CHAPTER THREE

Consciousness, Physicalism, and the Problem of Mental Causation

CHRISTIAN COSERU

1 INTRODUCTION

Is there such a thing as mental causation? Is it possible for the mental to have causal influence on the physical? Or has the old "mind over matter" question been rendered obsolete by the advent of brain science? Whatever our answers to these questions, it seems that we cannot systematically pursue them without considering what makes mental causation problematic in the first place: The causal closure of the physical world. The general idea is that since all things natural can be explained in terms of the most basic entities, events, and processes that physics postulates as fundamental, all explanation in terms of cause and effect is ultimately grounded in our most fundamental physics. So, if minds do appear to exercise any influence at all, not just on other minds but on bodies as well, that influence must admit of physical and/or biological explanation. The causal closure of the physical means that only physical and biochemical processes in the brain can have causal powers. Yet, the possibility of a physicalist explanation of the mental bolsters a view of the mind itself as physical. In short, the problem of mental causation does not go away by

assuming, as most philosophers and scientists nowadays do, that mental states and processes either are (on a token-identity view), or are reducible to, or at the very least supervene on, physical states and processes (Swinburne 2018, 2019). Even the physicalist must offer satisfactory explanations for the full range of beliefs, desires, intentions, and voluntary actions that are constitutive of human agency.

This chapter revisits the problem of mental causation by drawing on a classical debate between Buddhists and the Indian materialists, better known as the Cārvākas. The key issue concerns the relation between cognition and the body, specifically the role this relation plays in causal-explanatory accounts of consciousness and cognition. Here a number of questions arise: Does the central principle of Buddhist Abhidharma reductionism apply to consciousness? Is there a causal criterion for the presence of consciousness? If there is, can this causal criterion account for the specific features of consciousness, for example, its intentional, phenomenal, and subjective character? In short, can a causal account of phenomena be reconciled with the seeming irreducibility of consciousness? The Buddhist answer to the challenge of Cārvāka physicalism displays many of the common features of classical debates in metaphysics about grounding and the mind-body problem. This chapter proposes a philosophical reconstruction that builds on two important features of the Buddhist account: An expanded conception of causality and a robust account of phenomenal content that, taken together, cast new light on the problem of mental causation. It also argues that a satisfactory solution to the problem of mental causation must reject the completeness of the physical principle in favor of conceptions of causality that allow for phenomenal properties to play a causal-explanatory function.

2 PHYSICALISM AND ITS DISCONTENTS

The physicalist claims that to the extent that minds are taken to be non-physical, they cannot cause events in the physical world (Armstrong 1968; Lewis 1972; Smart 1959). However, if minds do appear to exercise such causal influence in the physical world, then that influence must have a natural explanation. The physicalist position on mental causation confronts us with the possibility that the mind itself is physical. This possibility was explicitly rejected by Descartes, whose solution to the problem of mental causation resulted in a dualist, interactionist picture that subsequent philosophers—beginning with Princess Elizabeth of Bohemia's calling into question of the very possibility that immaterial minds could causally interact with material bodies[1]—saw as untenable on both metaphysical and epistemological grounds. While early modern philosophers abandoned the problem in favor of the notion that

the mind and body exist independently (e.g., Leibniz and Malebranche's substantival dualism), a majority of philosophers, at least outside the German Idealism tradition,[2] have favored naturalist explanations that seek to understand mental events and processes in terms of their physical and/or neurobiological correlates.

What motivates this naturalist orientation is the recognition that consciousness must in some way or another be integrated into nature if mental causation is to be possible. And there are few satisfactory metaphysical accounts for how this integration may work. Among them, the most popular is a version of panpsychism or panprotopsychism better known as Russellian monism. Unlike dualism, which places the mental into a distinct domain, and physicalism, which is vulnerable to anti-materialist arguments (e.g., the knowledge and conceivability arguments), Russellian monism puts forth a view of consciousness as at least partly constituted by quiddities, intrinsic properties that categorically ground dispositional properties and that are not amenable to physical explanation. The Russellian monist holds that physical dispositions (associated with mass, charge, and spin) and macrophenomenal properties (e.g., sharp pain, the redness of a red tomato) are grounded in the same kinds of basic microproperties or quiddities, wherefrom they inherit their physical efficacy (Chalmers 2013; Goff 2015; Shani 2021; Strawson 2006). The key issue is whether Russellian monism can address the problem of naturalizing consciousness while retaining the transcendental stance that provides non-empirical access to consciousness as a domain of experience. Much of it, I argue, depends on whether a naturalism fine-tuned to accommodate mental phenomena is possible, and its implications for mental causation.

There is an obvious reason for this turn toward naturalism in the study of mental phenomena: Biology in the wake of Darwin's theory of evolution has been very successful in explaining a whole range of human functions (e.g., digestion, circulation, reflex responses) that do not depend on consciousness. If, *ex hypothesi*, mental processes (e.g., thoughts, emotions, sensations) are just as much a product of evolution as physical processes, there is good reason to assume that complex actions such as perception, the weighing of reasons, and intention, also could result from more refined neurobiological processes and functions. This is precisely the line of reasoning that Huxley (1874) adopts in challenging the Cartesian hypothesis that consciousness, specifically the capacity for self-consciousness, is what sets humans apart from animals, which are mere conscious automata. Had Descartes had the benefit of modern (by nineteenth-century standards) physiology—argued Huxley—he would have extended the automata hypothesis to human consciousness as well, given its dependence on a brain, which shares the same basic structure as other mammalian brains. The alternative hypothesis, namely that—as Huxley (1874: 80) famously put it—humans "are conscious automata endowed with free will" rests on

the notion that continuity of processes and events in the natural world is too well established to suppose that "any complex natural phenomenon" such as human consciousness could come into existence "suddenly, and without being preceded by simpler modifications."

