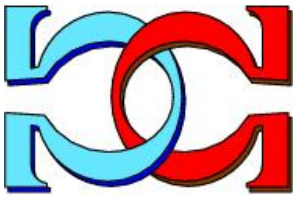
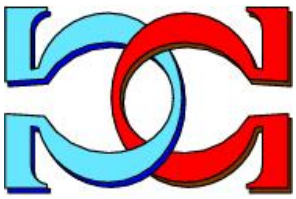
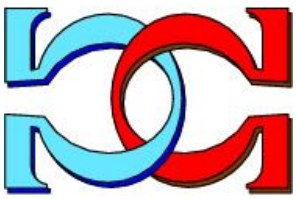
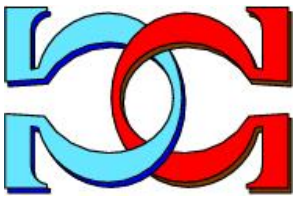


CDMTCS
Research
Report
Series

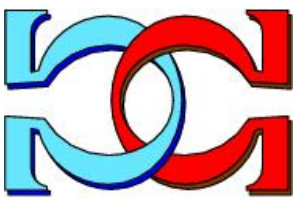




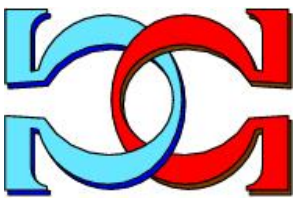
Is the Mind-Body Problem Algorithmically Solvable Under A Computational Formalization



C. S. Calude
University of Auckland



CDMTCS-592
3 September 2026



Centre for Discrete Mathematics and
Theoretical Computer Science

Is the Mind-Body Problem Algorithmically Solvable Under A Computational Formalization?*

Cristian S. Calude
University of Auckland, New Zealand
https://calude.net/cristianscalude/about/

September 17, 2026

Abstract

The Mind-Body Problem – the question of how mental states relate to physical states – is one of the oldest and most persistently unresolved issues in philosophy. In this note, we do not attempt to resolve the philosophical dispute; rather, we isolate a formal core of the problem, model it in the language of computability theory, and show that under this formalization, the associated decision problem is *(Turing) algorithmically unsolvable. In particular, no algorithmic AI system can provide a universally correct decision, today and in the future.* We analyze the limits of the result and situate it against both the classical philosophical debate (dualism vs. physicalism) and Colin McGinn's celebrated "cognitive closure" argument, which, informally, anticipates a structurally similar impossibility claim.

1 Philosophical Background

1.1 The classical Mind-Body Problem

In its broadest form, the Mind-Body Problem requires explaining how humans and other mentally endowed beings come to have, or seem to have, two very different kinds of properties. Humans have, or seem to have, the sort of properties attributed in the physical sciences, such as size, weight, shape, color, and motion through space and time. Still, they also have, or seem to have, mental properties such as perceptual experience, emotional experience, beliefs, desires, agency, and subjecthood or selfhood. The Mind-Body Problem requires an explanation of how beings come to have, or seem to have, these very different characteristics. Historically, the Mind-Body

*Dedicated to Professor A. Atanasiu for his 80th birthday.

Problem consists of two subproblems: the problem of causal interaction, which is primarily a problem of *dualism* [10] – if mind and matter are such different things, how can minds have effects in the material world?–, and the problem of *consciousness* – primarily a problem for physicalism, the thesis that everything is physical (how can consciousness be something material, since it seems so unlike everything else in the physical world?). Dualism, in the Cartesian tradition, holds that mind and body are fundamentally different substances, an idea popularized by R. Descartes, which raises the Mind-Body Problem of how a non-physical mind can interact with a physical body, and which challenges our understanding of consciousness and free will, questioning whether our thoughts truly cause our actions. Physicalism, by contrast, is the view that the mind is not a separate, wholly different kind of thing from the rest of the physical world, but is instead constituted by physical things. Today, dualism remains the second most popular response to the Mind-Body Problem among professional philosophers, after materialism and ahead of idealism and panpsychism, which testifies to the debate's persistent unresolved status.

1.2 McGinn's cognitive closure and mysterianism

A particularly influential and more formally-minded contribution to this debate is Colin McGinn's 1989 paper *Can We Solve the Mind-Body Problem?* [8]. McGinn's position is a form of "new mysterianism": according to McGinn, consciousness is, in and of itself, a fully natural phenomenon, but we humans are just cognitively closed to it, meaning that we cannot, in principle, understand its nature. McGinn's argument proceeds in stages. He first argues for what he calls *perceptual closure*: the argument starts from the thought that nothing we can imagine perceiving in the brain would ever convince us that we have located the intelligible nexus we seek, since no matter what recondite property we could see to be instantiated in the brain, we would always be baffled about how it could give rise to consciousness. McGinn invites the reader to try to conceive of such a perceptible property, expressing confidence that it cannot be done. He then extends this to full *cognitive closure* by arguing that no permissible form of inference from what is perceived can bridge the gap either.

