Article

Not peer-reviewed version

Rethinking Human and Machine Intelligence through Kant, Wittgenstein, Gödel, and Cantor

Jae Jeong Lee *

Posted Date: 13 October 2023

doi: 10.20944/preprints202310.0876.v1

Keywords: Ontology; AI; incongruent counterparts; computationalism; emergentism; determinism doi:10.20944/preprints202310.0876.v1

Article

Rethinking Human and Machine Intelligence through Kant, Wittgenstein, Gödel, and Cantor

Jae Jeong Lee

College of English, Hankuk University of Foreign Studies, 107 Imun-ro, Dongdaemun-gu, Seoul

Abstract: This paper proposes a new metaphysical framework for distinguishing between human and machine intelligence by drawing on Kant's incongruent counterparts as an analogy. Specifically, the paper posits two deterministic worlds that are superficially identical but ultimately different. Using ideas from Wittgenstein, Gödel, and Cantor, the paper defines "deterministic knowledge" and investigates how this knowledge is processed differently in those two worlds. The paper considers computationalism and causal determinism for the new framework. Then, the paper introduces new concepts to illustrate why human and machine agents display different causal characteristics in processing verbal information. Overall, the paper's framework provides a theoretical basis for the uniqueness of the human mind.

Keywords: ontology; AI; incongruent counterparts; computationalism; emergentism; determinism

---

Introduction

This paper was motivated in part by asking questions that arose from Nietzsche's concept of *amor fati* (or "love of fate") (Nietzsche 1990, 99). His suggestion to embrace and love fate has resonated with many of his readers seeking guidance in life. However, Nietzsche's advice seems to have certain inconsistencies. For instance, how can one learn to embrace fate if everything in one's life has been predetermined? If determinism is true, would it not be accurate to say that even an act of learning to love one's fate was also predetermined from the beginning? Was Nietzsche himself, perhaps, predestined to admonish his readers to love their fate? It is doubtful if it truly matters whatever people choose to believe or say if everything was predetermined. Furthermore, when determinists assert that the universe is deterministic, it is difficult to avoid the impression that they are putting themselves away from the universe that they belong to. Apparently, none of them have provided a convincingly strong basis for justifying the significance of their assertion while situating themselves within the universe.[i] In other words, no qualitative distinction has been established between the act of declaring the universe as deterministic and all the events of the universe that should also comprise the very act of declaration.

One may argue that such an issue has been addressed through compatibilism, which proposes that one's perceived sense of free will is compatible with determinism. However, the compatibilist view of human nature still relies on causal determinism, which is rooted in the notion of causality. This may reinforce the idea that humans are not essentially different from computing machines.[ii] Therefore, compatibilism itself may not be particularly helpful in clarifying what significant distinction lies between the determinist's mind and the events of the universe that are within the determinist's scope. If compatibilism is true, it is possible that the human mind differs from computers or other physical events of the universe only in terms of complexity.

To address this issue, this paper proposes a novel philosophical perspective by discussing different types of deterministic worlds. These deterministic worlds are carefully examined with respect to the human mind. Kant's "incongruent counterparts" (hereinafter, "ICs") was a major inspiration for the development of the metaphysical argument in this paper. The argument also builds upon the idea of Cantor's diagonal argument, Gödel's *proof strategy* for his incompleteness theorem, and Wittgenstein's first proposition in Tractatus Logico-Philosophicus (hereinafter, "TLP"). Admittedly, the use of ICs as an analogy may appear to be far-fetched, since the original idea concerns absolute versus relational space. In addition, given that Cantor's and Gödel's are mathematical theorems, their connections to determinism might be initially difficult to grasp. Moreover, one may question the relevance of the metaphysical argument to the empirical world. This paper will address these concerns.

1. The Incongruent Counterparts

Kant devised the concept of ICs in order to address the issue of absolute versus relational space (Kant 1994, 145-174). According to the theory of absolute space, even if the universe had only one body and nothing else, that body would still have a spatial background in which it could move (Asher 1987). However, the relational view of space denies the existence of absolute space and defines motion only in relation to other bodies.