The continuity thesis informs the view that the difference between organisms and inorganic matter is one of degree of complexity rather than kind. Does that mean conscious mental states are ultimately identical with brain states? Given that mental and physical phenomena fall under different categories, are identity theorists (Churchland 2013; Smart 1959) committing a category mistake in claiming that we—this subjective sense of self that is characteristic of the first-person stance we take on ourselves—and our brain are not separate? Armstrong famously argued that the identification of mental states with brain states is perfectly intelligible under a correct view of mental concepts grounded in a causal theory of the mental. Consider a causal concept such as poison, "the concept of something that when introduced into an organism causes that organism to sicken and/or die" (Armstrong 1981: 20). Since only certain substances act in such ways that can be labelled "poison," the concept in question stands for that, whatever it is, that can produce certain effects (which, in turn, creates the possibility of a scientific account of poisons). On a causal theory of mental concepts, "mental state" likewise stands for "something that is, characteristically, the cause of certain effects and the effect of certain causes" (Armstrong 1981: 23). Certain patterns of behavior, for instance, would count as the effects caused by a certain mental state: Hunger, as a mental state, is the cause for an individual seeking food.

Does a causal theory of mental concepts work for all categories of mental states or is it limited to the sort of mental states, such as desires and dispositions, that are relatively easy to accommodate? Would it work for perceptions and beliefs? Perhaps, but only if perceptions and beliefs are nothing more than mappings of the world. But perception discloses both the world and our presence to it, and belief covers intention, deliberation, and the weighing of reasons about matters (e.g., what to value, whom to love, how to achieve one's goals) that do not map neatly onto one's immediate surroundings, if at all. Furthermore, if purpose depends on perception and belief, and perception and belief depend, in turn, on purpose, the causal account of mental concepts faces the problem of circular reasoning. Armstrong thought this criticism can be avoided if the concepts in question are introduced together as correlative concepts (e.g., husband and wife, soldier and army). But it remains an open question whether a causal theory of mental concepts can work for intentional and introspective mental states since it is not clear whether these concepts are causally determined (hence *effects*) or *causes* in their own right. Nor is the thesis about the equivalence of statements about experience with statements about physics obvious: Whether "the sky is blue" is ultimately a statement about the

psychophysics of light perception depends on whether the subjective aspect of experience, that bluish way it is like for me to see the sky can just as easily be naturalized as the blueness of the sky (Kriegel 2005 thinks the difference between the subjective aspect of an experience and its intentional content is sufficiently salient to call into question the equivalence of statements thesis).

As Kim (1999) argued some time ago, the typical response to the problem of mental causation has been either to claim that there are no causal laws about psychological phenomena (e.g., Davidson's anomalous monism) or to propose that the mind is nothing more than a system that computes representations such that only its syntactical properties (not the semantic ones) have causal powers. Both proposals ignore the fact that mental phenomena exhibit properties (e.g., subjectivity, intentionality, qualia) and relations (e.g., logical entailment) that transcend the properties of the elements upon which they supervene. Higher-order cognitive processes such as introspection, reflection, and counterfactual reasoning enact ways of being and knowing—those constitutive elements of personal identity—that cannot be predicted on the basis of the laws governing simpler systems.[3] How are these apparently irreducible aspects of the mental to be reconciled with the causal closure of the physical principle? In short, given that every physical event that has a cause has a physical cause, how is a mental cause also possible? In seeking an answer to these questions, alternative proposals to reductionism,[4] identity theory, and anomalous monism invoke the concepts of emergence and mind-body supervenience.

A concept with roots in the nineteenth-century British emergentism movement, "emergence" stands for the notion that while mental states do emerge from physical states, they are not reducible to the latter. Based on the notion, first articulated in Mill's *A System of Logic* (1843), that some physical systems, when combined, exhibit new dynamic forces (e.g., chemical reactions), emergentism initially positioned itself as a rejection of mechanist ontologies. Whether or not emergence is fundamentally inexplicable and should be taken as a "brute empirical fact" (Alexander 1920, Vol. 2: 46–7) or as a "continued presence" that sustains the course of events specific to each physical and psychological level (Lloyd Morgan 1923: 17), it is clear that "as systems acquire increasingly higher degrees of organizational complexity they begin to exhibit novel properties" (Kim 1999: 3). Not unlike emergence, mind-body supervenience captures the notion that if two systems are wholly alike physically, we should expect the same mental properties to occur, or fail to occur, in each. Since in order for mental causation to work, the mental cannot float freely above the physical, supervenience brings the mental sufficiently close to the physical to make it intelligible and causally effective. But the possibility that mind-body supervenience could fail, for instance, due to the lack of an explicit correspondence between mental events and their neuronal correlates, leaves us with no clear way of understanding the possibility of mental causation.

The Buddhist-Cārvāka debate on the relation between cognition and the body confronts some of the same issues non-reductionist proposals set out to address: How can a model of mental causation predicated on the supervenience of mental phenomena on physical elements and processes account for the specific features of consciousness? But it departs from these other non-reductionist accounts in putting forward a distinctly proto-phenomenological conception of the mind, in which the irreducible elements of existence and/or experience (*dharmas*) are not essences or substances, but activities, properties, and patterns of connectedness. The identification and mapping of these irreducible elements are both descriptive and experimental and, as I have argued elsewhere (Coseru 2019, 2021), function ~~of~~ a kind of naturalized phenomenology (Roy et al. 1999), that is, as a method for bringing into focus, capturing, and categorizing variable mental operations and contents that are normally difficult to attend to, while also submitting to empirical scrutiny about their causal and conditioning factors. A naturalism open to phenomenological analysis seeks to enlarge the scope of naturalization beyond the confines of neurophysicalism.

3 NEUROPHYSICALISM AND THE PROJECT OF NATURALIZATION

The conceptions of naturalism that inform the background philosophical assumptions of emergentism and mind-body supervenience are not without their challenge. Physics and the natural sciences invoked at the beginning of the last century to support a conception of the nature of nature as material, and thus extra mental, have undergone a radical transformation. Quantum mechanics, as a set of mathematical principles for predicting the behavior of subatomic particles, may prove very effective in explaining what happens when phenomena at infinitesimal scales are subjected to the instrumentation of science. Nonetheless, the idea that we can give an account of what the world is like at this scale in and of itself independently of any observation and measurement thereof remains controversial (Chalmers 1996; Stapp 1993, 2007; von Neumann 1955).

