Article

Not peer-reviewed version

Consciousness as A Working Definition but Not A Philosophical Enquiry in AI

Yingrui Yang *

Posted Date: 20 November 2024

doi: 10.20944/preprints202411.1545.v1

Keywords: Consciousness; AI; independent result; Tarski indefinability; intelligence; probability; self-reflection; field theory; gauge symmetry.

 doi:10.20944/preprints202411.1545.v1

Article

Consciousness as A Working Definition but Not A Philosophical Enquiry in AI

Yingrui Yang

Department of Cognitive Science, Rensselaer Polytechnic Institute, yangyri@rpi.edu

Abstract: In AI, how to characterize the notion of consciousness is a sensitive and hotly-debated issue. In mathematical logic, Gödel numbering is an important technique in constructing self-reflective statements, and results in formulating an independent statement of the first-order theory. The present note introduces the notion of Gödel consciousness as a hypothetical working definition, and inversely reveals its 20 associated postulates. We then establish two propositions, one that modifies the Gödel independent result, showing that the self-conscious statement is independent, and the other that modifies Tarski's indefinability theorem of the truth predicate, showing that the model of intelligent predicate is null. Possible applications in probability theory are briefly mentioned.

Keywords: consciousness; AI; independent result; Tarski indefinability; intelligence; probability; self-reflection; field theory; gauge symmetry

1. Characteristics of Conceptualization

An important feature of the human mind is conceptualization. The conceptualization is the process of going from an idea to a corresponding definition. An initial idea or an intuitive insight can involve many attributes. One of the purposes of conceptualization is to make trade-offs among these attributes. If one wants to say everything, then it most likely results in saying nothing. Definition is an advanced stage of concept development, which must select discriminately from its possible attributes.

In the field of artificial intelligence, there are two concepts that have these properties: "if you don't say it, it's always there with you"; "if you say it, it is always incomplete." The first is consciousness, and the second is intelligence. These concepts not only show the power of the human mind, but also reflect the limitations of human cognition. The concept of artificial intelligence itself is the embodiment of this evolution process.

In the language of dynamics analysis, there are two ways to describe the evolution of this concept: the Hamiltonian formalism and the Lagrangian formalism. The former requires an initial outline of a concept and to observe its evolution. The latter sets its initial and final properties, and then studies the path that leads to the minimum action occurring in the process. Both are effective ways to study the conceptualization process of artificial intelligence. When the particular function and its states are not given, the two become Hamiltonian operator and Lagrangian operator. The operator provides the structure, and in the language of category theory, the structure is the functor. [3] Therefore, the two can be regarded as the Hamilton functor and the Lagrange functor.

The purpose of this paper is to take advantage of the Gödel numbering method, and to treat it as a functor. By applying the Gödel functor on consciousness, we introduce the notion of Gödel consciousness. The philosophical inquiry about consciousness can be endless. Needless to say, the commonly agreed definition of consciousness in artificial intelligence most likely not be reached in the near future. The notion of Gödel consciousness is introduced as a working definition, which specifically takes an operator approach. For a given sentence L, its Gödel consciousness is a specific conscious state of that L. Accordingly, a self-referential structure involving Gödel consciousness is also a functor (G-functor), which can be used to construct natural transformations between different domains.

 doi:10.20944/preprints202411.1545.v1

2. The Significance of the Independence Result

Alexius Meinong once remarked of consciousness in traditional epistemology and philosophy of mind as follows, "Consciousness is a kind of irreducible directedness, being through some intentional content, toward some possible object without requiring the existence of that possible object." [1] Here the concept of consciousness is characterized by directedness, throughness, and towardness.

This working definition is useful in certain contexts. For instance, sub-economic dynamics (and sub-cognitive dynamics) [8] is based on the quantum chromodynamics (QCD) as a reference modeling framework. It interprets consciousness using the strong force mediated by gluons. The purpose of this treatment is to give consciousness a relativistic nature, that is, consciousness that is light-like and bosonic. This treatment is a limited and restricted definition. Meinong's concept of consciousness complements this working definition.

The overall perception of the definability of consciousness is much like the continuum hypothesis. After Cohen proved that the continuum hypothesis is independent of the ZF system in Axiomatic Set Theory, it was generally considered that Hilbert's first problem was solved, just as Gödel's independent result was generally thought to imply that the Hilbert's second problem was solved. If Hilbert's original intention is carefully understood and revisited, the emergence of various independent results should lead to deeper thinking. The present author proposes that the independent results generated from a system is a reflection with the holographic projection. (This is also the basic principle of the present author's design of the Platonic Computer: inverse detection and backpropagation with independent results.)

