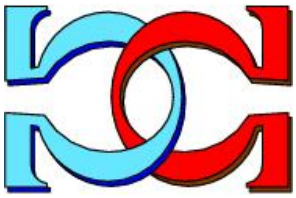
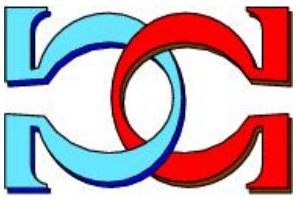
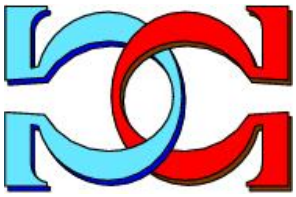
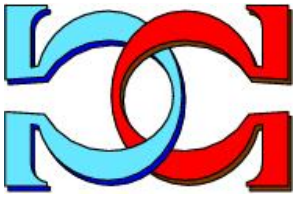


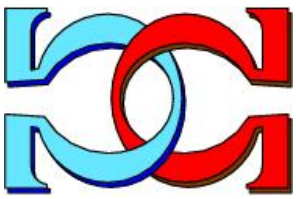




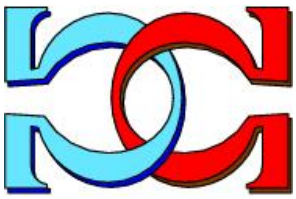
CDMTCS Research Report Series



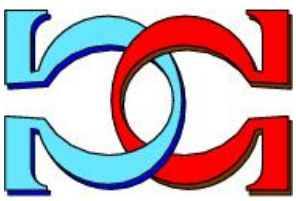
Is the Mind-Body Problem Algorithmically Solvable?



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?*

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

September 12, 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 purely formal core of the problem, model it in the language of computability theory, and show that under a natural formalization, the associated decision problem is *algorithmically unsolvable*, hence it is *unsolvable by any AI application, today and in the future*. We situate this result 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 Problem consists of two subproblems: the problem of causal interaction, which is primarily a problem of *dualism* – if mind and matter are such different things, how can minds have effects in the material world?–, and the

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

problem of *consciousness* – primarily a problem for physicalism (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 no *general, uniform, algorithm* can produce it, even though for every particular case a fact of the matter (an answer) exists. McGinn's mysterianism is thus an informal, cognitive-science-flavored cousin of a formal unsolvability theorem; the present note explicitly supplies the formal cousin, 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 and minds as computational devices and ask for a decision procedure that, given a body (a Turing machine), decides whether it realizes/implements a given mental property. Under a natural formalization – bodies as Turing machines and a "mental property" as a semantic property of the partial function the machine computes – *the decision problem is algorithmically unsolvable*. The following is a short, self-contained proof. For the elements of computability theory, we use the monograph [2].

3 Formal setup

- Fix a standard effective numbering (indexing) $(\varphi_e)_{e\in\mathbb{N}}$ of all Turing machines. The index e denotes the Turing machine computing the partial computable function φ_e.
- A "mental property" is modeled by a set S of indices, $S \subseteq \mathbb{N}$, such that $e \in S$ if the partial function computed by the Turing machine φ_e has the property. The set S is a *semantic property* as it depends only on the function computed, not on the syntactic code of the machine.
- The decision problem for a property S is: given an index e, decide whether $e \in S$.

We will show: *if S is nontrivial (i.e., $S \neq \varnothing$ and $S \neq \mathbb{N}$) and is semantic (depends only on the partial function computed), then the decision problem for S is algorithmically unsolvable.*

4 Key tool: the *s-m-n* Theorem

The *s-m-n* Theorem is a standard result in computability theory: *for every computable two-argument partial function, there is an effective procedure*

that, given the first argument, produces an index of a one-argument machine that hard-codes that first argument. In detail*, there exists a total computable function s such that for a fixed universal function Ψ,*

$$\varphi_{s(e,t)}(x) = \Psi(e, t, x).$$

In words, we can effectively build a machine whose behavior on input x depends on parameters e, t with e, t fixed in its description. We will use the *s-m-n* Theorem in the following concrete form: *given indices for the Turinng machines A and B and an index for a "switch" simulation, we can uniformly produce an index of a machine that first tests some predicate (by simulating another machine) and then behaves like A or B accordingly.*

5 Proof by reduction to the Halting Problem

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]. Since S is nontrivial, pick indices a and b with

$$a \in S \qquad \text{and} \qquad b \notin S.$$

We will construct a total computable function $f : \mathbb{N} \to \mathbb{N}$ such that

- if $e \in K$, then $f(e) \in S$,
- if $e \notin K$, then $f(e) \notin S$.

Construction of $f(e)$. For each e define a machine $M_{f(e)}$ (with index $f(e)$) whose behaviour on input x is:

1. Simulate $\varphi_e(e)$.
2. If that simulation halts, run $\varphi_a(x)$.
3. If that simulation never halts, run $\varphi_b(x)$.

By the *s-m-n* Theorem, from e one can uniformly produce an index $f(e)$ of such a machine. Observe that the partial function computed by $\varphi_{f(e)}$ is exactly φ_a in the case $\varphi_e(e)$ halts, and exactly φ_b otherwise. Therefore:

- if $e \in K$ (i.e., $\varphi_e(e)$ halts), then $\varphi_{f(e)} = \varphi_a$ and hence $f(e) \in S$, and
- if $e \notin K$, then $\varphi_{f(e)} = \varphi_b$ and hence $f(e) \notin S$.

Thus membership of $f(e)$ in S exactly mirrors membership of e in K. If S were decidable, we could decide K by computing $f(e)$ and testing membership in S, contradicting the undecidability of K. Hence S is algorithmically unsolvable. This completes the proof.

6 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 of this formal result to McGinn's informal mysterianism. 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 two facts are not identical: McGinn's is about human cognitive limits specifically, and the *s-m-n* 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.

7 Remarks and caveats

We presented a mathematical theorem about formal models. 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 solvability/decidability can change.

The theorem applies to any nontrivial behavioral (semantic) property: consciousness, belief, understanding, etc., modeled as such, fall under the

[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.*

undecidability result. This is, formally, an instance of Rice's Theorem [2] specialized (via an explicit *s-m-n* construction) to the "mind-as-semantic-property-of-a-body" reading of the Mind-Body Problem.

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 undecidability.

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 [10].

Acknowledgment

I was attracted to the Mind-Body problem after reading the book [5] a long time ago. I thank 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. 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/`.

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