As the most recent incarnation of these early conceptions of naturalism, neurophysicalism[5]—essentially the view that any and all mental events are some physical event of other, specifically some central nervous system event (Flanagan 1992, 2007; Koch 2004, 2009)—is predicated on the notion that the scientific method should be adopted in examining not only nature but human experience as well. A century later, we can state with confidence that the advice of these early champions of naturalism has been heeded. The claim that reality is exhausted by nature, however, remains problematic in light of ongoing debates about the meaning and extension of naturalism. This largely "semantic"

problem is further complicated by the varying degrees of commitment to naturalism: That is, those who operate with a rather unrestricted conception of nature embrace a less parsimonious ontology than stronger adherents, for whom naturalism serves as a platform for excluding most, if not all, of what belongs in the experiential domain. Among the latter, one encounters both eliminative physicalists, who seek to reduce all mental content to biological and neurobiological processes (Quine 1960; Sellars 1956; Smart 1959), and token-identity theorists who regard each mental state as some physical, specifically brain state. Both groups are equally diverse, and count among their constituents both realist physicalists, who claim that consciousness is part of the physical world, and type-identity theorists, who think the subjective and physical domains in effect, coincide (Churchland 1986, 2013). While the mind sciences do not rule out the possibility that mental states have nonphysical properties (indeed, the scope of cognitive science is precisely that of understanding the nature of such properties however they may be realized, rather than their reduction to more basic elements), the overwhelming evidence according to champions of neurophysicalism suggests otherwise. Much of it has to do with mental causation since, as Flanagan has noted on more than one occasion, "If mental events—for example, intentions to act—are, as they seem, causally efficacious, then the best explanation is that they are neural events" (Flanagan 2011: 65–6).

As I have argued at length elsewhere (Coseru 2019), grounding the efficacy of mental events on that of neural events is just what the token-identity view consists in: A particular feeling and a given brain state are really the same thing so long as they are constituted by the same token-event. But the token-identity account fails in one important respect: It does not and cannot explain how the phenomenal content of mental states is realized and its apparent capacity to impact physical events. Intentional behavior is an undertaking of the organism as a whole, not merely an outwardly directed mental state: We reach the glass on table by leaning forward, by stretching the arm, and by grasping. Furthermore, awareness of the position of the glass on the table is not an isolated event, but a function of the perceiver's location and sensorimotor contingencies, which disclose the glass as within reach and as graspable. These events certainly admit of two different levels of description: One that considers goal-oriented behavior and the other that accounts for the brain processes underlying this instance of intentional behavior. On the token-identity view, the agent's intention to do x is identical to a particular subset of neurons firing together. Since the dynamic brain event can be described intrinsically, without appeal to extraneous phenomena or events, it serves as the sole causal event. We are thus compelled to concede that descriptions in terms of dispositions and intentions are not any different than those of physiological states (e.g., hunger, satiety, arousal, alertness, body-temperature), and that they are merely shorthand descriptions

for brain processes and their somatic and voluntary functions. Thus, we can understand this particular instance of intentional behavior as token-identical with the brain process, which alone is responsible for the causal interaction we witness at the macro-level: It's not the individual person socialized into a world of table manners that picks the glass off the table, but her brain states. That leaves open the possibility of treating mental states as wholly epiphenomenal: They may well capture the *seeming* nature of experience, but they cannot exert any real *influence* on events occasioned by the only causes there are, brain processes. By conflating the descriptive and explanatory accounts, the token-identity theorist argues that the seeming nature of subjective phenomena (e.g., searing pain) has no explanatory purchase on the instantiation of purposeful and socially modulated behavior. As a result, types of inquiry that seek to figure out why, say, searing pain can be debilitating in one instance and, in another, a cause for resilient action are blocked.

Does neurophysicalism possess the explanatory resources necessary for making sense of such basic subjective and intentional behavior? Not knowing how such mental states are realized and yet assuming that they are realized as brain states takes a leap of faith. While neurophysicalism does offer a compelling strategy of naturalization, I will argue that we are better served by operating with a more capacious conception of nature, one that allows for non-supervenient mental causation, and for a theory of action that regards organisms as complex adaptive systems (where the mental is a self-organized structure with its own emergent dynamics that is not reflected in a change in the physical) to play a role. The naturalism I propose here, as more suited to the task at hand, is distinctly phenomenological. It argues that our conception of the mental must account for its phenomenal features in ways that capture their event-causal efficacy. These, in turn, become causally relevant in explaining how action is successfully accomplished with respect to criteria (e.g., deadlines, assent, opportunity) that are unintelligible in the third-personal language of neuroscience. Faced with such accounts of mental causation, critics typically invoke the causal closure of physical domain as evidence for the epiphenomenal character of mental states. But, as I will argue, what makes epiphenomenalism unattractive is the premise that cognitive events, which arise as a result of the tight causal coupling between perception, memory, reflection, and action, should themselves be causally inert. Arguments that invoke the closure of the physical domain have the peculiar distinction of lacking empirical grounding: Such closure is often assumed by a priori postulation (Buhler 2020).

As with modern attempts to explain mental states either in terms of their token identity or supervenient status, the classical Buddhist-Cārvāka debate on the relation between cognition and the body showcases some of the problematic aspects of reductionist accounts that preclude event-causal explanations of

consciousness itself. Unlike the modern debate, however, supervenience arguments against the autonomy of cognition, specifically as advanced by the Indian Buddhist philosopher Dharmakīrti, are meant to refute the existence of a distinct metaphysical realm of mental phenomena, not the efficacy of mental processes. These arguments are mainly a priori, and they aim to counter the Cārvāka's challenge that consciousness originates or has its causal basis in the body.[6] As should be obvious from the analysis that follows, Dharmakīrti's arguments against the Cārvāka offer interesting new ways to conceptualize the problem of mental causation.