Kant begins his argument by imagining two worlds. One world includes only a left hand ("LH world"). The other world includes only a right hand ("RH world"). If the relational view is correct, there should be no difference between these two. However, from an outsider's viewpoint, it is clear that the two worlds are different. Therefore, Kant concludes that the relational theory of space must be incorrect.

However, the objective of this paper is to use the IC analogy as a speculative tool for discussing the nature of determinism in relation to human reasoning, rather than address the absolute/relational space controversy. To accomplish this, the paper will consider the following cases based on Kant's original concept.

(LH1) A right hand cannot enter into an LH world. Also, the right hand is inconceivable in the LH world.

(LH2) A right hand can enter into an LH world, and if it does it will be perceived no differently than the existing left hand.

(RH1) A left hand cannot enter into an RH world. Nevertheless, its attributes can be hypothesized in the RH world.

(RH2) A left hand can enter into an RH world. Also, the RH world can hypothesize the attributes of the left hand before such entry takes place.

2. A Deterministic Knowledge Argument

This paper will use the following three key definitions for the discussion:

(1) Deterministic knowledge (D knowledge): The totality of facts associated with all the past, present, and future events in a deterministic world. The totality coincides with every time point of the world.

(2) Knowledge-in-hindsight (H knowledge): The totality of facts associated with all the events in a world ranging from the beginning up to a particular time point. The totality coincides with the particular time point and the time thereafter.

(3) Metaphysically open deterministic world: A deterministic world where there is a metaphysical sense in assuming a scenario in which its deterministic knowledge is provided to a cognitive agent of the world.

Definitions (1) and (2) were influenced by Wittgenstein's idea that the world is the "totality of facts" (Wittgenstein 1922, 25). In Definition (3), the idea of the cognitive agent receiving D knowledge bears a resemblance to the "circular-seeming idea of substituting a string's own Gödel number into the string itself" (Nagel & Newman 2001, 89).

But what is a string? In a formalized system of mathematics, "postulates and theorems" are "'strings' (or finitely long sequences) of meaningless marks, constructed according to rules" (Nagel & Newman 2001, 26). Further, a Gödel number is a "unique number [assigned] to each elementary sign, each formula (or sequence of signs), [or] each proof (or finite sequence of formulas)," which "serves as a distinctive tag or label" (Nagel & Newman 2001, 69). D knowledge for the world can be likened to the Gödel number assigned for a math theorem. Also, the cognitive agent can be compared

 doi:10.20944/preprints202310.0876.v1

to a variable within the math theorem. Just as the Gödel number for the theorem is substituted into the variable within the theorem, the D knowledge is fed back into the cognitive agent's information processing mechanism.

The result of such feedback in (3) can be illustrated using Cantor's diagonal argument as another analogy. According to Cantor, the proposition that real numbers correspond one-to-one with natural numbers is false because a distinct real number can be created that is not included in a hypothetical list where real numbers correspond to natural numbers (Simmons 2008, 20-22). See the following simplified illustration.

Suppose that natural numbers correspond one-to-one with real numbers within [0, 1]. For instance:

n = 1 and r_1 = 0.[1]0010…

n = 2 and r_2 = 0.0[0]100...

n = 3 and r_3 = 0.11[1]00…

n = 4 and r_4 = 0.001[1]0…

n = 5 and r_5 = 0.1111[1]…

… and so on.

r_d = 0.01000… can be generated by changing 0 to 1, or 1 to 0 in the digits in the brackets. This real number does not exist in the existing list. This proves that no one-to-one correspondence can be established between natural numbers and the real numbers within [0, 1].

Imagine what might happen when r_d = 0.01000… is paired with a new natural number and then added to the existing list of real numbers. Upon examination, a new real number would emerge that is not included in the updated list. The concept of D knowledge shows a similar characteristic. D knowledge should include all the descriptions of the physical events related to the agent. The verbal descriptions correspond one-to-one with the physical events. Suppose that the agent gains access to the content of the D knowledge. However, this access is not described in the existing verbal descriptions included in the D knowledge. Also, gaining access to it would generate a new derivative version of D knowledge that describes the agent's access to the original D knowledge.

