equivalent to a machine in terms of cognitive ability and privateness, i.e., the computationalist agenda is not valid.

Keywords

mind-body problem, Gödel's incompleteness theorem, epistemic arithmetic, privateness of mind

DOI: 10.47297/wspctWSP2515-470201.2018XX01

The mind-body problem is one of the main topics in philosophy of mind and cognitive science. For cognitive science, the answer to the mind-body problem will determine the orientation of its research. Computationalists see cognition as a computable process, so that mind and cognition are for them nothing more than a physical symbolic system with an external interpretation [A. Newell, H. Simon, 1963; J. Fodor 1975]. In terms of the expression and defense of cognition as computation, H. Putnam's argument is, in our opinion, very representative of his unique formulation of the mind-body problem, which captures the important features of the mind-body problem in terms of form and content, in a way that is developed in relation to the concept of the Turing machine. Many of the arguments of computationalism can be considered a variation of Putnam's arguments.

In this paper, we will actually refute Putnam's argument in two important ways: first, the application of Gödel's incompleteness theorem to the mind-body problem, which is a common way of discussing the mind-body problem (E. Nagel, J. Newman, 1960), and second, the privateness of the machine and the mind, which we will show lacks a solid basis for equating the mind with computation or Turing machines. On the other hand, we agree on the approach to the mind-body problem: we agree that the solution of the mind-body problem can be achieved by linguistic and logical argumentation without the help of subjective human experience, in other words, the appropriate answer to the mind-body problem can be given at the logical and linguistic level, and certainly without the need for metaphysical discourse (H. Putnam, 1960). Putnam, 1960), while empirical observation faces its own difficulties [R. Cummins, 2000]. It is worth mentioning that our argument comes from the latest research on the relationship between Turing machines and mind-self knowledge, which is reflected in the advances in research for Epistemological Arithmetic (EA), especially in understanding the modal-cognitive representations of Church's thesis. Gödel's incompleteness theorem is often cited to illustrate the difference between mind and machine, which cannot be understood quantitatively but qualitatively [E. Nagel, J. Newman, 2001]. This implies that it is contrary to Gödel's incompleteness theorem to consider the mind or thought as merely a mechanical process capable of action. This point was challenged by Putnam, who argued that Gödel's incompleteness theorem can be used to argue for the equivalence of mind and Turing machine. To refute Putnam's argument, we first analyze Putnam's argument and Gödel's theorem, then point out the conditions for the homogeneity of mind-machine in this problem, and conclude that the heterogeneity of mind and machine is a logical consequence of Gödel's theorem, i.e., that mind is not a machine and therefore not a computation, thus defending Nagel et al.'s understanding.

The mind-body problem concerns the existence and state of an individual's self-knowledge. For example, how does an individual know about his own pain? Putnam's attempt to use the behavior of a Turing machine as an analogy to the knowledge and state of the individual self seems to negate Putnam's analogy in our discussion. In this regard, the paper uses EA's results on the relationship between epistemic modality and Church's thesis to show that we cannot deny the existence and privateness of self-knowledge, which can be seen as a computational process, and that inferences about the minds of others are in fact closely related to self-knowledge.

Gödel's Incompleteness Theorem and the Limits of Machines. Analytic philosophers like to use Gödel's theorem to argue for their qualitative difference between mind and machine, and Putnam is no exception, trying to show that there is actually no difference in the ability to prove arithmetic theorems between mind and machine. Penrose, on the other hand, uses this theorem to argue the opposite of what he maintains, that the mind and the machine do differ in their ability to prove.

M. Detlefsen. agrees with R. Penrose that machines have in principle an insurmountable obstacle in simulating human proof ability, using Löb's theorem (a theorem closely related to Gödel's incompleteness theorem) as an argumentative tool [M. Detlfsen, 2002].

We conclude that the machine differs in principle from the mind in its ability to prove, not only on the basis of the Gödel incompleteness theorem, but also on the basis of the nature of the Turing machine itself, which is where this paper differs from many of the above. For this reason, we first describe the typical arguments that typically use Gödel's incompleteness theorem to show that mind and machine are not equivalent, and then our analysis proceeds to refute Putnam's ideas.