4 PHENOMENAL PRIMITIVES AND THE QUESTION OF EMERGENCE

Buddhist perspectives on the subjective aspects of consciousness emphasize what we may call—using the language of contemporary philosophy of mind—the dynamic, embodied, and embedded functioning of the five aggregates. In the schematic analysis of the five aggregates, however, only "body" (*rūpa*) is a physical aggregate *stricto sensu*. Sensations, apperception, and volitions can acquire an objectual aspect, but are not empirical phenomena proper. Nor are they abstract entities with well-defined properties and functional characteristics. For instance, a sensation of pain is not reducible to the physical substrate, say a finger, in which it is instantiated (nor presumably to a mere physiological response). Rather, as object-oriented cognitive aspects (*viṣayākāra*), sensations, apperception, and volitions are included in the broader Abhidharma category of mental factors (*caitasika*). Feelings may define the quality of the impressions that result from contact with an object, with the implication that they perhaps stand in a causal relation with these objects. But as internal mental states, they are also conditioned by habitual tendencies (*vāsanā*), which, in turn, they condition: Your physical condition after strenuous exertion may feel pleasant or unpleasant depending on your level of fitness and degree of exercise frequency. Likewise, apperception (*saṃjñā*), the capacity to make intelligible or cause to be understood, although dependent on a multiplicity of psychological factors, captures the datum of experience only as fused into a single percept. Volitions too fit the same profile, with one important difference: Rather than attending to the object at hand or providing a sort of transcendental unity of apperception, they bring forth future states of existence. As dispositions to act in certain ways, they cleave the mental domain into two classes of conditioned phenomena: Those that are internal to consciousness, such as, for instance, obsessive dispositions like greed and delusion, and those that are dissociated from it, usually taken to refer to latent dispositions typically comprising various biological and physical traits.[7]

It has been argued, most forcefully by Siderits (2003, 2011), that the reductionist framework of analysis in the Abhidharma entails physicalism, the view that everything is or supervenes on the physical (where "physical" stands for the world as described by our best physics). Although Dharmakīrti shares the empiricist stance of Abhidharma, the naturalism that informs his epistemological project is patently anti-physicalist. According to the Sautrāntika Abhidharma account of materiality that Dharmakīrti favors, entities reduce to their phenomenal primitives[8]: The particular is a token of a type, not blue in general, but this unique intensity of cerulean. Furthermore, the formal properties of material objects are analyzed either in terms of how they are impacted by contact or as factors that oppose resistance. These properties, however, do not extend to the atoms themselves, which according to the Abhidharma form the building blocks of materiality. As monadic units the atoms are seen as devoid of any formal properties. It is only as atomic compounds that atoms are subject to the same properties of resistance and destruction as composite material entities.

The reductionist model of Abhidharma, like all philosophical attempts to carve reality at its joints, works against the common conception that empirical awareness provides access to an external, stable, and self-sustaining world: A world as it is in and of itself or intrinsically (as captured by the notion of *svabhāva*) rather than as it appears to an observer. But the conception of the human mind that Abhidharma offers is grounded in a metaphysics of experience, rather than of momentary, point instant particulars. Indeed, the irreducible elements of existence and/or experience (*dharma*) are not fixed substances but activities, properties, or dynamic patterns of connectedness that are constitutive of the world as perceived. It is experience rather than the abstract schema of a causal web of impersonal elements that marks the boundary of what there is: The nexus of causes and conditions that set the boundaries of lived experience are determined by the operations of our cognitive architecture. Color, for instance, only exists for an organism that is sensitive to light.

How does this dynamic picture of what there is take on the characteristics of subjectivity and intentionality? And how do these emergent phenomena in turn create the conditions for grasping and attachment? For the Buddhist, the answer does not lie primarily in the patterns of conditioning that explain the aggregation of phenomena, but in certain defining characteristics that belong to the structure of experience[9] itself. Not only are the senses conceived as receptacles of experience, they also serve as ground or support, joining the external domain of sensory activity with the internal domain of subjective awareness. Of course, Abhidharma reductionism is predicated on the notion that things reduce to their component parts, which are ultimately real only if they are further irreducible. If something can be reduced either by breaking it down to more basic constituents or through conceptual analysis, then it is not

ultimately real. Pots are not ultimately real, nor are persons—as persistent, unchanging subjects—real in this ultimate sense. Does this principle of reductionism apply to consciousness? Certainly, consciousness is but a stream of momentary conscious events of different types (visual, auditory, introspective, etc.). But reductionism about consciousness is problematic. Why? Because it cannot explain its most basic features: Its intentional, phenomenal, and self-reflexive character.

Abhidharma metaphysics thus retains an irreducible notion of consciousness as a stream of mind-moments, even though the character and structure of these mind-moments are a subject of debate both within Buddhism and between Buddhists and their opponents. And it is the Cārvākas that pose the most serious challenge to the Buddhist attempt to accommodate phenomenal primitives within a generally atomist ontology. For the Cārvākas do not merely advance a new theory of causation. Rather, they put forward an austere ontological picture according to which every phenomenon reduces to the four basic material elements—earth, water, fire, and air—which alone are ultimately real. Emergent phenomena, like the fully formed body and discerning cognition, supervene on these basic elements as their functional properties. The Cārvāka physicalists thus take issue with the notion that consciousness originates outside the order of the material domain, perhaps in a previous instance of cognitive awareness associated with a different body (as demanded by the view of karmic rebirth). Of course, the causal closure of the physical domain at work in Cārvāka metaphysics does not preclude an event-causal explanation of consciousness itself. Indeed, the supervenience argument against the autonomy of cognition is meant to refute the existence of a distinct metaphysical realm of mental phenomena, not the efficacy of mental processes. On this emergentist picture, cognition is a scaled phenomenon that tracks closely the development of the body: The absence of sensory organs in the embryonic stage, according to Cārvāka, precludes the attribution of any sort of perceptual awareness at that level of development.