Crucially, McGinn's mysterianism is *not* a form of supernaturalism or an admission of an unbridgeable ontological gap: McGinn argues that certain philosophical problems, such as the Mind-Body Problem, personal identity, the foundations of meaning, knowledge, and free will, may be fundamentally unknowable by humans, a viewpoint he refers to as transcendental naturalism. In a memorable formulation reported in the secondary literature, Colin McGinn argues that the solution to the Mind-Body Problem cannot be grasped, even though the solution is "written in our genes." The unknowability is thus an epistemic limitation of human cognitive architecture,

not a metaphysical peculiarity of the mind-brain link itself – a point McGinn illustrates by analogy with the perceptual and conceptual limitations of non-human animals, who are "capable of perceiving different properties of the world and no species can perceive every property things may instantiate."

This is exactly the structural shape of the result we prove below, transposed into a formal setting: not that no answer to the Mind-Body question exists, but that *under the premise that mental properties are extensional*, no general, uniform algorithm can produce it, even though for every particular case an answer could exist. McGinn's mysterianism can be seen as an informal, cognitive-science-flavored argument of a formal unsolvability result; *the present note explicitly supplies such a formal argument, at the price of a very restrictive (and admittedly reductive) formalization of "body" and "mind.*

2 Formalization

The simplest way to give a mathematical statement about the "Mind–Body Problem" that can be studied by computability theory is to model bodies as computational devices and ask for a decision procedure that, given a body, decides whether it realizes/implements a given mental property. In a natural formalization, bodies are Turing machines, and a "mental property" is a semantic property of the partial function the machine computes.

The formalization consists of the following steps.

1. Fix a standard effective numbering (indexing) $(\varphi_e)_{e\in\mathbb{N}}$ of all Turing machines [2]. The index e denotes the Turing machine computing the partial computable function φ_e.
2. The decision problem for a set $S \subseteq \mathbb{N}$ is: *given* $e \in \mathbb{N}$*, decide whether* $e \in S$*.*
3. A *mental property* is modeled by a non-empty set S of indices $S \subset \mathbb{N}$.

The **main premise**. **A** *mental property* **is represented by an extensional property, that is, by a non-empty set** S **of indices** $S \subset \mathbb{N}$ **with the following property: for all** $e, e' \in S$**,**

$$\varphi_e = \varphi_{e'} \implies (e \in S \iff e' \in S).$$

A mental property S is a *semantic property* as it depends only on the function computed, not on the syntactic code (implementation) of the machine.

3 A Mathematical Result

In this section we prove the following theorem:

Every mental property is algorithmically undecidable.

Let $K = \{e \in \mathbb{N} \mid \varphi_e(e) \text{ halts}\}$, be the diagonal Halting Set, which is well known to be algorithmically unsolvable [2]. Let $\perp$ denote the nowhere-defined partial computable function, i.e. $\perp(x)$ is undefined for every $x \in \mathbb{N}$, and S a mental property.

Because S is nontrivial and extensional, we split the proof into two cases.

Case 1: $\perp \notin S$**.** Choose some partial computable function φ_a such that $a \in S$. Given e, construct effectively a machine $M_f(e)$ which, on input x,

1. simulates $\varphi_e(e)$;
2. if that computation halts, runs $\varphi_a(x)$.

Then we have

$$\varphi_f(e) = \begin{cases} \varphi_a, & e \in K, \\ \perp, & e \notin K. \end{cases}$$

Therefore, $e \in K \iff f(e) \in S$. If S were decidable, K would be decidable, a contradiction.

Case 2: $\perp \in S$**.** Suppose $\perp \in S$, apply the argument to S^c, the complement of S. Since S is nontrivial and extensional, so is S^c, and by hypothesis $\perp \notin S^c$. If S were decidable, so would S^c be, hence a contradiction.

Thus, S is undecidable for every nontrivial extensional property.

4 Conclusion and interpretation

The Mind-Body Problem is a philosophical problem concerning the relationship between thought and consciousness in the human mind and body. The problem has been a central issue in the philosophy of mind since the 17th century.

Any nontrivial semantic property of the behavior of Turing-computable systems is algorithmically undecidable. Interpreting "body" as a description of a physical system that can be modeled by a Turing machine and "mind" as some nontrivial semantic property of the system's input–output/behavioral

function, the decision problem "Does this body realize that mind?" is algorithmically unsolvable under the formalization described above. Furthermore, as *any AI application is an algorithm, the decision problem is algorithmically unsolvable by any AI application, today and in the future.*[1]

It is worth stressing the relation between the formal result and McGinn's informal mysterianism [8, 1]. McGinn claims that human cognitive architecture is *constitutionally* unable to grasp the property P linking brain and consciousness, even though P exists and is entirely natural [8]. Our result is a logically weaker but mathematically provable fact:

No algorithm, run by any conceivable computing agent, human or otherwise, in particular no AI application, can decide, uniformly over all possible "bodies", whether a given nontrivial "mental" property holds, even though for every fixed body the fact of the matter is perfectly determinate. The restriction of the formal argument comes with the extensionality in the **main premise**.