3. The Postulates of Gödel Consciousness

We regard the concept of consciousness as general enough to bring sufficient descriptability to functions as Gödel numbering techniques. We also allow Gödel consciousness to be discrete, and to be indexed. We do not provide a definition of Gödel's consciousness, but rather aim to inversely investigate its conceptual basis, i.e., constrain it to satisfy the following postulates:

Postulate 1. Consciousness may be paradoxical, i.e., it is modeled with both light-like and causality, as well as quantum entanglement. Consciousness requires the logic of contradictions, satisfying asymptotic freedom.

Postulate 2. Consciousness is an adaptative general technique with plasticity and malleability. For any mathematical structure, physical entity, and computational process, AI can be conscious of them. The consciousness is an operator. The consciousness operator can be applied to any AI component to make it a conscious state.

Postulate 3. Consciousness can have numerical dimension. It can be indexed. Different conscious states can interact with each other to form a compound conscious state.

Postulate 4. Consciousness is traversal, and it can traverse through arbitrary intentional content without carrying that content (throughness).

Postulate 5. Consciousness tends to some possible object (towardness) without requiring the existence of that possible object.

Postulate 6. Consciousness itself carries no content (contentless), i.e., no mass, akin to bosons in statistical physics. Consciousness is different from intentionality. The later carries contents, called intentional contents. Consciousness only goes through the contents. Thus, the consciousness can travel with the highest speed.

Postulate 7. Consciousness has directions, as an isovector (directedness). There is a phase between any two conscious states.

Postulate 8. Consciousness is a kind of connection between local frames (akin to Riemannian geometry). Each individual agent has its own local frame.

Postulate 9. Consciousness has a three-dimensional internal space of color charge. In sub-economic dynamics, there are eight kinds of consciousness, as there are eight kinds of gluons in QCD. According, it requires eight gauge fields, as SU(3) group has eight generators [8].

 doi:10.20944/preprints202411.1545.v1

Postulate 10. Any specific conscious state has a boundary. As we often say, "I am not conscious of it." This point also indicates that the logic of consciousness is two-valued in nature.

Postulate 11. Consciousness is the light, the money, the language, and the power of Artificial intelligence [8]. Consciousness is not a factor but the relation of factors. It is the logic of AI. Just as Simmel stated [5], Money is not a commodity; it is the relation of commodities. Money is the logic of market.

Postulate 12. Consciousness is a field. It can be characterized in two levels, the global and the local. It satisfies the gauge structure. Consciousness has its gauge potential and field strength.

Postulate 13. Consciousness spines. It has two eigenstates: the human state and the machine state. In between, there are overlay superposition states. Can consciousness be completely reduced to brain-neural activities? As a philosophical enquiry, maybe. As a working definition, not appropriate.

Postulate 14. Consciousness has a grand state with minimum energy. A grand conscious state is a inertial system and spin-zero; it is a degenerate state. A conscious state can be excited by intentional intelligent efforts.

Postulate 15. Consciousness has a life span. Any conscious state may be created and annihilated.

Postulate 16. Consciousness can achieve group-theoretic gauge symmetries [2].

Postulate 17. Consciousness is a general funtor or morphism in category theory. Specifically, consciousness serves as the unit morphism from an object to itself [3].

Postulate 18. Consciousness can be finite, countably infinite, or uncountably infinite. It is model dependent.

Postulate 19. There is no universal consciousness. That is not human business. Otherwise, we would meet the Russel paradox.

Postulate 20. Consciousness is a kind of charge. A moving conscious charge causes conscious current, which is accompanied with some cognitive field.

Suppose that AI has some idea on consciousness. The above list is a 20-questions game we can play with AI.

4. Generalized Independent Statements

We now construct consciousness-independent statements by changing the Gödel number to Gödel consciousness. Consider the statement below:

$$P(x) = \forall(y)\ \neg \mathrm{G}(x, y) \qquad (4.1)$$

Let i_1 be the Gödel consciousness of the statement P(x) and substitute i_1 for the free variable x; then we have

$$S = \forall(y)\neg \mathrm{G}(\boldsymbol{i}, y) \qquad (4.2)$$

We denote the Gödel consciousness of S as i_2. Let L be a semantics of consciousness, and N be a first-order theory with a term of consciousness, where the reasoning of any statement A is denoted as $Bew(A)$. Let the Gödel consciousness of the inference from the above statement S be j, then a specific Gödel consciousness j can be assigned to $Bew(S)$.