Undoubtedly, the idea of the cognitive agent receiving the D knowledge is unconventional and seemingly contradictory. How could someone know about one's future if it was preordained? One way of circumventing this contradiction is to assume that a particular deterministic world is contained within a larger system and that there exists a mathematical probability that the descriptions in the D knowledge will be at a particular time point provided to the agent from the larger system. Technically speaking, however, that would be an indeterministic world. Accordingly, this paper proposes to examine a hypothetical situation where the agent receives[iii] the D knowledge in a *metaphysical* sense.

Another unconventional aspect of the argument is the assumption of two apparently identical but different deterministic worlds. For example, Schwartz defines determinism as the view "that [possible] worlds cannot be the same up to a point and then diverge" (Schwartz 2012, 216). However, in this paper's thought experiment, it *is* possible for two deterministic worlds to be computationally the same up to a point and then diverge when D knowledge is provided to them. If one maintains that the human mind cannot be fully reduced into an algorithm, then it is necessary to assume that a divergence between the two worlds is possible. Specifically, this paper presents the following two deterministic worlds that are established as "ICs":

(i) The original world, and

(ii) A simulated world that replicates every aspect of the original world and realizes the human mind through computationalism (i.e., the computational theory of mind), characterized by an input-output system that may involve stochastic[iv] elements.

These two worlds are assumed to be *metaphysically open*. From a computational viewpoint, both worlds are the same. However, they are ultimately different as they produce different outcomes in response to D knowledge.

2.1. Predefined Deterministic Knowledge

If the D knowledge of the simulated world were provided to its cognitive agent, the agent would process it simply as one of the available inputs that is closest to the D knowledge and produce a corresponding output. This means that the agent's cognitive mechanism operates with **rigid processing**, as the agent cannot process in any other way an input that it was not configured to receive. Accordingly, this paper defines the simulated world as **trivially deterministic**. This world is governed by *predefined* D knowledge. This knowledge *dictates* how things should occur.

Based on the IC analogy, it is observed that the simulated world is physically characterized by "LH1." Recall that a right hand cannot enter into the LH1 world. Similarly, D knowledge cannot be provided to the simulated world. Additionally, the simulated world is metaphysically characterized by "LH2." If a right hand enters into the LH2 world, it will be perceived no differently than the existing left hand. Likewise, even if the D knowledge were provided to the simulated world, it could not be identified by its cognitive agent as distinct from all the other existing available inputs.

See below the input-output mappings for the time point that the agent receives the D knowledge. Since this is a deterministic world, the agent is originally designed to receive only one input from I_1 to I_n. The other inputs are provided only as hypotheticals that could have been processed from a computational viewpoint. Specifically, the other input-output pairs than the actual input-output pair serve to illustrate counterfactual cases. These cases are also described in the predefined D knowledge.

Input set: $I_1, I_2, \ldots, I_n$

Output set: $O_1, O_2, \ldots, O_n$

$1 \leq k \leq n$

I_D = D knowledge

$I_D = I_k$

$O_D = O_k$

However, the above mappings are based on a non-stochastic model, which does not allow for indeterminacy. Based on the notion of stochasticity, it is possible to construct the mappings below. Each subset of the output set is constructed such that the probabilities of the elements within each subset add up to 1.

$I_1, I_2, \ldots, I_n$

$\{O_{1[1]}, \ldots, O_{1[s1]}\}, \{O_{2[1]}, \ldots, O_{2[s2]}\}, \ldots, \{O_{n[1]}, \ldots, O_{n[sn]}\}$

O_D = One element from $\{O_{k[1]}, \ldots, O_{k[sk]}\}$

In the stochastic model as well, it is seen that the agent's response to the D knowledge remains trivial because its response cannot be anything other than the predefined outputs.

