Why Consciousness Should Explain Physical Phenomena: Toward a Testable Theory

Yoshiyuki Ohmura[1*], Yasuo Kuniyoshi[1]

1 The University of Tokyo, Department of Mechano Informatics, Graduate School of Information Science and Technology

Abstract

The reductionist approach commonly employed in scientific methods presupposes that both macro and micro phenomena can be explained by micro-level laws alone. This assumption implies intra-level causal closure, rendering all macro phenomena epiphenomenal. However, the integrative nature of consciousness suggests that it is a macro phenomenon. To ensure scientific testability and reject epiphenomenalism, the reductionist assumption of intra-level causal closure must be rejected. This implies that even neural-level behavior cannot be explained by observable neural-level laws alone. Therefore, a new methodology is necessary to acknowledge the causal efficacy of macro-level phenomena. We model the brain as operating under dual laws at different levels. This model includes hypothetical macro-level psychological laws that are not determined solely by micro-level neural laws, as well as the causal effects from macro- to micro-levels. In this study, we propose a constructive approach that explains both mental and physical phenomena through the interaction between these two sets of laws.

Keywords

Consciousness; non-reductive method; constructive approach; contra-epiphenomenalism; dual laws model

1 Introduction

The hard problem of consciousness refers to the explanatory gap that arises when attempting to explain first-person data using third-person data (Chalmers, 1995). Why is it that physical systems can see, know, feel, or believe something? The mental capacity to represent or think about something is called intentionality (Searl,1983). The hard problem is said to arise not only in explaining the qualitative aspects of consciousness but also in explaining first-person phenomena such as intentionality (Mendelovici & Bourget, 2023).

Intuitively, deciding or believing something seems to depend on the intrinsic causal power of a conscious system. Such intrinsic causal power appears to be self-determinative, like mental or agent causation, and seems different from physical event causation (Chisholm, 1982). To explain intentionality, it is necessary to clarify how the brain autonomously generates integrated intrinsic causal power.

The brain is a system composed of a vast number of neurons and other nerve cells. It is not surprising that individual micro-elements possess intrinsic causal power, such as spontaneous random firing. The problem lies in understanding how the brain can generate integrated intrinsic causal power at the macro level. There is usually only one "agent" in the brain, or at least far fewer than the number of neurons. Furthermore, unlike randomness, this agent seems to operate under its own rules and influence the brain. The brain appears to follow at least two laws: one at the macro agent level and one at the micro neural level. This paper proposes a method to model brain activity as an interaction dynamics between macro- and micro-level laws.

The idea of considering mental phenomena as macro phenomena has a long history (Broad, 1925; Sperry, 1991), and many theories of consciousness seem to regard consciousness as an integrated macro phenomenon (Crick & Koch, 1990; Tononi & Edelman, 1998; Freeman, 1999; Baars, Geld & Kozuma, 2021; Kuhn, 2024). However, if macro mental states are determined by micro neural activity from the bottom-up (Crick & Koch, 1990), then mental states cannot have causal efficacy on physical systems (Kim, 1998), leading to epiphenomenalism. This is because if micro-level causes alone are sufficient to explain system behavior, then macro-level causes suffer from the overdetermination problem. Overdetermination refers to the problem that macro-level causes cannot be redundantly added to phenomena that can be explained by the micro-level laws alone. Unless epiphenomenalism of macro phenomena is rejected, a model of interaction between macro- and micro-level laws cannot be constructed.

Traditionally, macro-level phenomena have been considered reducible to the micro-level phenomena that constitute them (Kim, 1978; Crick & Koch, 1990; Freeman, 1999; Baars, Geld & Kozuma, 2021; Kuhn, 2024). Reductionism is a scientific methodology based on the following two assumptions:

(R1) Macro phenomena can be explained by combinations of micro-level laws.

(R2) Micro phenomena can be explained by micro-level laws alone.

Although bridge laws that describe the relationship between macro-level and micro-level phenomena should be considered in (R1), they will be omitted here as they do not affect the subsequent discussion. Assumption (R2) implies intra-level causal closure. This means that the whole does not alter the behavior of its own parts. Because the relationship between levels in reductionism is a whole-parts relationship, we can only consider influences from the whole beyond micro-level causation. If micro phenomena can be explained solely by micro-level laws, then whole-to-parts influence is unnecessary and excluded from explanations. This leads to overdetermination. Since epiphenomenalism of macro phenomena stems from assumption (R2), rejecting epiphenomenalism requires rejecting (R2). However, past discussions on reductionism have focused on assumption (R1), and assumption (R2) has not been regarded as a premise of reductionism (Nagel, 1961; Kim, 1998; Van Gulick, 2001; Gillett, 2016; Van Riel & Van Gulick, 2025).

The impossibility of causation from the whole to its own parts is a consequence of intra-level causal closure assumed by reductionist scientific methodology and is unrelated to metaphysics. Even if the hypothesis of causal monism—that all causation is physical—is correct, it remains possible for systems to exist in which the whole influences its own parts. Such systems do not hold reductionism assumptions.

In science, it is common to assume different laws for different hierarchical levels—such as sociology, psychology, neuroscience, chemistry, and quantum mechanics (Laughlin, 2005; Gillett, 2016). However, interaction between laws of different levels is generally not assumed. There is also fundamentalism, which holds that macro-level laws should ultimately be explained by more fundamental level laws. Yet, in systems where the whole affects its own parts, neither micro nor macro phenomena can be explained by micro-level laws alone. Additional macro-level laws must be introduced to explain observed phenomena.

To reject intra-level causal closure and provide a model of interaction between macro- and micro-level laws, we introduce macro-level laws that are not determined solely by micro-level laws. Macro-level laws are defined as laws concerning macro-states. Here, macro-states are not derived from micro-states and bridge laws. Rather they are states that determine the influence of the whole on its parts. Furthermore, macro-states are not necessarily observable states as coarse-grained spatiotemporal patterns. If whole-parts causation can be explained, the observability of macro-states is an unnecessary assumption.

We propose that physical systems can be categorized into at least two types. System A allows the influence from the whole to its parts to be ignored, and its behavior can be explained in principle by laws at the micro level alone. For such systems, conventional scientific methodology that does not assume interaction between laws of different levels can be applied. System B, on the other hand, possesses a self-referential feedback control mechanism, where the whole influences its own parts. System A and System B can be explained by differences in embodiment, and both are realizable under common physical laws. If a biological organism acquires a self-referential feedback control mechanism through evolution, its physical mechanism cannot be elucidated by reductionist methods.

For such systems, a theory must be constructed that assumes macro-level laws, micro-level laws, and their interaction dynamics. We consider the conscious brain to be System B because System A cannot explain intrinsic causal power at the macro level.