The usual argument goes like this:

(1) If the mind is equivalent to a Turing machine, then the mathematical propositions that must be proved by humans must be proved for Turing machines; (2) According to the Gödel incompleteness theorem, I can find a proposition that is true but cannot be proved in the Peano arithmetic system; and

for the Peano arithmetic system. (3) According to Gödel's Incompleteness Theorem, I can find a proposition that is true but cannot be proved in the Peano arithmetic system; and for a Turing machine, since a proposition cannot be proved, of course it can no longer be judged true or false. But the human mind can even prove Gödel's incompleteness theorem (outside the Peano arithmetic system, of course), while the Turing machine cannot. (4) Thus, the corollary of (2) rejects (1), or rather, the mind has a proof power that Turing machines do not have. In fact, (1) and (2) of this argument, which is often used by many philosophers, are not so reliable and are not well thought out, mainly because we think that: (a) The proof of "I" is not the same as the proof of a Turing machine in the sense that the former, as a rational agent, involves its rational condition, i. e. a mathematical theorem prover, which can modify the premises, change the conclusion and the existing proof process, i. e. it can have extra-mathematical control over a mathematical proof process, while the proof of a Turing machine should be its ability to recognize the process of mathematical proof. The Turing machine itself cannot change anything about the proof process, it follows a formalized reasoning procedure and instruction set. This difference in the concept of proof leads to a fundamental difference between the mind and the Turing machine: namely, that the Turing machine is only a symbolic system for the operation of the mind and humans, and that the converse is not true, indicating a hierarchical difference between the mind and the machine, which supports in some way the argument that the mind is different from the machine. (b) The proof of the consistency of the Peano arithmetic system cannot be done within the Peano system, but the human mind must take a super-exhaustive order to do so, but this is beyond the philosophical boundaries of exhaustivism and potential infinity, and it is impossible for a Turing machine to prove the consistency of the Peano arithmetic system in this way, which is why we disagree with Potuam that a Turing machine can also prove the consistency of the Peano arithmetic system. (3), one of the main reasons why we disagree with Potuam that Turing machines can also prove (3), which apparently only minds and humans do! So the proof capability of Turing machines seems to be more limited in scope.

Putnam counters that this is the result of a misapplication of Gödel's incompleteness theorem [Putuam, 1961], and that he is defending the argument of (1), which he believes should really mean:

In fact, for any given Turing machine T, what I do is to find a proposition U such that I can prove:

(4) If T is consistent, then U is true.

Here U is undecidable in T if T is in fact consistent.

(5) However, T is perfectly capable of proving that (4) holds.

We believe that this rebuttal by Putnam is unsatisfactory. To do so, we first analyze Putnam's argument, which proceeds as follows: In order for the Turing machine T to prove (4), first assume that the Turing machine can prove that "T is consistent", and then prove that "U is true" under this premise. (b) Turing machine can prove (4):

(c) Putuam's assumption that (3) is provable by the Turing machine T is in fact implicitly based on the fact that if proposition A is provable in the Peano arithmetic system and P[A] and P[A→B] hold (P[A] means that [A] is provable in the Peano arithmetic system and [A] means the Gödel encoding of A. P is a provability predicate of the Peano arithmetic system for P[A → B], analogously).

Thus, T introduces (3) and it must be shown that T can prove the consistency of T. The extremely important premise here is that the proof of T is related to the provability predicate in the PA system, i.e.

(d) T proves that $\Phi \Leftrightarrow PA \vdash \Phi$, PA is shorthand for the Peano arithmetic system, $\vdash$ denotes the notion of being a provable meta-mathematics in PA, and Φ is any mathematical proposition which, for convenience, can be set to be an arithmetic formula in PA.