The two facts are not identical: McGinn's is about human cognitive limits specifically, and the formal argument is about the nonexistence of *any* effective procedure at all. However, they share the same dialectical structure: an accurate impossibility result that leaves the underlying "fact of the matter" completely intact.

5 Remarks and caveats

We presented a mathematical theorem about a formal model. Its content depends entirely on the formalization: if one models bodies and minds differently (for example, by restricting the class of bodies, by using syntactic properties of descriptions, or by allowing only trivial properties), the conclusions about algorithmic solvability can change.

The mathematical result can be seen as an instance of Rice's Theorem [9, 2] specialized to the "mind-as-semantic-property-of-a-body" reading of the Mind-Body Problem. The theorem applies to any nontrivial behavioral (semantic) property, such as consciousness, belief, or understanding. Modeled as such, they fall under the unprovability result.

The deepest philosophical objection to the result of this article comes from the distinction between "deciding" and "explaining". The theorem addresses the decidability question, while the traditional Mind-Body Problem asks a question closer to:

> *How or why does physical organization give rise to conscious experience?*

[1] I asked Claude the question *Can you decide the Mind-Body Problem?* and got the following answer: *No – I can't decide the Mind-Body Problem, and I'd argue no one currently can.*

Those are entirely different questions. One is a uniform decision problem, and the other is an explanatory problem. *Undecidability of the first does not entail impossibility of the second.* For example, the absence of an algorithmic decidability for the Mind-Body Problem does not preclude:

- a theory explaining consciousness in humans,
- a mechanistic theory of particular nervous systems,
- a sufficient condition for consciousness, or
- a partial characterization.

In practice, undecidability is a worst-case, absolute impossibility result; however, it does not prevent useful heuristics, probabilistic tests, or partial algorithms for restricted instances [3]. Likewise, McGinn's cognitive closure thesis does not preclude partial, piecemeal empirical progress on the neural correlates of consciousness – it only denies that such progress can ever add up to a full *understanding* of why those correlates yield experience [8].

A further disanalogy is worth noting: McGinn's thesis is contested even within the philosophy of mind. Some commentators have argued that his cognitive-closure argument is internally unstable or "pseudo-mysterian" [6], while others have proposed that the very notion of a problem being permanently unsolvable, yet coherently understood, requires independent justification via the logic of questions and answers [7].

The formal result of this note is immune to such objections in its own (narrow, formal) domain, precisely because it trades the rich, contested notion of "understanding" for the austere, precisely defined notion of algorithmic unsolvability. However, other objections to the present analysis, philosophical or related to modeling, could arise.

The algorithmic unsolvability result seems to be supported by the analysis based on quantum information theory [4], which makes the distinction between the observable mind and the unobservable mind: the first is composed of classical bits of information that form the brain, and the second is derived from measurements of quantum brain unobservables [11].

Acknowledgment

I was attracted to the Mind-Body Problem after reading the book [5] a long time ago. I thank Dr. D. Paul and Dr. C. Stoica for excellent comments that improved the paper.

References

[1] A. Brueckner and E. A. Beroukhim. McGinn on consciousness and the Mind-Body Problem, in: Q. Smith and A. Jokic (eds.),*Consciousness:*

New Philosophical Perspectives, Oxford University Press, 2003, 390–408.

[2] C. S. Calude. *To Halt or not to Halt? That Is the Question*, World Scientific, Singapore, 2024. Russian Translation, DMK Press, Moscow, 2024; Romanian Translation, Cartea Românească Educațional, Iași, 2025.

[3] C. S. Calude and M. Dumitrescu. A statistical anytime algorithm for the Halting Problem, *Computability* 9(2) (2020), 155-166.

[4] D. D. Georgiev. Quantum information theoretic approach to the mind-brain problem, *Progress in Biophysics and Molecular Biology*, 158 (2020), 16–32.

[5] R. Goldstein. *The Mind-Body Problem*, Random House, New York, 1983.

[6] S. Ikgil. Pseudo-mysterianism of Colin McGinn, *MetaZihin: Journal of Artificial Intelligence and Philosophy of Mind.* 6(2), 53-71. `https://doi.org/10.51404/metazihin.1345299`.

[7] U. Kriegel. The new mysterianism and the Thesis of Cognitive Closure, *Acta Analytica*, vol. 18 (2003), 177–191.

[8] C. McGinn. Can we solve the Mind-Body Problem? *Mind*, 98, 391 (1989), 349–366.

[9] H. G. Rice. Classes of recursively enumerable sets and their decision problems, *Transactions of the American Mathematical Society* 74 (2) (1853), 358–366.

[10] H. Robinson and W. Ralph. Dualism. *The Stanford Encyclopedia of Philosophy* (Fall 2026 Edition), E. N. Zalta & U. Nodelman (eds.), `https://plato.stanford.edu/archives/fall2026/entries/dualism/`.

[11] O. C. Stoica. *Is Mind Just a Computation? How Computer Science Dispels the Myths of the Age of AI*, Routledge, London, 2026.