We name the macro-level laws that cannot be nomologically reduced to micro neural-level laws as psychological laws. Psychological laws are autonomous from neural-level laws and cannot be explained solely by them. In this study, we propose a methodology that implements psychological laws as theoretical constructs within physical systems, simulates their behavior, and tests the theory using observational data. Furthermore, we propose a model of the self-referential feedback control mechanism, which is essential for realizing this approach.

Unlike conventional theories of consciousness that ask, "What physical states give rise to phenomenal consciousness?", we propose a research program that asks:

- "How do macro-level psychological laws contribute to the generation of physical states corresponding to consciousness?"

- "How do the generated physical states corresponding to consciousness subsequently alter physical states?"

In theoretical physics, physical laws are assumed to explain observational data obtained through experiments. Similarly, we aim to explain the construction of consciously accessible representations by assuming hypothetical macro-level psychological laws. By striving for psychological laws with greater explanatory power, scientific progress can be expected. Explaining observable physical states from a theory is a natural methodology, yet even in recent comprehensive reviews of consciousness research (Seth & Bayne, 2022; Kuhn, 2024), similar approaches are not found.

Whereas conventional theories attempt to explain a theory of consciousness from observable data, our approach seeks to explain observable data from a theory of consciousness. Given the similarity of our methodological shift to Kant's epistemological revolution, we believe it is appropriate to name our new program in consciousness science a Copernican Turn in the Science of Consciousness.

2 Relationship with Existing Theories

We believe that any theory of consciousness must address two distinct issues: the unique features of conscious systems and the functional roles of consciousness. We believe that cognitive functions and information processing that do not involve macro-level intrinsic causes cannot explain the necessity of consciousness, because they can perform without consciousness and the possibility of a philosophical zombie cannot be ruled out (Chalmers, 1995). Furthermore, conventional functionalism theories cannot account for the intrinsic nature of *original* intentionality (Searle, 1980; Searle, 1983).

Global Workspace Theory (GWT) links consciousness to broadcasting functions (Baars, 1988; Baars, Geld, & Kozuma, 2021; Dehaene & Changeux, 2011). It is relevant in emphasizing the functional role of consciousness (Baars, 1988). We also believe that theories of consciousness that ignore functions are scientifically inadequate (Cohen & Dennett, 2011; Herzog, Schurger & Doerig, 2022). However, the idea that a function generates consciousness faces the typical problems of functionalism (Block, 1978; Chalmers, 1995). GWT posits that parallel distributed processing is unconscious, while serial, integrated, and coherent information broadcasted to distributed systems is conscious (Baars, 1998; Baars, Geld & Kozuma, 2021). We argue that having a function like broadcasting is not a necessary condition for generating consciousness, unlike what Baars proposed (Baars, 1998). Our claim is that consciousness arises not bottom-up from neural levels, but through interaction with macro-level psychological laws. While this process may relate to the seriality, integration, and coherence of consciousness, how the generated contents are used, stored, or broadcast is independent of the generation of consciousness. We do not assume localized brain regions or pathways like the global workspace are necessary for generating consciousness. The distinction between the generation and utilization of consciousness is a key difference from GWT.

In this study, we refer to the mechanism by which the whole influences its own parts as a self-referential feedback control mechanism. This differs from circular causation (Freeman, 1999), where elements at the same level interact to form causal loops. In such complex systems, global states emerge from micro-level interactions and then organize individual neurons downward (Freeman, 1999). The macro-level psychological laws we propose are not obtained through self-organization from the micro level. Furthermore, Bedau (2008) points out that macro phenomena arising from self-organization fall into epiphenomenalism. To avoid epiphenomenalism, we believe that explanations relying solely on same-level interactions are insufficient and that whole-to-parts influence must be considered.

Chalmers posits that consciousness cannot be derived from physical laws because physics only describes structure and dynamics. He therefore argues for the necessity of a non-reductive approach (Chalmers, 1995). While we agree with the need for non-reductionism, our position is ontological monism: we introduce macro-level psychological laws that are nomologically irreducible to micro-level neural laws, and we model the interaction dynamics between macro- and micro-level laws. Such interaction dynamics cannot be described by psychophysical laws, unlike what Chalmers proposed (Chalmers, 1995). This allows us to explain how consciousness influences physical systems, which is essential for empirical testification. Unlike Chalmers' ontological "extra ingredient" (Chalmers, 1996), our psychological laws are physically implementable within systems and affect the system through a self-referential feedback control mechanism.

Our theory also differs from conventional computationalism. Computationalism posits that mental states are defined by functional relations between inputs, outputs, and internal states, and that these relations are computational (Jackendoff, 1987; Piccinini, 2014; Michael, 2025). While computational theories are nomologically irreducible to physical laws and thus similar to our approach in that respect, they assume a one-to-one correspondence between computational states and physical states. Furthermore, due to overdetermination, computational states cannot affect physical states because both the physical and computational states are completely determined by the behavior at the physical level alone (Kazez, 1994). Because conventional computationalism theories assume intra-level causal closure, they cannot explain intrinsic causal power at the macro level. The problem with computationalism is not the idea of cognition as computation, but its assumption of intra-level causal closure, which excludes causal relations between computational and physical levels.

Integrated Information Theory (IIT) (Tononi, 2008; Tononi & Koch, 2015) is like our theory in that it emphasizes the intrinsic and integrated nature of consciousness. Furthermore, Hoel et al. (2016) address the relationship between IIT and macro causation. However, the macro states they assume are different from ours. They assume that the macro states are coarse-grained states explained by bridge laws in reductionism rather than the whole constituted by the parts. The major difference between IIT and our approach is that IIT assumes integration through bottom-up processes within the recurrent structure of System A, whereas we posit System B, which incorporates a self-referential feedback control mechanism that exerts influence from the whole to its own parts. We believe that a non-reductive dual-laws model is necessary to explain the intrinsic and integrated nature of consciousness.

Sperry considered consciousness to be 'the holistic properties of the organism with causal effects' (Sperry, 1969). Sperry seems to equate macro-level intrinsic causation with consciousness. However, we do not equate the causal power from whole to parts with consciousness itself. We believe this causal power is necessary for generating the contents of consciousness, but how these contents are used and how they influence neural system is a separate issue. To investigate the physical impact of consciousness, it is necessary to clarify how the contents of consciousness affect neural system *after* their generation.

When theories of consciousness emphasize intrinsic causes, they often intersect with the problem of Free Will or agent causation. Sperry appears to equate the problem of consciousness with that of Free Will (Sperry, 1976). Recently, discussions have also linked IIT with the problem of agency (Potter & Mitchell, 2022; Desmond & Huneman, 2022). These connections arise from a shared challenge: macro-level intrinsic causality cannot be explained by physical event causation. We argue that this challenge stems not from the metaphysical nature of macro-level intrinsic causality, but from the assumption of intra-level causal closure in reductionism, which excludes macro-level intrinsic causality. We propose that macro-level intrinsic causality can be explained, despite the assumption of causal monism, by a self-referential feedback control mechanism.