For the Cārvāka, thus, consciousness is a manifestation of the body's development, functional organization, and responsiveness to objects, situations, and things. Just like fermented grain yields a liquid with the capacity to intoxicate, so also consciousness must be regarded as nothing more than a product of the type of material organization that is constitutive of biological organisms. Indeed, the Cārvāka offers a rather stark picture of the human condition, as we see in one of their earliest accounts:

> We shall now examine the principles of reality. Earth, water, fire, and air are the principles (of reality), nothing else. Their combination is called the 'body', 'senses', and 'objects'. Consciousness arises out of these elements, as the power to intoxicate arises out of fermenting ingredients. The human

> being is nothing but the body endowed with consciousness. Cognition arises from the body itself, because of the presence of consciousness when there is a body.
>
> (Bhattacharya 2002: 603ff)

If consciousness is nothing but an emergent property of the body, then consciousness can be present neither without a faculty of apprehension, nor in the absence of suitable objects of apprehension. For the Cārvāka, then, seeking to circumvent these objections, as the Buddhist does, by pointing to the momentary nature of phenomena undermines the "continuity of consciousness" thesis. That is, the Buddhist wants to claim both that each moment of consciousness has as its cause a previous mind moment in the stream of consciousness (on the principle that consciousness cannot emerge from insentient matter) and that specific types of conscious mind moments (e.g., visual awareness of a blue patch) nonetheless arise as a result of the coming together of objects and their corresponding cognitive faculties (on the principle of dependent arising). The main reason for the objection concerns the idea of momentariness itself, which entails the serial dissolution of both the object and the cognitive process by which it is apprehended.

If the body is taken to be the cause of consciousness, as the Cārvāka claims, the question naturally arises: is it "cause" in the form of an aggregated whole, as a composite of various elements, or as an aggregate of atoms? And, given the requirement that the sensory systems are fully developed, does the body (as a physical or biological substrate) serve as a cause along with, or independently of, the senses? Finally, should the body (with its physical and functional properties) be regarded as a material cause or simply as a condition for the possibility of consciousness?

5 AN ERROR ARGUMENT FOR MENTAL CAUSATION

What makes Dharmakīrti's response to the Cārvāka physicalist challenge relevant to contemporary debates are the metaphysical considerations that ground his causal account, and the specific conception of consciousness that thus emerges. As already noted, Buddhists who take Dharmakīrti's lead hold that consciousness is but a stream of conscious episodes of different types (visual, auditory, etc.). If the subjective and intentional dimensions of consciousness are nothing but aggregates of discrete mind moments, a dilemma arises: What accounts for the sense of continuity of awareness and, more importantly, what could serve as the basis for the arising of each instance of cognitive awareness from one moment to the next? The bundle theory stipulates that every phenomenon is part of a complex causal web. Indeed, the Sanskrit notion of *skandha* (lit., "heap") captures rather well the aggregated nature of phenomena—something

fashioned by the collective combination of multiple causes and conditions (as Vasubandhu glosses it in AKBh ad I, 7). The constitutive factors themselves exist only as part of a causal continuum of interdependently arising phenomena. Of course, these constitutive factors do not all contribute in equal measure: Some are basic or necessary and some are merely contingent. The Cārvāka claims that the body alone is the source of cognition. But on the aggregate model of personal identity put forth by the Buddhist, the body is just one among the five constitutive factors of agency, even as statements of personal identity have no corresponding states of affairs.[10] The principle that establishes effects as markedly different from their causes runs counter to empirical evidence. We observe that like causes like: Cows give birth to calves, and fermented milk yields yoghurt.

On this principle of causal homogeneity, then, cognitive awareness cannot arise from something non-conscious, such as the physical body. As Dharmakīrti notes (PV II vv. 35–36a), there could be "unwarranted consequences" for presupposing otherwise. Indeed, Dharmakīrti is committed to a strict ontological difference between "cause" (*kāraṇa*) and "condition" or "conditioning factor" (*pratyaya*): The former can only give rise to a specific type of effect, while the latter can serve as a basis for the arising of multiple effects. The acorn can only grow into an oak tree, but the same soil and climactic conditions may provide support for various tree species. On the reductionist Abhidharma model, all aggregate entities reduce to two kinds of basic constituents: Elemental atoms comprising the four primary elements, and the atomic totality, which includes the secondary elements associated with each of the four sense spheres (with the exception of sound). Although there is some debate among Abhidharma philosophers about how best to draw the lines between primary and secondary existents, conscious mental states, as a domain of phenomenal primitives, do belong in the Buddhist's ultimate ontology. Their reality is premised on their causal and pragmatic efficacy. For instance, in perceiving a pot, it is not the pot itself that serves as the basis for the arising of the cognitive event but rather the causal efficacy of phenomenal primitives that ground a conception of pots as material entities.

Furthermore, middle-size dry goods such as pots are not perceived in a vacuum. Rather, conditioning factors play an important role. Just as the potter fashions his clay in whatever form the vessel will take by increasing the preponderance of the water element, so also, under certain conditions something solid may become liquid, like heat causing the melting of a block of ice. Given the speculative nature of the Abhidharma metaphysics, there should be no surprise in finding disagreements about the specific ways in which properties attach to each aggregated entity. For instance, while for the Vaibhāṣikas entities borrow their physical properties from the elements themselves, Sautrāntikas take them to be present only as mere potentialities.

In a block of ice, the fire element is only potentially present, for without it, ice cannot melt into water. Dharmakīrti works out his account of causal efficacy in terms of the strict regularities that must obtain between elements in a causal series. These regularities act as a kind of "restriction in causal potential" (*śaktiniyama*)—a notion that Dharmakīrti uses to argue for the limited or restricted efficacy of causal elements. For instance, a lotus seed cannot produce a cow and oil cannot be extracted from sand. The so-called essential nature of the causally efficient element in a causal chain suggests that entities are not simply the product of a given causal chain or causal complex. Rather, they are the product of specifically active elements within that chain and of the conditions that make it possible for those active elements to manifest their potentiality.

While Dharmakīrti does in principle concede that the material elements could serve as a basis for the arising of cognition, his conception of causation as limited to relations between homogeneous elements within the causal chain prevents the establishment of a direct link between consciousness and the body. Likewise, while the principle of preponderance may apply to all kinds, a cow is not just a collection of elements with a certain predominant property like solidity, heat, or capacity to produce milk. Nor is it a conceptually constructed entity like a forest, or a cart, that is analytically reducible to its constitutive parts. There must be more than just the configuration of matter that accounts for the arising of cognitive awareness and its specific intentional content (PV II vv. 37–8).