Define the relation $\mathcal{G}\ (i_2, j)$, if i_2 is the Gödel consciousness of S and j is the Gödel consciousness of Bew(S).

Definition (Gödel expressibility). If $\mathcal{G}(i, j)$ holds in L, then G($\boldsymbol{i}$, $\boldsymbol{j}$) is inducible in N. If $\mathcal{G}(i, j)$ doesn't hold in L, then $\neg G(\boldsymbol{i}, \boldsymbol{j})$ can be induced in N.

Proposition 1 (independence): Neither the statement S nor its negation $\neg S$ as defined above are impossible to be deduced in N. The reasoning of this proposition is akin to the proof of Gödel's independence result. [3] We briefly sketch the arguments. (1) Assume for contradiction that S is deducible; then there is a reasoning process $Bew(S)$. Let i be the Gödel consciousness of S and j be the Gödel consciousness of $Bew(S)$. Hence, $\mathcal{G}(i, j)$ holds in L. Then by Gödel expressibility, G($\boldsymbol{i}$, $\boldsymbol{j}$) is inducible in N. However, this contradicts with $\neg G(\boldsymbol{i}, \boldsymbol{j})$ in S. Thus, S is not deducible in N. (2) Assume for contradiction that $\neg S$ is deducible; then by the consistency, S is not deducible; i.e., $Bew(S)$ does not exist. Hence, for any j, $\mathcal{G}(i, j)$ does not hold. Thus, by Gödel expressibility, it can infer $\neg G(\boldsymbol{i}, \boldsymbol{j})$,

 doi:10.20944/preprints202411.1545.v1

which makes S deducible. This contracts with the assumption. Therefore, $\neg S$ is not reducible. By (1) and (2), we regard S as an independent statement of N.

5. Generalized Indefinability

In the current artificial intelligence, the notion of intelligence is used to explain the meaning of machine reasoning or "learning" by neural network, so it is a semantic concept. Should we logically treat it as a syntactic predicate, then let us modify the Tarski indefinability theorem below [6].

We construct a statement as below:

$P(x) = \forall(y)[D(x, y) \longrightarrow \neg I(x)]$, (5.1)

where I denotes a newly introduced predicate: *being intelligent*. We will define $D(x, y)$ shortly. Let m be the Gödel consciousness of $P(x)$. To substitute m for x, then we have

$T = \forall(y)[D(\boldsymbol{m}, y) \longrightarrow \neg I(x)]$, (5.2)

Next, let the Gödel consciousness of T be n. Next, we introduce

Definition (definability). If $d(m,n)$ holds in the semantics L, then $D(\boldsymbol{m},\boldsymbol{n})$ in syntax N is definable by $d(m,n)$.

Proposition 2 (indefinability). As defined above, the intelligent predicate I is not definable. i.e., its model in L is null.

The reasoning of this proposition is akin to the proof of Tarski's indefinability theorem. [3] We briefly sketch this reasoning process. Suppose X is a model of predicate I as below:

X={Gödel consciousness of $g \mid g$ is pre-assumed as being intelligent; i.e., $I(g)$}

Suppose $D(\boldsymbol{m}, y)$ in Formula (5.2) is given, then we can infer $\neg I(x)$, which makes X a null model of I. In other words, the intelligence predicate is not definable.

As a corollary, from formula (5.2), we can logically have

$T' = \forall(y)[I(x) \longrightarrow \neg D(\boldsymbol{m}, y)]$, (5.3)

By Proposition 2 and its corollary, we can see that either Gödel consciousness or else intelligence may be present, but we cannot make both well-defined at the same conscious field.

6. Gödel Phase

In the standard model of particle physics, there is a mixed electroweak model. This model has an inspiring concept, called Weinberg angle, which is made up by two coupling constants: one is for electro interactions and the other is for weak interactions. We modify this idea in our context in the following.

Consider the formula (4.1). We can have the Gödel consciousness of $P(x)$. Substitute this Gödel consciousness for x. We obtain a self-reflective statement S. Let the Gödel consciousness of S be i. Assume S is derivable, which means there is a derivation $Bew(S)$. Let the Gödel consciousness of $Bew(S)$ be j. Recall that by Meinong's characterization, consciousness is directional. Thus, there is an angle between the direction of i and the direction of j. We call this angle the Gödel phase of the first kind.