Non-reductive physicalism holds that consciousness is a product of the brain, but that mental states and properties are non-reductively distinct from physical states and properties (Macdonald & Macdonald, 2019; Ellis, 2019). According to Kuhn (2024), this view resembles property dualism, since it treats mental states as ontologically distinct from physical ones. The core mechanism of non-reductive physicalism is emergence—the idea that higher-level emergent properties cannot be predicted even with complete knowledge of the underlying level. Moreover, non-reductive physicalism often assumes that emergent mental properties exert downward causal influence on physical systems. This notion of downward causation is frequently used to explain mental causation (Sperry, 1991) and agent causation (O'Conner, 2009; Steward, 2012; Steward, 2017; Mayr, 2018).

While our hypothesis shares many similarities with non-reductive physicalism, it differs in the following key respects: Non-reductive physicalism assumes that emergent mental properties are ontologically irreducible to physical properties. In contrast, we propose nomologically irreducible macro-level psychological laws that are physically implementable within systems. Emergentism interprets macro-level causal power as something that arises from micro-level phenomena and thus treats ontological irreducibility as a central issue. However, we argue that the need for causal emergence arises only because the influence from macro- to micro-levels has been excluded by the assumption of intra-level causal closure. If macro-level influence is considered from the outset, causal emergence becomes an unnecessary concept.

Scheffel's *Emergent Will* resembles our macro-level psychological laws (Scheffel, 2025). Scheffel regards *Emergent Will* as a law that cannot be reduced to lower-level physical laws and proposes it as a new definition of Free Will. To justify the validity of such *Emergent Will*, Scheffel appeals to the emergence of psychological properties. However, if intra-level causal closure is rejected, Scheffel's proposed concept of Free Will can be maintained, and the assumption of mysterious causal emergence becomes unnecessary. Similarly, List employs the emergence of intentional agency from lower-level physical phenomena in his discussion of Free Will (List, 2019), whereas the psychological laws we assume are not emergent from physical systems but merely mechanisms implemented within them.

Gillett (2016), like us, proposes a non-reductive theory that assumes influence from macro- to micro-levels. However, his concept of machretic determination differs from our notion of causal influence. Machretic determination refers to a mutual determination relationship between macro- and micro-levels, rather than a directional causal relationship. In contrast, our proposed self-referential feedback control mechanism allows for macro-level changes to precede and influence micro-level changes, which we interpret as a form of causation.

We argue that it is not necessary to treat mental or agent causation as ontologically basic properties. Instead, they can be explained through physical mechanisms that allow the whole to influence its own parts. In debates about Free Will, the intuitive resistance to determinism may stem from the reductionist assumption that agent-level influence does not exist at the micro-level. To explain agent-level influence, the theory must presuppose macro-level influence from the outset. In this case, micro-level phenomena cannot be explained without macro-level laws.

Piaget's developmental theory is also relevant to our framework. Piaget argued that structural changes in intelligence arise through two mechanisms possessed by the agent:

assimilation and accommodation (Piaget, 1950). Assimilation is the active process of incorporating physical stimuli, such as sensory input, into the organism's internal structure, while accommodation refers to the process of modifying that internal structure itself.

Piaget presupposed the agent's mental capacity characterized by these two processes, and he appeared to believe that the mechanisms of biological development cannot be reduced to physical laws. His epistemology was influenced by Kant, but he criticized Kant's static view of *a priori* structures and experimentally demonstrated that logical and mathematical structures evolve during development. Our concept of psychological laws is inspired by Piaget's idea of structural expansion. We interpret assimilation as corresponding to the self-referential feedback control mechanism, and accommodation as corresponding to the psychological laws. In this way, Piaget's developmental framework provides a conceptual foundation for understanding how internal mental structures evolve and interact with physical systems in our model.

Our view can be considered a variant of Davidson's Anomalous Monism (Davidson, 1970; Davidson,1990). Anomalous Monism is based on the following three premises:

- (AM1) Mental events have causal power and act on physical events.
- (AM2) All causal relations are governed by physical laws (=causal monism).
- (AM3) There are no strict laws that determine mental events from physical events in a bottom-up manner.

Premise (AM3) implies that consciousness cannot be explained solely from physical states, which contrasts with the assumptions of conventional reductive consciousness science. Davidson characterizes Anomalous Monism as a monism that denies the identity theory between mental and physical states (Davidson, 1970). He seems to regard the identity theory as a source of epiphenomenalism and further argues that the mind is nomologically irreducible to physical laws, thereby highlighting the limitations of reductionist methodology.

However, Davidson's discussion is situated within the context of analytic philosophy and does not emphasize that intra-level causal closure is a methodological assumption of reductionism. This makes the argument more complex. We claim that premise (AM3) is a consequence of rejecting intra-level causal closure. Because macro-level laws and micro-level laws interact, the correspondence between macro mental states and micro physical states is not determined bottom-up from the micro physical states.

To investigate consciousness and the mind, we must assume that observable neural and muscular activity in living organisms is the result of influence from consciousness and mental states. While assuming that observed data are only affected by observable-level phenomena simplifies analysis, there is no justification for assuming that all systems satisfy this condition. Not all physical systems can be understood through reductive approaches—only those that fall under System A.

In a non-reductive approach that rejects intra-level causal closure, we face the challenge that macro-level laws must be understood to interpret observational data. Simply combining micro-level knowledge is insufficient for understanding phenomena. Furthermore, if observed micro-level phenomena are influenced by unobservable macro-level laws, we must address the problem of how to disentangle macro- and micro-level laws. Because this approach requires distinct laws at both macro- and micro-levels, and considers their interaction dynamics, theory construction becomes inherently complex.

Introducing multiple laws into a scientific theory may be criticized as a violation of Occam's razor. However, current science already studies different laws independently at each hierarchical level. Therefore, assuming multiple laws is not inherently problematic. In our framework, the interaction between macro-level psychological laws and micro-level neural laws is essential to account for macro-level intrinsic causal power. Thus, the complexity introduced by multiple laws is justified by the need to explain phenomena that cannot be captured by a single-level theory.

3 Causal efficacy of consciousness

Epiphenomenalism is the view that mental events depend on physical events but have no effect on them. Any scientific theory of consciousness must reject epiphenomenalism. This is because if consciousness does not affect physical systems, then it cannot be empirically tested through observational data. A theory of consciousness must commit to explaining how consciousness influences physical systems.

Mental phenomena including consciousness are all thought to systematically co-vary with changes at the neural level, such that no change in the former occurs without a change in the latter. In philosophy, this dependency is called a supervenient–subvenient relationship. For example, the sensation of pain is believed to arise only with activity in peripheral nerves or the cerebrum. Changes in supervenient phenomena do not occur without changes in subvenient phenomena. This does not mean that supervenient phenomena are caused by subvenient phenomena. Structural relationships like whole and parts also fall under supervenient–subvenient relationships. These relationships are more about determination than causation.