The significance of (d) is that it turns the notion of a Turing machine proof from a non-explicitly defined intuition into a mathematical concept with a definite meaning, which immediately contradicts the following fact when we try to show that T states the consistency of T: If PA is consistent, then PA $\vdash$Con and Con denotes $\neg P([O \neq O])$, i. e., PA consistency. Thus, although the antecedent of (3) is disproved, and (3) can thus be proved by T, this is of little significance to the argument. Because at this point U can be any formula in PA.

Even if (d) proves (3), there is no clear meaning for T to prove that "U is true", because according to (d), P A ├óΓé¼Γäós U is true does not conform to the language specification in the PA system, because "U is true" is not a PA formula. Putnam's intention may be that U is not testifiable in T, i.e., that it is not testifiable in PA, but that U is true. But the question is how does T determine that U is true? There is a fundamental difference between U being provable, which is grammatical, and U being true, which is semantic, and U, for which T can only give grammatical concepts, while semantics can only be given by the human mind. This has been pointed out by many philosophers [S. Harnad, 1997].

It is important to note that the argument in this paper differs from that of R. Penrose, although we both agree that the mind is not equivalent to a machine, since our main argument against the mind being a machine is, as we have just pointed out, that machines do not have the concept of "meaning". It can only be given concepts of meaning, such as truth and falsehood, by the mind from outside. When we give the machine an axiomatic set and a scheme of reasoning, it would be wrong to assume that we also give the machine the corresponding notion of mathematical truth. This formalist view of mathematics is also proved wrong by Gödel's theorem. In fact, what Putnam is saying is that a formula that cannot be proved in PA can be recognized as true or false by comparing its content with the relevant facts, and that machines do the same, but this does not lead to his philosophical conclusion that the mind is equivalent to a machine.

Another reason we have against Prnarose is that it actually contradicts Löb's theorem, as pointed out by H. Deutsch [Deutsch, H, 1996].

The private nature of the machine and the mind

The mind-body problem is also plagued by the privateness of the self. It seems that the privateness of the self is a strong evidence for dualism, but this is only half true, because dualism does not directly imply privateness, and privateness can be discussed on the basis of mind-body homology. The point we are trying to make here is that in philosophical analysis, to emphasize the private nature of the self and to argue that it cannot be discussed at the level of observation and logic is contrary to the goal of the analysis and falls into mysticism. Like Putnam, we agree with the position of studying the privateness of the self: it is really a purely linguistic and logical problem. The difference, however, is that we consider the self-privacy of the machine (as opposed to the privateness of the mind) to be suspect.

First, let us outline what is the privateness of the self and how to define the privateness of the corresponding machine (in this case, of course, the Turing machine).

For the ego, the following proposition expresses a privateness:

(6) I know that p,

p is a proposition that determines the state, the nature of the ego. Is what is expressed in (6) true with respect to privateness?

Or to diminish one step, is there a truth-value condition that might make (6) hold? The truth condition here must be in the sense of being observable and logically deducible. Many, such as K.R. Stueber [2002] and R. Jacobsen [1997], assert that we can answer this question within the context of observation and experience, and for Jacobsen, the sentence of first-person self-statement qualifies as an observational sentence class for testing because various first-person self-statement sentence classes can be compared and screened at the level of physical events. Thus, our declarative sentences are equivalent to the principles of laws in the natural sciences, and although they are tentative conjectures, they stand up to many rigorous tests. For Stueber, on the other hand, he sees natural declarative sentences as the result of a default reasoning mechanism, although each person asserts its own intrinsic state, the reasonableness of which is that the jump of its own default reasoning mechanism is acceptable. Thus, in order to answer these two questions, it is inevitable to go back to oneself: the observation itself justifies the observation and the default mechanism itself justifies itself. Still, we need to point out that these arguments often do not avoid the circular question: why is the generalization of self-referential sentences justified? Why is the default reasoning mechanism reasonable?

In this paper, we wish to avoid this, and therefore we reformulate the self-privacy thesis (6). Indeed, this must be appended.