In the simulated world, suppose that there is an AI philosopher named Susan. She loves coffee but often hesitates whether to have a coffee every morning. She loves coffee for its taste. Besides, its caffeine helps fuel her insights when developing a line of philosophical thinking. However, she also worries about a potential side-effect of caffeine such as insomnia. One morning, she decides to have a coffee anyway without knowing that it would cause her insomnia later that night. She starts drinking it while reading her philosophical essay draft through a tablet device. In this case, suppose that there is a 60% chance that she will stop drinking her coffee if she is somehow convinced that she will not be able to fall asleep at night. The following mappings are established for a specific time point in the morning:

I_1 = She feels thirsty (possibly due to the caffeine).

I_2 = Nothing happens other than the continuous visual influx of texts from her tablet.

I_3 = She is convinced that she will not be able to sleep tonight because of her coffee.

O_1 = She drinks a glass of water on her desk.

O_2 = She continues to read.

O_{31} = She stops drinking her coffee.

O_{32} = She continues to drink her coffee.

However, since the world is deterministic, it can be assumed that only a particular input such as I_2 was configured to occur at the specific time point. Meanwhile, in a metaphysical sense, it is possible to assume that specific descriptions in the D knowledge could be provided to her at the specific time

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 October 2023 doi:10.20944/preprints202310.0876.v1

point. For instance, suppose that her tablet suddenly displays detailed descriptions involving all of her activities that occurred in the morning (such as having breakfast or checking the weather outside), her inner thoughts and emotions throughout the morning, and a subsequent scenario of the day to unfold involving her loss of sleep due to the coffee. How would she respond to this? She will most likely be "convinced that she will not be able to sleep tonight because of her coffee." Then, given the 60% chance, she will probably stop drinking the coffee.

2.2. Reflective Deterministic Knowledge

If the D knowledge of the original world were provided to its cognitive agent, the agent would perceive it as a different input than the available inputs and generate a new corresponding output. This means that the agent's cognitive mechanism exhibits **emergent processing**,[v] as the agent can distinctly identify a particular input that it was not supposed to receive. Accordingly, this paper defines the original world as **non-trivially deterministic**. Using the IC analogy, this world can be physically characterized by "RH1" and metaphysically by "RH2." Further, it is possible that the above D knowledge is *reflective* D knowledge. This knowledge only *reflects* every physical event across time. Unlike predefined D knowledge, it does not describe counterfactual cases. Also, reflective D knowledge is compatible with the block universe theory.

In the block universe model, "[w]hether past, present or future, all events 'lie frozen' in the four-dimensional block, much like the scenes from a movie are fixed on the film roll" (Thyssen 2020, 6). If one were to see the events of the universe like fixed scenes on a film roll from an omniscient viewpoint across time, that person might be able to extrapolate counterfactual cases in relation to those events. However, the scenes themselves do not include such information. In that sense, the reflective D knowledge only mirrors the physical events.

Meanwhile, it is assumed that emergence of a new output in response to D knowledge is *necessary*, considering that the agent's cognitive mechanism is usually assumed to be governed by causality. However, the *content* of the new output may be deterministic or non-deterministic. This is highlighted by the question mark in the input-output mappings below. The input-output pairs other than the actual input-output pair are provided as dummies whose contents are unknown. "I_{n+1}" is enclosed in the parentheses to indicate that it is only a latent input for the agent.

$I_1, I_2, \ldots, I_n, (I_{n+1})$

$O_1, O_2, \ldots, O_n, (?)$

$I_D = I_{n+1} \qquad O_D = ?$

If the Susan scenario happened in the original world, she might have been struck to the core and asked, "Am I living in a Matrix?"

2.3. Causal Deterministic Knowledge

Based on the notion of causality, this subsection defines *causal* D knowledge. Specifically, the D knowledge of a causally deterministic world is *generated* by the first cause of the world.

Causal determinism holds that everything that has happened could not have happened otherwise and that everything will happen the way it is supposed to (Hoefer 2023). Such an idea of strong causal connections is applicable to a non-trivial world without inconsistency. Assume that everything is deterministic in the metaphysical as well as physical realm of the non-trivial world. Then, in the metaphysical realm, it follows that the agent should produce a new corresponding output whose content is deterministic in response to each derivative version of D knowledge (i.e., D', D'', and so on).