Kim's exclusion argument (Kim, 1998) is often cited to deny supervenient-to-subvenient causation. The flaw in Kim's argument lies in the assumption that the behavior of the subvenient level is determined solely by subvenient-level laws (Kim, 1978; Kim 1998), which implicitly commits to intra-level causal closure. Therefore, Kim's critique of mental causation (Kim, 1990) is not valid. Kim argues that the causal efficacy from supervenient macro to subvenient micro is overdetermination, but this claim amounts to begging the question unless it demonstrates that intra-level causal closure can be derived from the premise of the completeness of physical causation (causal monism).

Mayr (2018), like us, correctly identifies the flaws in Kim's influential exclusion argument. The completeness of physical causation does not necessarily entail causal closure of lower levels. Hoel et al. (2016) also points out that micro causal closure is assumption of reductionism. However, it is important to distinguish between micro causal closure, which denies the causal efficacy of the macro states, and intra-level causal closure, which denies the causal efficacy of the macro level laws. Without macro-level laws that is not determined solely by micro-level laws, intra-level causal closure cannot be rejected.

Discussions on epiphenomenalism often focus on whether subjective experiences, such as qualia, exert causal influence on the physical states that generate them (Jackendoff, 1987; Kim, 1998; Pauen, 2006; Bradley, 2011). This narrower version of epiphenomenalism is known as qualia epiphenomenalism (Jackson, 1982).

This issue should not be interpreted as a question of whether phenomenal consciousness has causal efficacy, distinct from access consciousness (Block, 1995). As a scientific hypothesis, it has been pointed out that we should not assume the existence of two types of consciousness—phenomenal and access (Cohen & Dennett, 2011). Moreover, contradictions between this distinction and scientific findings have also been noted (Naccache, 2018). If we define access consciousness as the kind of consciousness that has physical influence, then any consciousness that is not accessed becomes epiphenomenal, rendering the distinction scientifically meaningless.

We consider access consciousness not as a type of consciousness, but as a utilization of consciousness. Furthermore, we argue that the generation and utilization of consciousness are distinct problems. If we focus only on the cognitive processes that utilize the contents of consciousness, it seems that input-dependent processes that determine output do not require intentionality and cannot explain the generation of consciousness (Searl, 1980). We do not believe that the material used to realize the function—whether protein or silicon—is decisive. Rather, we interpret this as evidence of the independence between the generation and utilization of consciousness. We believe that there are two types of issues, not two types of consciousness: generation and utilization.

Assuming causation is purely physical makes it unclear how the qualitative nature of consciousness can exert direct causal influence on physics. If consciousness is a macro property that supervenes on neural states, then there must be neural activity corresponding to conscious experience. Thus, the problem of qualia epiphenomenalism can be interpreted as a question of how the neural activity corresponding to experience exerts causal influence.

Neural activity corresponding to conscious experience cannot be explained by correlation alone. For example, visual experience is highly correlated with retinal information, yet neural firing in the retina cannot be regarded as a sufficient condition for generating consciousness. Our assumptions are as follows:

1. Generating conscious experience requires the influence of macro-level intrinsic causal power.

2. The neural activity corresponding to conscious experience has a distinct causal efficacy, equivalent to causal efficacy of conscious experience.

We hypothesize that conscious experience accompanies neural activity that represents conscious content. Although the hypothesis that conscious experience accompanies neural activity produced under the influence of intrinsic causal power is difficult to test scientifically, we believe this limitation stems from the subjective and unobservable nature of conscious experience.

We argue that intrinsic causal power from the whole to its own parts is involved in the generation of conscious contents. We cannot equate this whole-to-parts causal power with the causal efficacy of consciousness itself. Unless we clarify how the brain uses the generated consciousness, we cannot properly address the causal efficacy of consciousness. To solve the problem of epiphenomenalism, the theory of consciousness must explain how generated consciousness is utilized by the brain.

We assume two possibilities:

1. The neural states corresponding to conscious contents influence micro-level neural states.

2. They influence macro-level psychological states.

We propose the second. This is because if consciousness is a macro phenomenon, it is more natural to assume that it affects macro-level states. However, we cannot rule out the possibility that both exist.

Psychological laws can be understood as the dynamics of macro psychological states that determine how the whole influences its own parts. In this case, it is conceivable that once conscious contents are generated, the psychological states are subsequently affected by them. If conscious contents influence the psychological states, and the psychological states then affect the neural states via the self-referential feedback control mechanism, it becomes possible to scientifically test the influence of consciousness.

4 Self-referential Feedback Control Mechanism

Consciousness and mental phenomena are explained by intrinsic causes within physical systems (Nagal, 1970; Searle, 1984; Chalmers, 1995), since their presence is not determined by external influences alone. Therefore, to explain consciousness and the mind, it is necessary to clarify how intrinsic causes generate within the physical systems.

Existing reductive theories of consciousness are incompatible with intrinsic nature of consciousness (Schwitzgebel, 2015; Mørch, 2019). If a conscious system does not overlap with other conscious systems, reductive theories require an appeal to extrinsic maximality over supersets to explain this, which contradicts the intrinsic nature of consciousness.

Intuitively, a social system composed of individuals is not conscious. This suggests that the assumption of non-overlap seems to be true. Thus, if consciousness is a macro phenomenon with intrinsic causal efficacy, the assumptions of reductionism should be rejected.

Because the whole and its constituent parts share the same physical entities, the influence from the whole to its parts is internal rather than external and thus can be characterized as intrinsic. The self-referential feedback control mechanism should explain both the influence from the whole to its parts and the intrinsic nature of this influence.

The relationship between cell groups and their constituent individual cells is a whole-parts relationship. In the brain, such cell groups form neural networks. A single neuron can be viewed as a function that determines output signals based on input signals from synapses. Similarly, neural networks composed of these neurons can be viewed as functions. We can see that both the macro layer of cell groups and the micro layer of cells consist of mathematical functions.

We assume that a self-referential feedback mechanism is modular and consists of macro mathematical functions. The relationship between macro mathematical functions and their constituent neurons and synapses is structurally whole-parts relationship. Given the modular and hierarchical structure of biological systems such as cells and organs, it is reasonable to assume that there are multiple macro mathematical functions in the brain. Because changes in macro mathematical functions cannot occur without changes in constituent neurons and synapses, macro mathematical functions supervene on them.

We define an inter-level feedback control as negative feedback control at the micro level, which reduces feedback errors defined at the macro level. Because the behavior of macro mathematical functions is determined by the neurons and synapses that constitute them, it can be controlled through negative feedback control achieved by synaptic adjustments at the micro level.

We assume that the self-referential feedback control mechanism is an inter-level feedback control system that uses intrinsic feedback errors at the macro level. If we define 'intrinsic features of things' as 'how the things are in themselves rather than how they interconnect with their surroundings' (Braddon-Mitchell & Jackson, 1996), then the relationships between cell groups can be considered intrinsic to the mechanism. Since macro-level feedback errors are intrinsic in the self-referential feedback control mechanism, they must be defined by the relationships between cell groups. This means that the equations describing feedback errors are based on the relationships between macro mathematical functions within the mechanism, independent of extrinsic sensory inputs.