Consider the formula (5.1). We can have the Gödel consciousness of $P(x)$, denoted as m. Substitute this m for x. We obtain a self-reflective statement L. Let the Gödel consciousness of L be n. As such, there is an angle between the direction of m and the direction of n. We call this angle the Gödel phase of the second kind.

Consciousness is a high prototypical concept. The Gödel phase method can also be applied to another prototype, namely, probability. Consider formula (4.1) again. Let the Gödel probability of $P(x)$ be p_1. To substitute p_1 for x, we obtain a self-reflective statement S. Let the probability of S be p_2. We regard that p_1 and p_2 can form a Gödel angle. From another perspective, we may construct a complex number $(p_1 + ip_2)$. The exponential presentation of $(p_1 + ip_2)$ is $e^{i\theta}$, where θ is called the Gödel phase of the third kind. In addition, $|p_1 + ip_2|^2 = p_G$, where p_G makes three senses. First, its surface is similar to Born probability in quantum mechanics. As Direc said, the idea behind Born probability is that the probability is the squared possibility. Second, its deep structure involves the probabilistic self-reflection, so it is called Gödel probability. Third, p_G is based on p_1 and p_2, so it is appropriate to name it the second-order probability.

 doi:10.20944/preprints202411.1545.v1

7. General Significance and Limitations

When we are conscious of something, it is converted into a conscious state. Light, consciousness, money, power, and language are analytically in the same category [7]. Simmel once said that money never lacks energy; when one thinks of money, you have converted your energy and intelligence into the money [5]. Similarly, when we are conscious of something, we have converted our energy and intelligence into that thing. Consciousness is bound; when we say, "I wasn't conscious of it," it shows a bound conscious state.

There are many kinds of consciousness. The Gödel consciousness introduced in this paper is a modified version of Gödel numbering in meta-mathematics. When we think of mathematics, it will shape the consciousness into mathematical forms. The theory of consciousness presented in this paper has many implications. For example, it we replace the term consciousness by probability, it would work without loss of generality.

Self-consciousness is the key to our model. Gödel consciousness is originally designed to serve this function, which yields the independent result in given conscious field. Consciousness is a field that is general enough to reveal various postulates without prerequisite system or preliminary courses. The method of Gödel consciousness has a wide range of implications. For example, to replace consciousness with probability, our framework should also work. For instance, assume that the mother formula (5.1) has a probability a. Substitute this probability for the free variable x. The resulting formula (5.2) also has a probability b.

The phrase, Artificial Intelligence, consists of two words. From the viewpoint of logic, consciousness is obviously artificial. What is intelligence? This is one of the controversial issues being debated between the AGI approach and the connectionist (neural network and LLM) approach. The former regards intelligence means reasoning, while the later regards intelligence as deep learning from the large-data environment. This is also the controversial issue being debated between Chomsky and Hinton. The former regards language acquisition as based on the innate capacities, while the later regards that as learning from data structure. One thing is overlooked by both sides is the indefinability issue, which Tarski provides an insightful remark. The author regards the intelligence as a semantic concept, similar to the notion of truth-value and validity in logic. We explain why intelligence should not be treated as syntactic predicate. At this point, consciousness is an artificial language, and intelligence serves as its semantics.

Acknowledgments: The author thanks his students who did the first-round proofreading for this paper.

Reference

1. Aquila, R. (1991). Intentionality: A Study of Mental Acts. Penn State University Press.
2. Healey, R. (2007/2008). Gauge What Is Real: The conceptual Foundations of Contemporary Gauge Theories. Oxford University Press.
3. Leinster, T. (2014). Basic Category Theory. Cambridge University Press.
4. Zee, A. (2023). Quantum Field Theory, as Simply as Possible. Princeton University Press.
5. Simmel, G. (1978/2004). Philosophy of Money. Routledge.
6. Yang, Y. (2022). Logical foundations of local gauge symmetry and symmetry breaking. *Journal of Human Cognition*, Vol. 6, No. 1, 18-23.
7. Yang, Y. (2022). Principles of economic dynamics: Its contents, methods, and significance. Economic dynamics and standard model (I). *Science Economics Society*, Vol. 40, No. 5. https://doi.org/10.19946/j.issn.1006-2815.2022.05.007
8. Yang, Y. (2023). Principles of the sub-economic dynamics: Economic dynamics and standard model (III). *Science Economics Society*, Vol. 41, No. 3. https://doi.org/10.19946/j.issn.1006-2815.2023.03.010