This paper defines such property as **hard causality**. It renders deterministic both physical and metaphysical scenarios at the very beginning. In that sense, it can be described as an extreme version of causal determinism. Specifically, hard causality suggests that infinitely many derivative versions of D knowledge are causally generated at once, which would make it *impossible* to construct a simulated world through computationalism. In Susan's situation, she would have to give a different response to each derivative version of D knowledge. Specifically, she should not keep only saying

"Am I living in a Matrix?" with regard to every derivative version of D knowledge. If she does, this would show she relies on rigid processing.

Accordingly, there can be no *predefined* D knowledge that *dictates* a non-trivial world. In order for such knowledge to exist, the contents of new outputs would have to be predefined in response to infinite derivative versions of D knowledge before the world could begin. This is not a plausible idea. See the mappings below.

$I_1, I_2, \ldots, I_n, I_{n+1}, I_{n+2}, \ldots$ and so on

$O_1, O_2, \ldots, O_n, O_{n+1}, O_{n+2}, \ldots$ and so on

$I_D = I_{n+1}$, $O_D = O_{n+1}$

$I_{D'} = I_{n+2}$, $O_{D'} = O_{n+2}$

… …

2.4. Knowledge-in-Hindsight as a "Fail-Safe"

Unlike in the above cases, suppose now that the original world is indeterministic. Then, the concept of D knowledge would be useless, rendering the D knowledge argument futile. In this case, H knowledge can be used as a replacement. In a metaphysical sense, a cognitive agent of the past would still allow for generation of an output in response to the H knowledge that the cognitive agent is subject to. Further, even if there were exceptional events in the history of the universe where causality failed, this problem can be addressed, because H knowledge provides information on a continuous sequence of events regardless of whether or not they were causally interrelated. As long as causality works in regards to the agent's reception of H knowledge, the thought experiment remains valid. The point is that H knowledge would also be interpreted distinctly by the human agent that is subject to the H knowledge.

3. Diagonalization through Concatenation

This section introduces new terms that can help explain different causal characteristics exhibited by human and machine agents in processing verbal information. Additionally, it presents a theoretical basis for the human agent's way of processing verbal information.

3.1. Definitions

When a sentence is input into a machine agent, it processes the sentence as a mere **concatenation** of words. The machine has no sense of a temporal flow when executing the process. It simply moves from one bit to another during its information processing. On the other hand, when a sentence is presented to the human agent, the agent forms a mental image of the subject word and *retains*[vi] it up to the point of recognizing the predicate. Ultimately, the images of the subject and predicate are combined to create a holistic image of the sentence itself. This process can be defined as **diagonalization**.[vii][viii]

3.2. Continuity of Space and Time

The process of "retaining" mental images raises the question of how any spatial/temporal transitions can occur if space and time are continuous. For instance, what does it mean to move within continuous space when no immediately subsequent coordinate can be defined with respect to an origin? This issue can be resolved through an "ontological" argument. Specifically, transitions can happen because they should happen in order for the notion of continuity to be established. As illustrated in Zeno's paradox, continuity is discovered *retroactively*[ix] through endless transitions. Without relying on these transitions, it is impossible to identify continuity. Therefore, transitions must exist. Moreover, the very initial distance between the two points that is to be split in two ensures the presence of a discrete leap in real space. Ultimately, it can be proposed that the human agent's perceptual mechanism *proactively* achieves a discrete leap in real space and time, by retaining relevant information (e.g., perceptible spatial/temporal coordinates) along the way. This enables the human agent to process verbal information in an inherently different way.

4. Further Implications

This section investigates one possible empirical case of using the metaphysical concept of D knowledge.[x] Further, it explores how the initial question posed in the introduction of this paper – how is it possible that the determinist can declare the universe to be deterministic while remaining part of the deterministic universe – can be resolved based on the D knowledge argument.