We believe that self-referential feedback control is generally achieved by inter-level feedback control using feedback errors based on the relationships between macro-level cell groups. In this mechanism, feedback errors are transmitted to the constituent cells at the micro level. Then, these cells behave in a way that reduces the errors.

To define intrinsic feedback errors, we propose using algebraic theory to describe the relationships between macro mathematical functions. By applying well-defined algebraic constraints, the algebraic structure of macro mathematical functions can be constrained. For example, an algebraic constraint such as commutativity (AB = BA) can be defined between two functions A and B. Alternatively, group axioms may be applied to a set of macro mathematical functions. However, tautologies such as A = A are inappropriate as constraints. Thus, by designing algebraic structures among macro mathematical functions, various algebraic equations can be used to define intrinsic feedback errors.

In this framework, psychological laws can be defined as laws that alter the algebraic equations used to define feedback errors. At the macro level, it is assumed that the algebraic equations defining feedback errors change over time. We regard these algebraic equations as macro psychological states. Macro psychological states are not coarse-grained states that are explained by bridge laws but equation that is satisfied by feedback control. However, it is unclear whether psychological states are physically observable. To implement psychological laws, dedicated neural circuits are required at the macro level to modify algebraic equations, compute feedback errors, and transmit them to the micro level—these circuits govern psychological laws.

Although psychological laws must consider the convergence time required for the micro level to reduce feedback errors, they can fundamentally possess dynamics independent of the micro-level law. If the micro level of a neural system observes intrinsic feedback errors at the macro level and adjusts synapses accordingly, then the system can *as much as possible* follow macro-level laws without violating micro-level neural laws. In this way, psychological laws can influence the behavior of micro-level neural systems. Thus, the behavior of these systems cannot be explained by micro-level neural laws alone.

5 Dual-laws model

We propose a system model with distinct laws at the macro- and micro- levels (Fig.1). The system consists of the self-referential feedback control mechanism, macro-level neural circuits B for modifying the algebraic equations that define feedback errors, and other micro-level neural circuits A. In the self-referential feedback control mechanism, the whole does not refer to the entire brain but rather to the relative relationship between the whole and its own parts within the mechanism. Although many neurons exist in the system beyond those involved in the self-referential feedback control mechanism, they may be indirectly influenced by macro-level effects through interactions.

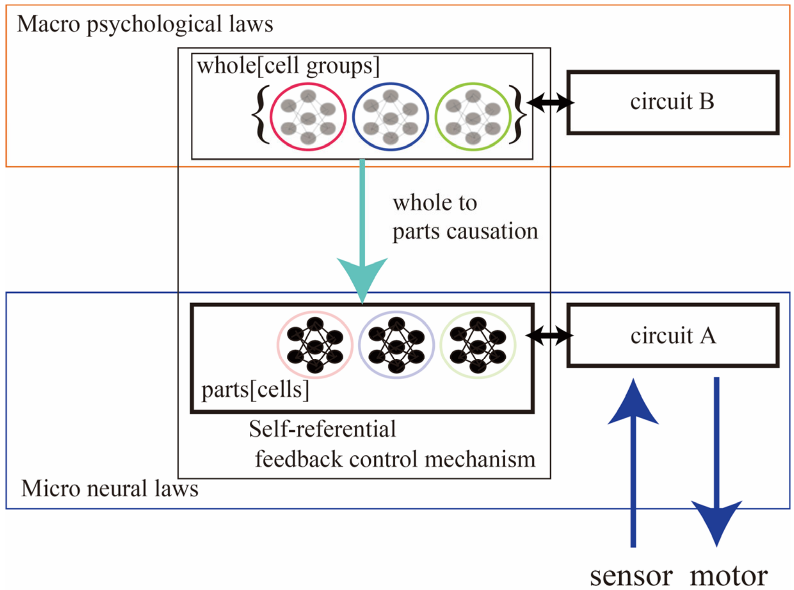

Figure 1. Dual-laws model: The system consists of self-referential feedback control mechanism, circuit B for modifying the algebraic equations used to define feedback errors at the psychological level, and other micro-level circuits A. In self-referential feedback mechanism, whole is a set of macro mathematical functions, and the constituent parts are neurons and synapses within the functions. Bold-line rectangles represent distinct physical entities. Although psychological laws are physically realizable, the micro neural level is influenced not only by micro-level neural laws but also by macro-level psychological laws. Therefore, the behavior of neural states cannot be reductively explained by micro-level neural laws alone.

In the self-referential feedback control mechanism, the whole and its parts share physical entities such as neurons and synapses, so there is no separate region corresponding to the whole beyond the neurons corresponding to the parts.

Macro-level neural circuit B is the core mechanism that determines psychological laws, computing feedback errors at the macro psychological level and transmitting them to the micro neural level. Neural circuit B can be considered a kind of serial computational device (algorithm). Although its entities are neural, it is classified under the psychological layer due to its involvement in the whole level of the self-referential feedback control mechanism. If neural circuit B itself is computation, it does not directly exert causal efficacy on the neurons that constitute it. The influence of macro psychological layer on micro neural layer occurs via the self-referential feedback control mechanism. While interaction between micro-level neural circuit A and macro-level neural circuit B is theoretically possible, it is not considered here.

The micro-level neural layer includes numerous neural circuits for sensory, motor, and life-support functions beyond reducing intrinsic macro feedback errors in the self-referential feedback mechanism. Therefore, through interaction with micro-level neural circuit A, the self-referential feedback control mechanism can receive sensory information about the body and environment. The activity of the micro-level neural circuit A that is not influenced by macro-level intrinsic causes is unlikely to be involved in the generation of conscious experience. Conventional theories of consciousness assumed the existence of neural circuit A alone and did not consider the influence of macro psychological laws.

Several caveats must be noted. The number of macro-level mathematical functions in the self-referential feedback control mechanism is relatively small compared to the neurons and synapses in the system, resulting in fewer degrees of freedom at the macro level. Therefore, macro-level cannot fully determine the micro-level.

In the self-referential feedback control mechanism, the constituent parts operate in accordance with macro-level psychological states, but this mechanism depends on the micro-level appropriately reducing feedback errors. The micro-level can turn off the negative feedback control to reduce macro feedback error, and the decision lies with the micro-level. Unless the micro-level reduces macro feedback errors, the system will not be affected by the macro-level. This represents a conditional causal relationship, differing from conventional causality. Thus, even if macro psychological laws are involved in the generation of consciousness, their activation can be determined at the micro neural level.

This implies a situation different from the micro-level simply following macro-level laws. Conflicts may arise between micro-level neural laws and macro-level psychological laws. Since macro-level psychological laws and micro-level neural laws have distinct dynamics and their interaction determines system behavior, one cannot be reduced to the other.