(7) I know that P(I/you), P(I/you) is the result of replacing the free variant I (an agent symbol) in P with you (another agent symbol). For example, if P stands for "I have a headache", then (6) says: "I know I have a headache" and (7) says "I know you have a headache". What is proposed in (7) is that the expression of privateness should not be a barrier to the communicability of the ego-subjects, and that the ego-subjects should not be able to do to others what they do to themselves. This ensures that the analysis of self-privacy can be carried out smoothly. However, in order to be able to guarantee (6), (7) there must be (8), (9)

(8) I know that P implies P(I/you)

(9) I know that P,implies q, then if I know thatp,then I also know that q

q is the same as p for propositions involving ego states. Thus, if there are (6), (8), and (9), it follows that (7).

For (9), given our semantics about "knowing", it should hold; the question is whether (8) holds. Here, it is important to mention the fact that if "I know that p" is said to be Kp, then

(10) Kp↔Kp (you/I)

Note that p is a proposition, i.e., it is either true or false, and we consider that it is not a free variant, and that I and you appear as bound variables in p, (10) as one of the theorems of the so-called Epistemological Arithmetic (EA) system. The so-called EA is in fact a new system that expands the first-order language by introducing the "I know" K-operator, whose atomic formulas include, in addition to the usual first-order form, atomic formulas of the form Kp. The basic "I know" notation is as follows:

· If I know φ and I also know that φ implies φ, then I know that φ

· If I know φ, then φ is true.

· If I know φ, then I know that I know φ

· If φ is a substitution special case of first-order reduplication or first-order reduplication in EA, then I know that φ

In short, the privity of the self becomes part of the inference of the EA logic, and thus it is fully logically deducible within EA.

· Peano's arithmetic axiom on EA [Shapiro, S, 1988]

At the same time, this qualification of self-privacy makes it possible that the formula provable in EA is not known to the self, e.g., the formula $K\forall x\varphi \rightarrow \forall xK\varphi$ is provable in EA but does not satisfy our qualification of self-privacy, i.e., the self does not know the truth or falsity of this formula. [T.J. Carlson, 2000]. This may be the cost of our qualification of self-privacy.

So what does it mean to say that Turing machines have self-privacy? On this issue, we are less extreme than, for example, J. Searle, who argues that a Turing machine cannot be self-aware because it does not generate its corresponding intentionality, i.e., a Turing machine cannot be conscious, and certainly not private with respect to self-awareness [J. Searle, 1984], and, like Putnam, we understand privateness in the sense of a functional analogy. The formulation is as follows:

Let T be a Turing machine, then the privateness problem can be formulated as follows:

(11) How does T know that it is in state A?

(12) How does T know that another Turing machine, T′ , is also in state A?

Assuming that T has some Accessible Relation with T′

"Know" here is understood in the sense of EA mentioned above. In fact, Turing knows its state in the following sense machine knows its own state: i.e., it has a description of itself, i.e., it is self-reflective, so that it can also recognize every state that T experiences, which answers (11), for (12). Similar to the self-replication of the Turing machine T involved in (11), it can take T′ as input and then check whether the content on the tape wraps around T′ and if so stops, then the output is T′, which shows that T knows T′, which in Putnam is done with T and T′ with the help of some device in T similar to vision of T. If the criterion of privity like (11), (12) is simply satisfied, then there is no functional difference between the mind and the graph machine, as Putnam argues. However, as we shall see, there are in principle limits to the self-privacy of Turing machines, given their performance in EA. We should point out that Putnam's argument implicitly presupposes that T is automatically informed about its state as it goes through states A, B, C, and so on. This argument is made in such a way that the concept of knowing lacks a descriptive definition under a Turing machine, and this definition must be similar to the one mentioned above about human knowing, otherwise the argument for difference or equivalence of mental machines becomes ambiguous. For example, Putnam's difference between the following two sentences:

(13) T knows that it is in state A

(14) Jones knows that he is having a headache

In fact, Putnam takes knowing in this sense: it is knowing based on observational evidence, which is obviously not what we have proposed a logical concept with an EA meaning. Otherwise, comparing the similarities and differences between the two is a meaningless matter. For (14) is of course based on observation, and for (13) is of course based on introspection, but (14) has to be derived based on introspection, because headaches are after all the result of introspective extrapolation.