Indeed, for Dharmakīrti the reflexive dimension of awareness cannot be a function of the body arising together with all the senses because its occurrence is observed even when one or more of the senses are impaired (PV II v. 47). The wealth of empirical evidence from clinical neuroscience about such phenomena as the "locked-in syndrome" or the persistence of "minimal consciousness" in patients diagnosed as being in a vegetative state gives credit to the hypothesis that consciousness enjoys a greater degree of autonomy than basic cognitive function (Koch 2012; Laureys et al. 2005; Laurey, Perrin, and Brédart 2007; Naci, Sinai, and Owen 2017; Noirhomme et al. 2007;). This sort of evidence, it seems, lends support to Dharmakīrti's thesis that basic sentience enjoys a certain degree of causal autonomy from more specific higher-order modes of cognitive awareness. It also suggests that, given the difficulty of diagnosing whether a patient is in a minimally conscious rather than a permanent vegetative state, the distinction between unconscious mental states and states of consciousness with minimal cognitive and behavioral function is less clear than it may seem. Rather than being unconscious, a cognitively and behaviorally non-responsive individual could simply be minimally conscious. Most importantly, in the absence of a better understanding of the tight correlation between mental and physical (e.g., brain) states, such evidence sets the stage for developing a wider conception of causality than physicalism allows.

Consider the occurrence of a sensation, say of pain, resulting from a wound in the body. The pain has both qualitative features or qualia (sharp, stinging) and intentional content, insofar as it discloses the body as the locus of tissue damage. The co-occurrence of bodily processes and specific mental states, however, does not suggest that the body actually causes it, but only that it serves as a contributing factor in the arising of cognition. What we see here is a clear example of Occam's Razor: Dharmakīrti argues against taking cognitive awareness to be a product of bodily functions because he thinks the mental domain is the natural place for cognitive awareness. The mental domain is sufficiently complex to support its own operations. So, while the body and its environment both support and constrain its operations (e.g., functional eyesight enables vision, while night restricts it), cognitive activity functions on the basis of a different set of principles (e.g., of illumination, representation, logical entailment) (PV II vv. 33–44). Nothing is closer to each instance of cognitive awareness than a cognition immediately preceding it. Why not postulate that each state of cognitive awareness serves as the antecedent cause for cognition? Hence Dharmakīrti's dictum: "let only what is observed as the cause always be considered the cause" (PV II v. 44cd). And what is observed is the constant stream of conscious mental states.

Dharmakīrti's attempt to carve out a space for the autonomy of the reflexive dimension of cognitive awareness from material causation while retaining the efficient-causal model showcases not only his conceptual ingenuity but also his keen phenomenological sense. His case for the autonomy of consciousness draws on an error argument based on impaired cognitive function: "Nor are the senses, or the body together with the senses, the cause of cognition, [for] even when every single one of the senses is impaired, the [corresponding] cognitive awareness is not impaired. But when [the cognitive awareness] is impaired, their (i.e., the senses') impairment is observed" (PV II. v. 39). Phenomena such as phantom limbs and the locked-in syndrome are perfect examples of the sort of impaired cognitive function that does not impact cognitive awareness in the respective domain. But central to the error argument is the premise that cognitive awareness is nonetheless in some kind of dependency relation to the body, as demanded by the causal principle of dependent arising. For instance, visual awareness can only emerge in organisms that are sensitive to light. Can Dharmakīrti answer the physicalist challenge while retaining a causal-explanatory framework that is necessary for explaining the relation between cognition and the body?

As I have argued at length elsewhere (Coseru 2020), models of causation in the material domain face certain limitations when extended to consciousness and cognition. For instance, on a strict account of causal generation, cognitive error would track closely deficient causation. But that does not always happen. One might perceive a sparkling lake where there is only a naturally occurring

optical illusion. This perceptual illusion is not simply a case of misapprehension, for the illusion persists even after it has been disambiguated (i.e., after one has come to apprehend the appearance of the lake as a mirage). What the error argument targets is strict causal generation: The notion that each mental state is instantiated by a suitably relevant combination of physical elements and processes. The persistence of perceptual illusion even after disambiguation, and the possibility of effective action such disambiguation affords (by not chasing after a mirage), works against the strict causal model of the Cārvāka physicalist, which reduces human agency to changes in the microphysical structure of each individual.

When Dharmakīrti claims that a trustworthy cognition (*avisaṃvāda*) is not merely epistemically salient but also causally effective, he advances a naturalistic account of cognition, one that accounts for the intentional structure of awareness and its phenomenal character: Perception is not simply the apprehension of a unique particular as such; rather it is the apprehension of a particular *as perceived*, which also discloses the perceiver's intentional stance. In the case of perceptual illusions such as mirages, it is not only the perceiver's vantage point but also the phenomenal character of the experience itself that ensures successful action: Illusory water can neither quench thirst nor afford immersion.

Has the Buddhist satisfactorily answered the challenge of physicalism? It is clear that by rejecting the notion that intentional objects are causally related to the experience of unique particulars (e.g., PV III v. 320), Dharmakīrti showcases the importance of phenomenological considerations (specifically about the structure of awareness) in settling the debate about mental causation. This point is necessary if his account of the efficacy of cognition, which takes causal explanation to contain an element of ontological subjectivity, is to succeed.

6 CONCLUSION

The debate about mental causation has been primarily driven by efforts to explain how mental properties could be causally relevant to bodily behavior. Neither agency, nor free will and moral responsibility can be satisfactorily explained without appealing in one way or another to mental causation. Indeed, if behavior was not the result of mind's activities—its reasoning, deliberation, and decision to act—there would be little scope for a notion of responsibility. Isolating the mental from the physical or rendering it epiphenomenal by attributing all causality to the physical, makes it difficult to explain how what goes on in our minds relates to what the body does. To these concerns, the Buddhist-Cārvāka debate on the relation between cognition and the body adds another, specifically about demonstrating the possibility of freedom, which

is essential to overcoming the limitations of the human condition. For the Cārvākas the most probable explanation for the existence of the universe is a series of random events. In rejecting both the law of karma and the concept of destiny, their argument is that implicit in these notions is a view of existence as inherently purposeful. In response, the Buddhist emphasizes not only the reality of karmic action but also the efficacy of individual effort. Indeed, against the claim that nothing is done either by oneself or another, the Buddha pointed out that articulating any view whatsoever shows that there is an element of initiative, that one either strives to overcome some resistance or to reach the sort of reflective equilibrium that comes with understanding and insight (Bodhi 2012: 901). Insofar as they eschew such concerns, the Cārvākas also discount the importance of efficient causation, focusing instead on material causes and conditions, as their emergentist account of consciousness demonstrates. In that regard, they both align with, and face the same challenges as, present-day physicalists. Indeed, from a modern standpoint, it may be objected that consciousness is a subjective phenomenon and thus not amenable to scientific (hence, efficient causal) explanation.