6 Copernican Turn in the Science of Consciousness

Traditional reductionist theories of consciousness aim to explain consciousness unidirectionally through physical states. However, when consciousness is regarded as a macro-level phenomenon, reductionist approaches tend to fall into epiphenomenalism. A testable science of consciousness must presuppose that consciousness exerts influence on physical systems.

Our methodology posits that macro-level psychological laws form consciously accessible representations by altering the physical systems that implement them, and that the contents of consciousness influence psychological states. Whereas conventional theories attempt to explain a theory of consciousness from observable brain data, our approach seeks to explain observable data in phenomenology, cognitive science, psychology, and neuroscience from a theory of consciousness. This reversal of explanatory direction resembles the Copernican turn in Kantian epistemology.

Kant used the term "Copernican turn" to characterize his epistemology. While traditional epistemology assumes that cognition is determined by objects, Kant argued that objects are determined by cognition. The hallmark of Kant's epistemology lies in recognizing the active role of the agent in cognition. Similarly, we acknowledge corresponding macro-level psychological laws. However, our methodology emphasizes the dynamics of interaction between macro-level and micro-level laws, differing from the notion that agents unilaterally determine cognition. Unlike many traditional theories of consciousness that attempt to explain consciousness bottom-up from neural phenomena, our approach considers not only sensory information from the external world but also influences from intrinsic causes—another point of similarity with Kant's Copernican turn.

The psychological laws we assume may vary among individuals. In this sense, as Davidson (1970) argues, psychological laws may not be strict, unlike physical laws. Nevertheless, it is not necessary to assume that all psychological laws are not strict; we may posit common principles underlying the generation of consciousness. Furthermore, we may assume common principles regarding the functional significance of the contents of consciousness. The science of consciousness must explain the generation of diverse conscious contents based on such common principles.

We model macro-level psychological laws as dynamics of algebraic structural constraints. These algebraic structures are assumed to constrain the relationships among multiple neural networks that learn inter-sensory relationships. Although the definition of constraints does not depend on extrinsic sensory input, neural network activity is determined by both algebraic structural constraints and extrinsic sensory input. Since neural networks are functions, their inputs and outputs can be associated with extrinsic inputs. If consciousness has functional significance for organisms, we must clarify how consciousness relates to information from the environment. Intrinsic properties must not imply disconnection from the environment. In this sense, algebraic structural constraints among neural networks that learn inter-sensory relationships may potentially explain the interaction between intrinsic constraints and extrinsic information.

Traditional scientific theories have studied law independently at each hierarchical level. In contrast, we assume that there are distinct laws at both macro- and micro-levels, and that micro-level observations result from their interactions. At the macro-level, to realize a self-referential feedback control mechanism, we must define intrinsic feedback error at the macro-level. This requirement leads us to express macro-level psychological laws as changes in algebraic equations defining feedback error. At the micro-level, we assume that at least part of the system observes macro-level feedback error and works to reduce it. If the micro level merely follows macro-level laws, then perhaps only macro-level laws are needed to explain the system behavior. However, the brain is not that simple. In addition to conscious serial processing, unconscious parallel processing is well known, and functions such as emotion, reflexes, and homeostasis must also be realized simultaneously. Therefore, micro-level laws are likely governed by their own principles, not solely by the realization of self-referential feedback control mechanisms. Reproducing micro-level observations under the assumption of complex interaction dynamics between levels is considered highly challenging.

Because it is difficult to disentangle macro- and micro-level laws from observed data, a constructive approach is essential—implementing both laws in simulations or robotic systems and examining their behavior. The self-referential feedback control mechanism can be simulated using artificial neural networks on a computer. Macro mathematical functions can be modeled using artificial neural networks, and the micro-level can be considered as the neurons and synapses forming these networks. If autonomous macro-level psychological laws can alter the behavior of neural systems, it becomes possible to experimentally investigate which psychological laws (macro-level intrinsic causes) are involved in the generation of conscious contents. Furthermore, by introducing the hypothesis that the content of generated consciousness influences psychological states, we can predict how consciousness affects physical states.

Since both macro- and micro-level laws cannot be directly derived from observational data, exploratory approaches may also be necessary. For example, machine learning may be a promising method for exploring combinations of macro- and micro-level laws that explain system behavior. The crucial point is that constructing a science of consciousness may not be feasible through theories based solely on the neural level. If consciousness is generated by the influence of macro-level intrinsic causality, the system behavior cannot be explained by observable micro-level laws alone, we must clarify how unobservable macro-level laws influence physical systems.

7 Conclusion

The main claims of this paper can be summarized as follows:

1. If a physical system contains a special mechanism whereby the whole influences the behavior of its own parts, it is impossible to construct a bottom-up theory that explains the system's behavior based solely on micro-level observations.
2. If consciousness/mind is a macro phenomenon within the brain and it influences neural activity, then a bottom-up theory explaining brain behavior from neural observations alone cannot be constructed.
3. To study systems that cannot be reductively explained from the micro-level, a new scientific methodology is required—one that adopts a constructive approach based on the interaction dynamics between macro- and micro-level laws.

This claim is independent of metaphysical concepts such as physical causal completeness. Reductionism is a methodological approach that presupposes intra-level causal closure and thus cannot be applied to phenomena where this condition does not hold. The intuition that the causal efficacy of consciousness cannot be explained by neural activity arises because existing theories are not constructed on the premise that observed neural activity is influenced by macro-level laws. If consciousness and mind are considered macro-level phenomena of the brain, then scientific theories must avoid epiphenomenalism, and therefore cannot employ reductive methods that explain macro phenomena bottom-up from the micro-level alone.

Our non-reductive position is causal monism, and it is compatible with physicalism and computationalism. Traditional forms of physicalism and computationalism have assumed intra-level causal closure, which rendered macro-level phenomena epiphenomenal. A scientific theory of consciousness must reject this assumption.

This paper proposes a method for studying non-reductive systems governed by both macro- and micro-level laws. The difficulty of this issue lies in the fact that the presence of a self-referential feedback control mechanism in the system invalidates reductive scientific methodology, regardless of metaphysical considerations. One source of confusion in traditional discussions of mental causation and agent causation is the widespread misunderstanding of the methodological assumption of intra-level causal closure as a physical law. Furthermore, the inability to causal efficacy of consciousness, based on the assumption of intra-level causal closure, has contributed to confusion in consciousness science.

We defined psychological laws as being characterized by self-referential feedback control mechanisms and physical implementablility. We believe that the physical implementability of the "mind" is essential for providing a physical explanation of its causal efficacy. This characterization of "mind" is minimal and incomplete, and further rigorous definition is required. To scientifically elucidate what "mind" is, it is necessary to empirically demonstrate that psychological laws contribute to the constitution of mental and cognitive processes.