So this brings us to the question of the self-privacy of the mind we are discussing: how to understand or know the mind of others? According to our proposed criterion, so that T in φ is a proposition about the state of itself and K is "T knows", the criterion on the self-privacy of the Turing machine can be written or:

(15) $\vdash_{EA} K\varphi \rightarrow K\varphi(T/T')$

$\vdash_{EA}$ denotes the deductive relation in EA, and (15) is actually saying that T knows that it is a Turing machine, then T′ also knows that it is a Turing machine.

knows that it is a Turing machine, and at the same time, T must also know that T′ is a Turing machine in order for Kφ(T/T′) to make sense. Alternatively, T is a Turing machine and knows which one it is, but this contradicts Reinhardt's conclusion: Reinhardt has shown that this conclusion is inconsistent with EA [W. Reinhardt 1986], which suggests that if we trust EA sufficiently. Then, (15) is not trustworthy in any case, which suggests that if we want to argue that Turing machines have human-like self-privacy we face logical inconsistencies, and therefore we have to give up that any Turing machine can experience self-privacy as well as the human mind. It is worth noting that the strong machine thesis SMT (or Post-Turing thesis): "I know I am a Turing machine" has been shown to be consistent with EA, suggesting that Turing machines have some degree of self-awareness [T. Carlson, 2000], although their self-awareness also has inherent insurmountable obstacles. However, its self-awareness also has an inherent insurmountable obstacle, since it cannot distinguish which kind of self it is, and thus its self-privacy is in principle limited. This also proves that the self-privacy of a Turing machine is not comparable to the human ego, and therefore the computationalist equivalence of mind to computation is not firmly grounded.

We argue that the mind is different from computation or Turing machines in two main ways: first, in its ability to prove, and second, in its self-privacy, an aspect often discussed by computationalists, as exemplified by Putuam's arguments. Our rebuttal is also based on logical and linguistic grounds, but we cite much recent evidence from the philosophical analysis of Gödel's Incompleteness Theorem and new developments in EA. Rather than discussing what the mind actually is, we show that the mind is never a computation, or a Turing machine.

References

[1] T. Carlson Knowledge, machines, and The Consistency of SMT, Annals of Pure and Applied logic, 105 (2000), 51-82

[2] R. Cummins, D. Cummins, introduction to Minds, Brains, and Computers. Blackwell Publishing Ltd,2000

[3] M.Detlfesen, Lob’s Theorem as a Limitation on Mechanism, Minds and Ma-

chines, 12(3), 20024. H. Deutsch, Deconstructing mathematics and mind, http//:www.ptproject.ilstu.edu
[5] J. Fodor: The Language of Thought: First Approximations, available from The Language of Thought, Harvard University Press, 1975
[6] S.Harnad, Computation is just Interpretable symbol manipulation, Cognition isn't Minds and Machines, Vol4, 1994
[7] R.Jacobsen, self-Quotation and Self-knowledge, Sythese, Vol 110, 1997
[8] E. Nagel，J.Newman, Gödel's Proof (revisted) New York University Press, 1960
[9] A.Newell, H.Simon, GPS, a Program that simulates human thought,available from. Feigenbaum, J. Feldman(eds.) Computers and Thought, McGraw-Hall, 1961
[10] R.Penrose, shadows of the Mind, Oxford University Press,1994
[11] H.Putnam, Minds and Machines, available from S.Hook(ed.),Dimensions of Minds a: Symposium, New York University Press,1960
[12] W.Reinhardt, Epistemic theories and the interpretation of Gödel incompleteness theorems, J. Philos. Logic 15(1986),427-474
[13] J.searle, minds, Brains and Science, Harvard university Press,1984
[14] S.Shapiro, lntensional Mathematics,North-Holland Publishing,1991
[15] K.Stueber, the Problem of self-Knowledge, Erkenntnis, N0,3, 2002