The Buddhists may well admit that aggregated entities reduce to their ontological primitives, which alone are real. But causally describable series of phenomena are not incompatible with treating some basic phenomena as irreducibly mental, so Buddhist reductionism does not necessarily entail physicalism. However, on a distinctly Kantian line of argumentation about the irreducibility of normative relations (relations that obtain in the logical space of reason), conscious mental events will not admit of efficient-causal description. Against such a normative framework, Dharmakīrti's appeal to causal explanation as a criterion for the efficacy of epistemic practices would arguably render his account indistinguishable from that of his physicalist opponent. The efficacy of reasons lies primarily in the relevance of their content rather than the fact that they fall under some efficient-causal description. But the Buddhist does not think the relevance of such content can be ascertained on logical or conceptual grounds alone since what makes a cognition veridical is its causal efficacy, the fact that it can lead to successful action. Hence, the Buddhist response to the problem of how to think of consciousness using the language of efficient-causal explanation is framed by two sorts of considerations: First, about the basis or support of consciousness (that is, about what sorts of factors might be responsible for the arising of different aspects of consciousness given a metaphysical commitment to momentariness), and second, about the structure and character of consciousness (which reflects a commitment to the reflexivity thesis). On the account put forth here, consciousness is thus constitutive of a constant and continuous stream of discrete cognitive events, not independently of, but rather alongside, various conditioning and dispositional factors. The problem is not how consciousness could arise from purely causal interactions

in the psycho-physical domain. Rather, the problem is how this domain, which also includes irreducible phenomenal properties, conditions the arising of specific cognitive events.

The Buddhist account of mental causation sketched here insists that facts about one's subjective experience are not empirically accessible in the way that facts about the external world are. The epistemic subjective-objective distinction thus rides on a more basic, ontological distinction in modes of existence. Pains and pleasures manifest only as phenomenal qualities of awareness. Their experience-dependent status, however, does not render them any less real. Likewise, just because we are physiologically limited to perceiving only a narrow range of light frequencies does not mean our color experiences lack objective properties. The capacity to unambiguously apprehend phenomena not only as they seem, but as they presumably are, suggests that we can have an epistemologically objective account of the subjective and intentional dimensions of consciousness. In short, ontological subjectivity is no bar to epistemic objectivity. Framing mental causation as an observer-relative phenomenon does not mean, however, that it is not a real feature of our ontology. Rather, its observer-relative status simply suggests that it contains an element of ontological subjectivity.

NOTES

1 As she famously puts in one of her letters to Descartes, after acknowledging that it is largely due to our ignorance of what causes bodies to move that we attribute such causal power to the soul, "I admit that it would be easier for me to concede matter and extension to the soul than to concede the capacity to move a body and to be moved by it to an immaterial thing" (Princess Elizabeth of Bohemia and Descartes 2007: 68).

2 One notable exception is Schelling, whose natural philosophy (*Naturphilosophie*) puts forward a dynamic conception of nature that seeks to bridge the chasm opened up, on the one hand, by the Newtonian mechanistic conception of nature as a domain of causality and, on the other, by the Kantian notion that the spontaneity of practical reason places the self-constituting subject beyond the conditioned realm of determination. Rejecting the Kantian conception of freedom as mere noumenal spontaneity, Schelling argues instead for the possibility that genuine causality expresses itself through reason: "the empirical I cannot possibly realize itself, because the empirical I as such does not exist through itself, through its own free causality" (Schelling 1976: I, 2, 166). Rather, as he claims, "'reason' is a mere play of higher and necessarily unknown natural forces" such that "there is nothing impossible in the thought that the same activity by which Nature reproduces itself anew in each successive phase, is reproductive in thought through the medium of the organism" (Robertson 2004: xviii–xix). See also Woodard (2019: 18), who argues that Schelling integrates human thought regarding it as a "species of motion."

3 Predictive processing models of cognition, according to which the brain makes sense of the sensory data by making probabilistic inferences about the world and

correcting them in order to minimize predictive errors, advance the claim that minds are "inferentially secluded from the world … and more neurocentrically skull-bound than embodied and extended" (Hohwy 2013: 259). Critics, however, argue that without embodiment and interaction it is hard to make sense of the dynamic variants of predictive coding (Clark 2016).

4 One way to understand the difference between the non-reductionist and the reductionist views of personal identity is along the simple/complex divide: The non-reductionist favors the simple, soul or Cartesian Ego, view, whereas the reductionist prefers the complex view that entails relations among physical and psychological states. Holding a soul view, of course, does not necessarily amount to holding a brute fact view, although in the absence of non-circular criteria for personal identity (of the sort required by the complex view) it is hard to tell them apart. What motivates recent defenders of the simple view (e.g., Baker 2013; Lowe 2010, 2013; Nida-Rümelin 2013; Swinburne 2013) is not commitment to a Cartesian Ego, but rather the notion that a specific, perhaps non-conceptual and pre-reflective, type of self-awareness seems indispensable to framing any account of personal identity.

5 I use "neurophysicalism" here mainly in its *token* sense as the view that for every mental state there is a suitable set of neural correlates. Correspondingly, *type* neurophysicalism stands for the view that different types or kinds of mental events (e.g., perception, belief, emotion) are similarly realized by each member of a species (cf. Flanagan 2007: 27).

6 Detailed analyses of these arguments can be found in Vetter (1964), Hayes (1993), Taber (2003), and Arnold (2008).

7 Detailed accounts of this twofold analysis of phenomena are found in Vasumitra's *Pañcavastukavibhāṣāśāstra* [Wu shih p'i-p'o-sha Zun], T 28 (1555), p. 989b2, Vasubandhu's *Abhidharmakośabhāṣya*, II, 23–34, and Yaśomitra's *Spuṭārthā Abhidharmakośavyākhyā*. Cox (1995: ch. 4) offers the most detailed account to date of the *citta-viprayuktasaṃskāra*. For a broader discussion of the process by which mental factors that arise in conjunction with a given intentional object come to be associated with the qualities of the respective object, see Waldron 2003: 57ff.