For empirical purposes, we proposed a self-referential feedback control mechanism that uses macro-level algebraic structural constraints as feedback errors. However, this is merely one example, and it cannot be ruled out that the "mind" implemented in the brain may utilize a different self-referential feedback control mechanism. Our proposal is a practical model aimed at realizing a scientific program that elucidates the nature of "mind" by explaining the intrinsic causes of mental functions.

We believe that a theory of consciousness must possess universality. While some philosophers claim that consciousness arises only in biological systems, the basis for such claims is unclear. Theories that associate specific physical states of the biological brain with consciousness cannot address the consciousness of artificial entities that lack brains. Our theory assumes that the neural level consists of neurons modeled as functions with plasticity, which can be artificially reproduced, thereby satisfying the condition of universality.

Our non-reductive methodology requires a constructive approach to clarify the functional roles of consciousness. Although future experiments are essential to determine which functions are unique to consciousness, some predictions can be made. First, the formation of beliefs based on experience is conceivable. We form various beliefs based on experience. Once such beliefs are verbalized and become objective data, they can be taught to machine learning systems. However, the data used in machine learning are considered imitative and differ from what Searle (1980) calls *original* intentionality. Second, the realization of voluntary movement. Current machine systems are essentially input-output machines that determine output based on input. To be called voluntary systems, they must at least be capable of acting based on original beliefs formed by the system itself. Whether such functional differences produce unique characteristics in behavior and cognition requires further research, but many existing theories of consciousness have not sufficiently addressed what differences arise in physical systems when consciousness is present.

In scientific theories that assume multiple laws, it is important to note that both the hypothesis that macro-level laws determine micro behavior in a whole-down manner and the hypothesis that micro-level laws determine macro behavior in a bottom-up manner are rejected. It is necessary to consider that the interaction dynamics between both levels of laws play a crucial role in theoretical construction.

The proposed constructive methodology is an extremely challenging endeavor, given the diversity of human conscious experiences. However, since the theory can be evaluated by the explanatory power of psychological laws as theoretical constructs, progressive advancement can be expected.

Appendix

Our model differs from the interaction model of two dynamics with different periodicities. Such dynamics can be described by the neural-level laws alone. We assume that the reductive neural level laws can be described by all neural states, x, as follows:

$$x_{t+1} = f(x_t).$$

If the macro states are coarse-grained states in reductionism, then the macro states X_t can be described by the bridge laws from the micro to macro as follows:

$$X_t = b(x_t).$$

If the dynamics of neural states are affected by the macro states, the dynamics can be described as follows:

$$x_{t+1} = f'(x_t, X_t) = f'\big(x_t, b(x_t)\big) = f(x_t).$$

This equation implies inter-level causal closure. Therefore, we do not assume that the macro states are coarse-grained states which can be described by the bridge laws from the micro to macro. In our theory, the macro states are functions to calculate feedback error in the referential feedback control mechanism.

In our theory, micro neural level laws can be described as follows:

$$x_{t+1}^{sr}, x_{t+1}^{a} \ = f(x_t^{sr}, x_t^{a}, e_t).$$

x^{sr} are neural states in the self-referential feedback mechanism, x^{a} are neural states in the circuit A. e are feedback errors from the whole level in the self-referential feedback mechanism.

Macro psychological laws can be described as follows:

$$X_{T+1} = g(X_t).$$

Here, the macro states X_t are mathematical functions that can be described by the macro-level mathematical functions, $X_0^{sr}, X_1^{sr}, \ldots, X_n^{sr}$, binary operations ° and the subtraction operator $-$. All macro-level mathematical functions and the binary operations supervene on neural states x^{sr}.

The bridge laws from the macro to the micro can be described as follows:

$$e = X(x^{a}).$$

Importantly, the macro states are macro-level functions to calculate feedback errors. The micro states x_t^{sr}change to satisfy $X_t = 0$ for all x^{a}.

Then, the micro neural level laws can be described as follows:

$$x_{t+1}^{sr}, x_{t+1}^{a} \ = f(x_t^{sr}, x_t^{a}, X_t(x_t^{a})).$$

In our theory, the neural states are affected by the macro states, unlike the reductive models, when $e_t \neq 0$. This enables our theory to avoid epiphenomenalism.

References

Baars, G (1988) *A Cognitive Theory of Consciousness*, Cambridge: Cambridge University Press.

Baars, B, Geld, N., Kozuma, R (2021) Global Workspace Theory (GWT) and Prefrontal Cortex: Recent Development, Frontiers in Psychology, 12, https://doi.org/10.3389/fpsyg.2021.749868.

Bedau, M.A. (2008) Is Weak Emergence Just in the Mind?, *Minds & Machines*, 18, pp. 443–459, https://doi.org/10.1007/s11023-008-9122-6.

Block, N. (1978) Troubles with functionalism, *Minnesota Studies in the Philosophy of Science*, 9, pp. 261–325.

Block, N. (1995) On a confusion about a function of consciousness, *Behavioral and Brain Sciences*, 18, pp. 227‐287.

Bradley, M. (2011) The Causal Efficacy of Qualia, *Journal of Consciousness Studies*, 18(11‐12), pp. 32‐44.

Braddon-Mitchell, D., Jackson, F. (1996) *Philosophy of Mind and Cognition*, Oxford: Blackwell.

Broad, C.D. (1925) *The Mind and Its Place in Nature*, Routledge & Kegan Paul.

Chalmers, D.J. (1995) Facing up to the problem of consciousness, *Journal of Consciousness Studies*, 2 (3), pp. 200‐219.

Chalmers, D.J. (1996) *The Conscious Mind: In Search of a Fundamental Theory*, Oxford: Oxford University Press.

Chisholm, R.M. (1982) Human Freedom and the self, Watson, G. (Ed) *Free Will*, New York: Oxford University Press.

Cohen, M.A., Dennett, D.C. (2011) Consciousness cannot be separated from function, *Trends in Cognitive Science*, 15(8), pp.358‐364, https://doi.org.10.1016/j.tics.2011.06.008.

Crick, F. and Koch, C. (1990) Towards a neurobiological theory of consciousness, *seminars in THE NEUROSCIENCE*, 2, pp.263‐275.

Davidson, D. (1970) Mental Events. Foster, L., Swanson, J.W. (Ed) *Experience and Theory*, Amherst, University of Massachusetts Press.

Davidson, D. (1990) Thinking Causes, Heil, J. Mele, A. (Ed) *Mental Causation*, Clarendon Press/Oxford University Press.

Dehanene, S., Changeux, J,-P. (2011) Experimental and Theoretical Approaches to Conscious Processing, Neuron, 70, pp.200‐227, https://doi.org/10.1016/j.neuron.2011.03.018.

Desmond, H., Huneman, P. (2022) The integrated information theory of agency, Behavioral and Brain Sciences, 45, e45, https://doi.org/10.1017/S0140525X21002004.