4.1. Quasi-Deterministic Knowledge

Consider a prototype AI named "TARS." Assume that TARS relies on *non-stochastic* processing, which simplifies the experimental setup. With regard to the replicas TARS0 and TARS1, conduct the following experiment: (1) Place TARS0 in a controlled environment, (2) provide it with various inputs from $t=t_0$ to $t=t_n$, (3) collect its corresponding outputs, and (4) compile the data. Next, place TARS1 in the same environment and provide it with the same inputs at the exact time points as TARS0 only up to $t=t_k$ ($0<k<n$). Finally, at $t=t_n$, provide TARS1 with the data compiled from TARS0. The key idea is that the compiled data serves as an input that TARS1 was not supposed to receive. If TARS1 produces an emergent output, as described in Section 2.2 on emergent processing, it can be considered to have surpassed conventional AIs.[xi]

In the above case, the compiled data can be regarded as an approximation to a narrow breadth of D knowledge specific to TARS1. While the compiled data partially constitutes H knowledge for humans, it can serve as **quasi-D knowledge** for TARS1.[xii] Admittedly, it must be very difficult to determine if the above output truly represents a new response to an input other than the available inputs, due to the technical difficulties involved. Nevertheless, the metaphysical argument remains significant as it provides a standard for evaluating an AI's level of enhancement.

The TARS experiment has an advantage over one existing method of testing an AI on its enhancement. In her book "Artificial You," Schneider proposes an AI Consciousness Test (ACT). To test if an AI has consciousness, she suggests "'boxing in' an AI – making it unable to get information about the world" (Schneider 2019, 53). Schneider emphasizes that "the AI's vocabulary must lack expressions like 'consciousness,' 'soul,' and 'mind'" (Schneider 2019, 54). Then, the AI can be asked a question like "Could you survive the permanent deletion of your program?" (Schneider 2019, 55). If the AI's answer is similar to what a human being might provide, it can indicate some evidence that the AI is conscious. However, in the TARS experiment, an AI can be taught all those words and still be tested on whether it has achieved emergent processing.

Regarding the TARS experiment, consider adding a third replica, which will be referred to as TARS2. If TARS2 is provided with the same quasi-D knowledge that was received by TARS1, in the same manner that it was provided to TARS1, will TARS2 produce the exact same output as TARS1's? The outcome will have to be interpreted differently according to each world model (e.g., a metaphysically open/closed, deterministic/indeterministic world).

4.2. The Determinist versus the Deterministic Universe

In "Freedom Evolves," Dennett notes that "confusion [over determinism] arises when one tries to maintain two perspectives on the universe at once." One perspective is the "God's eye" perspective, and the other is the "engaged perspective of an agent within the universe" (Dennett 2003, 93). His description of the former perspective coincides with the Parmenidean view of the universe. Specifically, he states that "[f]rom the timeless God's-eye perspective nothing ever changes" as "the whole history of the universe is laid out 'at once'" (Dennett 2003, 93). It appears that Dennett is giving equal weights to both perspectives but cautions against assuming them at the same time. He stops short of providing a philosophical scheme in which both perspectives can coexist. Also, he does not explicitly state that when he is expressing support for determinism, he is doing that from a provisional God's-eye perspective. As a matter of fact, every philosopher and scientist who makes a declarative statement about the universe at large assumes such a perspective. However, every such individual is also part of the universe. How to reconcile this discrepancy?

According to the D knowledge argument, a human agent is distinguished by its capacity to process even "otherworldly but comprehensible" inputs (i.e., D knowledge).[xiii] This temporarily sets a determinist apart from the objects of the determinist's investigation that are to include the determinist when the determinist declares the universe to be deterministic. This relationship between the two creates a dialectic circle[xiv] that grows as the determinist and the objects/events continue to encircle each other in an alternating manner. This expanding circle provides a more sophisticated illustration of the dynamics involving the determinist and the universe, compared to the diagram in Wittgenstein's TLP of the metaphysical subject's eye that remains encapsulated within the world's periphery (Wittgenstein 1922, 75). It is this dialectic circle that provides a holistic scheme for investigation of the universe.