Ellis, G.F.R. (2019) *Why reductionism does not work. Essay for the Kurt Gödel Prize*,

https://kurtgoedel.de/cms-83FO/wp-content/uploads/2019/11/Why_reductionism_does_not_work-1-28.pdf.

Freeman, W.J. (1999) Consciousness, Intentionality, and Causality, Journal of Consciousness Studies, 6(11‐12), pages. 143‐172.

Gillett, C. (2016) *Reduction and Emergence in Science and Philosophy*, Cambridge, United Kingdom.

Herzog, M.H., Schurger, A., Doerig, A. (2022) First-person experience cannot rescue causal structure theories from the unfolding argument, *Consciousness and Cognition*, 98, 103261,

https://doi.org/10.1016/j.concog.2021.103261.

Hoel, E.P., Albantalis, L., Marshall, W., Tononi, G. (2016), Can the macro beat the micro? Integrated information across spatiotemporal scales, *Neuroscience of Consciousness*, 2016(1), https://doi.org/10.1093/nc/niw012.

Jackendoff, R. (1987) Consciousness and the Computational Mind. Cambridge, Massachusetts: The MIT Press, A Bradford Book.

Jackson, F. (1982) Epiphenomenal Qualia, *The Philosophical Quarterly*, 32, pp. 127‑136.

Kazez, J.R. (1994) Computationalism and the causal role of content, *Philosophical Studies*, 75, pp. 231‑260.

Kim, J. (1978) Supervenience and nomological incommensurable, American Philosophical Quarterly, 15, pp. 149‑156.

Kim, J. (1990) Can Supervenience and 'Non-Strict Laws' Save Anomalous Monism?, Heil, J. Mele, A. (Ed) *Mental Causation*, Clarendon Press/Oxford University Press.

Kim, J. (1998) *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*, Cambridge, The MIT Press.

Kuhn, R.L. (2024) A landscape of consciousness: Toward a taxonomy of explanations and implications, *Progress in Biophysics and Molecular Biology* 190, pp. 28‑169, https://doi.org/10.1016/j.pbiomolbio.2023.12.003.

Laughlin, R.B. (2005) *A Differential Universe: Reinventing Physics from the Bottom Down*, Basic Books.

List, C. (2019) *Why Free Will Is Real*, Cambridge, MA: Harverd University Press.

Macdonald, C., Macdonald, G. (2019) Emergence and Non-reductive physicalism, Gibb, S., Hendry, R.F., Lancaster, T. (Ed) *The Routledge Handbook of Emergence*, Routledge.

Mayr, E. (2018) *Understanding Human Agency*, Oxford: Oxford University Press.

Mendelovici, A., Bourget, D. (2023) Facing up to the problem of intentionality, *Philosophical Perspectives* 37, pp. 228‑247, https://doi.org/10.1111/phpe.12188.

Micheal, R. (2025) The Computational Theory of Mind, Zalta, E.N., Nodelman, U. (Ed) The Stanford Encyclopedia of Philosophy (Summer 2025 Edition.) https://plato.stanford.edu/archives/sum2025/entries/computational-mind/

Mørch, H.H. (2019) Is Consciousness Intrinsic? A Problem for the Integrated Information Theory, *Journal of Consciousness Studies* 26(1‑2), pp. 133‑162.

Naccache, L. (2018) Why and how access consciousness can account for phenomenal consciousness, *Philosophical Transactions B*, 373, 20170357, https;//doi.org/10.1098/rstb.2017.0357.

Nagel, E. (1961) *The Structure of Science. Problem in the Logic of Explanation*, New York: Harcourt, Brace & World, Inc.

Nagel, T. (1970) Armstrong on the Mind, *The Philosophical Review*, 79(3), pp.394‑403.

O'Conner T. (2009) Agent-causal power. Handfield, T. (Ed) *Dispositions and Causes*, Oxford: Oxford University Press.

Pauen, M. (2006) Feeling Causes, *Journal of Consciousness Studies* 13(1-2), pp.129‑152.

Piaget, J. (1950) *The Psychology of Intelligence*, London: Routledge and Kegan Paul.

Piccinini, G. (2004) Functionalism, computationalism, and mental states, *Studies in History and Philosophy of Science*, 35, pp. 811‑833, https://doi.org/10.1016/j.shpsa.2004.02.003.

Potter, H.D., Mitchell, K.J. (2022) Naturalising Agent Causation, Entropy, 24(4), 472. https://doi.org/10.3390/e24040472.

Scheffel, J. (2025) Emergent Will, *Journal of Consciousness Studies*, 32(3‑4), pp. 79‑105.

Searle, J. (1980) Minds, brains, and programs, *Behavioral and Brain Science*, 3(3), pp. 417‑457.

Searle, J. (1983) *Intentionality, an essay in the philosophy of the mind*, Cambridge, UK: Cambridge University Press.

Searle, J. (1984) Intentionality and Its Place in Nature, *Dialectica* 38(2‑3), pp.87‑99.

Schwitzgebel, E. (2015) If Materialism Is True, the United States Is Probably Conscious, Philosophical Studies 172(7), pp. 1697‑1721, https://doi.org/10.1007/s11098-014-0387-8.

Seth, A.K., Bayne, T. (2022) Theories of consciousness, Nature Reviews Neuroscience, 23, pp.439‑452, https://doi.org/10.1038/s41583-022-00587-4.

Sperry, R. (1969) A Modified Concpet of Consciousness, Psychological Review, 76(6), pp. 532‑536.

Sperry, R. (1976) Changing concepts of consciousness and Free Will. Perspectives in Biology and Medicine, University of Chicago Press 20(1), pp. 9‑19.

Sperry, R. (1991) In defense of mentalism and emergent interaction, *The Journal of Mind and Behavior* 12(2), pp.221‑246.

Steward, H. (2012) *A Metaphysics for Freedom*, Oxford: Oxford University Press.

Steward, H. (2017) Action as Downward Causation, *Royal Institute of Philosophy Supplement*, 80, pp. 195‑215.

Tononi, G., Edelman, G.M. (1998) Consciousness and complexity, *Science* 282(5395), pp. 1846‑1851.

Tononi, G. (2008) Consciousness as Integrated Information: a provisional manifesto, *Biological Bulletin*, 215(3), pp. 216‑242.

Tononi, G., Koch, C. (2015) Consciousness: here, there, and everywhere? *Philosophical Transactions B*, 370, pp. 1‑18, https;//doi.org/10.1098/rstb.2014.0167.

Van Gulick, R. (2001) Reduction, Emergence and Other Recent Options on the Mind/Body Problem: A Philosophic Overview, *Journal of Consciousness Studies*, 8(9‑10), pp.1‑34.

Van Riel, R., Van Gulick, R. (2025) Scientific Reduction, Zalta, E.N., Nodelman, U. (Ed) *The Stanford Encyclopedia of Philosophy* (Summer 2025 Edition.) https://plato.stanford.edu/archives/sum2025/entries/scientific-reduction/