5. Conclusion

The major ideas of this paper can be outlined as follows.

(1) Deterministic knowledge ("otherworldly but comprehensible")

- Predefined D knowledge: dictates the world (computationalism)
- Reflective D knowledge: reflects the world (block universe theory)
- Causal D knowledge: generated by the world (hard causality)

(2) Knowledge-in-hindsight

- A "fail-safe" in case causal determinism fails
- Quasi-D knowledge for an AI machine

(3) Metaphysically open deterministic world

- Trivial determinism: rigid processing (concatenation)
- Non-trivial determinism: emergent processing (diagonalization)

Based on the above conceptual scheme, this paper has sought to preserve the uniqueness of the human mind while allowing for hard determinism. Instead of discrediting computationalism and causal determinism, it has integrated them into a comprehensive framework. Recall that the investigation in this paper began by critically questioning Nietzsche's *amor fati*. As such, it has an underlying humanistic motivation as well. Ultimately, the above metaphysical model can provide a foundation for investigation of the human mind and its place in the supposedly deterministic universe.

However, this paper is subject to limitations, including heavy reliance on metaphysical speculation and lack of empirical evidence. For instance, in Section 2.3, the idea that the human mind is capable of providing a distinct response to each of the infinite derivative versions of D knowledge may not be deemed plausible by several readers. Further, the conception of D knowledge may face challenges from quantum physicists, who argue that describing physical events through exact spatial/temporal coordinates on the quantum level is impossible in principle. Additionally, the D knowledge argument cannot explain the phenomenon of qualia or a sense of agency and free will. These problems require further study.

References

Asher, W. O. "Berkeley on absolute motion." History of Philosophy Quarterly 4, no. 4 (1987): 447-466. http://www.jstor.org/stable/27743831.

Dennett, D. C. Freedom Evolves. Penguin Books, 2003.

Dodd, J. "Reading Husserl's time-diagrams from 1917/18." Husserl Studies 21, no. 2 (2005): 95-115. https://doi.org/10.1007/s10743-005-4375-5.

Hoefer, C. "Causal determinism." In The Stanford Encyclopedia of Philosophy, edited by E. N. Zalta and U. Nodelman, Spring 2023 edition. Stanford University, 2023. https://plato.stanford.edu/archives/spr2023/entries/determinism-causal/.

Kant, I. "Concerning the ultimate ground of the differentiation of directions in space." In Symmetries in Physics 3, 145-174. Springer Netherlands, 1994.

Maybee, J. E. "Hegel's dialectics." In The Stanford Encyclopedia of Philosophy, edited by E. N. Zalta, Winter 2020 edition. Stanford University, 2020. https://plato.stanford.edu/archives/win2020/entries/hegel-dialectics/.

Nagel, E., and J. R. Newman. Gödel's Proof. Edited by D. R. Hofstadter. New York, NY: NYU Press, 2001.

Nietzsche, F. "The Anti-Christ, Ecce Homo, Twilight of the Idols." In Friedrich Nietzsche: The Anti-Christ, Ecce Homo, Twilight of the Idols, edited by R. J. Hollingdale, 1-199. Cambridge University Press, 1990.

Piccinini, Gualtiero, and Corey Maley. "Computation in Physical Systems." In The Stanford Encyclopedia of Philosophy, edited by Edward N. Zalta, Summer 2021 Edition. Retrieved from https://plato.stanford.edu/archives/sum2021/entries/computation-physicalsystems/.

Rescorla, M. "The Computational Theory of Mind." In The Stanford Encyclopedia of Philosophy, edited by E. N. Zalta, Fall 2020 edition. Stanford University, 2020. https://plato.stanford.edu/archives/fall2020/entries/computational-mind/.

Schneider, S. Artificial You: AI and the Future of Your Mind. Princeton University Press, 2019.

Schneider, S. "What is consciousness? ChatGPT and advanced AI might redefine our answer." NBC News, March 1, 2023. Accessed March 29, 2023. https://www.nbcnews.com/tech/tech-news/chatgpt-ai-consciousness-rcna71777.

Schwartz, S. P. A Brief History of Analytic Philosophy: From Russell to Rawls. Wiley Blackwell, 2012.

Simmons, K. Universality and the Liar: An Essay on Truth and the Diagonal Argument. Cambridge University Press, 2008.

Thyssen, P. The Block Universe: A Philosophical Investigation in Four Dimensions. Doctoral dissertation, KU Leuven, Humanities and Social Sciences Group, Institute of Philosophy, 2020.

Wittgenstein, L. Tractatus Logico-Philosophicus. Project Gutenberg, 1922. https://www.gutenberg.org/ebooks/5740.

Žižek, S. The Most Sublime Hysteric: Hegel with Lacan. John Wiley & Sons, 2014.

---

[i] Similarly, Wittgenstein expresses the view that the "[metaphysical] subject does not belong to the world" (Wittgenstein 1922, 74). It appears that neither Wittgenstein nor compatibilists fully addressed how to distinguish the "metaphysical subject" from the world *by including the subject within the world at the same time.*

[ii] According to Piccinini and Maley (2021, Section 3.4), some scholars support "pancomputationalism," which proposes that the whole universe is computational (Piccinini and Maley 2021, Section 3.4).

[iii] It is assumed that the cognitive agent receives only a "small breadth" of D knowledge that is relevant to the agent. The entirety of the D knowledge would be too immense to be processed by any agent.

[iv] Rescorla states that "[i]n a stochastic model, current state does not dictate a unique next state. (Rescorla 2020, Section 3.0)"

[v] Schneider observes that "current chatbots [such as ChatGPT] use existing human writing to describe their internal state" (Schneider 2023). She suggests that "one way to test if a program is conscious" is to "not give it access to that sort of material and see if it can still describe subjective experience" (Schneider 2023). This idea inherently relies on the concept of emergent processing.

[vi] Supposing that space and time are continuous, Husserl's diagram can provide a useful illustration for how the "retaining" takes place (Dodd 2005). If this "retaining" process cannot be implemented in machines, no amount of machine training may achieve consciousness for an AI.

[vii] The word "diagonalization" has been coined in this paper by drawing inspiration from Cantor's diagonal argument again. As explained in Section 2.0, there is always a new real number that turns out to be not included in a list where every real number is supposedly matched with a corresponding natural number. Note that all the existing real numbers have left their "marks" in the single new real number. Similarly, in the process of diagonalization, the distinct images of the subject and predicate are merged together to create a holistic meaning of the entire sentence.

[viii] One's image of a word arises out of one's subconscious corpus in which the word has formed sophisticated

interconnections with other words. These connections are also established through diagonalization. Then how does one build a corpus from scratch? It starts by matching a particular spoken word with a physical object and on and on. This matching process must also rely on diagonalization.

[ix] The "ontological" argument was influenced by Žižek, who mentions "a retroactive realization that the solution can be found in what we originally saw as the problem" (Žižek 2014, 29).

[x] The concept of an imaginary number contradicts common sense as it appears to be an "intangible" number in the realm of human experience. Nonetheless, its use has been instrumental in establishing quantum mechanics and telecommunications. Although D knowledge is a purely metaphysical concept, it can meaningfully relate to the empirical world.

[xi] It is possible that the (human) experimenter, whose mind is characterized by emergent processing, can coexist in the same non-trivially deterministic world with an AI whose mechanism is characterized by rigid or emergent processing.

[xii] The practical application of quasi-D knowledge as an incomplete representation of D knowledge specific to the AI can be likened to the conventional use of 3.14 as an approximation for the mathematical constant π.

[xiii] D knowledge is "otherworldly but comprehensible" in that it can never be accessed but exists in comprehensible form. In that sense, D knowledge is unlike Kantian things-in-themselves, which are "otherworldly and incomprehensible."

[xiv] This dialectic circle is associated with the diagram of "The Absolute Idea" in Section 1 of Maybee's article on "Hegel's Dialectics" (2020).