8 One way to understand "phenomenal primitives" is by analogy to the sort of phenomenal properties that Russelian panpsychists associate with certain fundamental entities that exhibit conscious experience. Since these entities are assumed to be microphysical, they thus exhibit "microphenomenal" or "protophenomenal" properties (Alter and Coleman 2019; Chalmers 2013).

9 By "structure of experience" I mean things like interiority, directedness toward an object, background awareness, and experiential horizon, all of which can result from the aggregation of both non-phenomenal properties (shape, texture, boundary conditions, luminosity, etc.) and phenomenal properties that give experience its structure or aspectual shape (sharpness, brightness, edge, etc.). One way to avoid the physicalist view that the structure of experience could be entirely constituted of non-phenomenal properties is to distinguish between protophenomenal and structural properties, and to require that there be "an a priori entailment from truths about protophenomenal properties (perhaps along with structural properties) to truths about the phenomenal properties that they constitute" (Chalmers 2013: 260).

10 Although "states of affairs" and "facts" are often used synonymously, the latter typically designates *actual* (hence, existing) states of affairs. Many philosophers (Armstrong 2009; Meixner 2009; Olson 1987) take *states of affairs* to be the constituent, building blocks of reality, and many, though not all (Voltolini 2006), distinguish them from *propositions*: The former are taken to be *truthmakers* (e.g., entities that make true propositions true), while the latter are *truthbearers* (the vehicles of truth or falsity). See Reicher (2009) for a detailed discussion of how to think of states of affairs relative to facts, propositions, events, and tropes, and of various arguments for and against their very existence.

REFERENCES

Primary Sources

AK *Abhidharmakośa*: See AKBh.

AKBh *Abhidharmakośabhāṣya*: *Abhidharmakośa* and *Bhāṣya* of Āchārya Vasubandhu with *Sphuṭārtha* Commentary of Āchārya Yaśomitra, ed. P. Pradhan. Vārāṇasī 1975.

PV *Pramāṇavārttika*: Ācāryamanorathanandivṛttiyutam (Dharmakīrtinibandhāvaliḥ), ed. Swami Dwarikadas Shastri. Vārāṇasī 1968.

PVSV *Pramāṇavārttikasvavṛtti*: The *Pramāṇavārttikam* of Āchārya Dharmakīrti with the Commentaries "*Svopajñavṛtti*" of the Author and "*Pramāṇavārttikavṛtti*" of Manorathanandin, ed. Ram Chandra Pandeya. Delhi: Motilal Banarsidass, 1989.

TS *Tattvasaṃgraha*: *Tattvasaṃgraha* of Āchārya Shāntarakṣita with the Commentary "*Pañjikā*" of Shrī Kamalaśīla, vols. I–II, ed. Swami Dwarikadas Shastri. Vārāṇasī 1968.

TSP *Tattvasaṃgrahapañjikā*: See TS.

Secondary Sources

Alexander, S. (1920), *Space, Time, and Deity: The Gifford Lectures at Glasgow 1916–1918*, Vols I–II, London: Macmillan.

Alter, T and S. Coleman (2019), "Russelian Monism and Mental Causation," *Noûs*, 55 (2): 409–25.

Armstrong, David M. (1968), *A Materialist Theory of the Mind*, London: Routledge & Kegan Paul.

Armstrong, D. M. (1981), *The Nature of Mind*, Brighton: The Harvester Press.

Armstrong, D. M. (2009), "Questions about States of Affairs," in M. E. Reicher (ed.), *States of Affairs*, 39–50, Frankfurt: Ontos Verlag.

Arnold, D. (2008), "Dharmakīrti's Dualism: Critical Reflections on a Buddhist Proof of Rebirth," *Philosophy Compass*, 3 (5): 1079–96.

Baker, L. R. (2013), "Personal Identity: A Not-so-simple View," in G. Gasser and M. Stefan (eds), *Personal Identity: Complex or Simple*, 179–91, Cambridge: Cambridge University Press.

Bhattacharya, R. (2002), "Cārvāka's Fragments: A New Collection," *Journal of Indian Philosophy*, 30 (6): 597–640.
Bodhi, B., trans. and ed. (2012), *The Numerical Discourses of the Buddha: A Translation of the Aṅguttara Nikāya*, Boston: Wisdom Publications.
Broad, C. D. (1925), *The Mind and Its Place in Nature*, London: Routledge & Kegan.
Buhler, K. (2020), "No Good Arguments for Causal Closure," *Metaphysica*, 21 (2): 223–36.
Chalmers, D. J. (1996), *The Conscious Mind*, New York: Oxford University Press.
Chalmers, D. (2013 [2015]), "Panpsychism and Panprotopsychism," *Amherst Lecture in Philosophy* 8; reprinted in T. Alter and Y. Nagasawa (eds), *Consciousness in the Physical World: Perspectives on Russellian Monism*, (Philosophy of Mind), 246–76, New York: Oxford University Press.
Chalmers, D. J. (2015), "Panpsychism and Panprotopsychism," in T. Alter and Y. Nagasawa (eds), *Consciousness in the Physical World: Perspectives on Russellian Monism*, 246–76, New York: Oxford University Press.
Churchland, P. (1986), *Neurophilosophy: Toward a Unified Science of the Mind/Brain*, Cambridge, MA: The MIT Press.
Churchland, P. (2013), *Touching a Nerve: The Self as Brain*, New York: W. W Norton & Company.
Clark, A. (2016), *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*, Oxford: Oxford University Press.
Coseru, C. (2019), "Consciousness, Naturalism, and Human Flourishing," in B. Seok (ed.), *Naturalism, Human Flourishing, and Asian Philosophy: Owen Flanagan and Beyond*, 113–30, New York: Routledge.
Coseru, C. (2020), "Consciousness and Causality: Dharmakīrti against Physicalism," in B. Kellner, P. McAllister, H. Lasic, and S. McClintock (eds), *Reverberations of Dharmakīrti's Philosophy: Proceedings of the Fifth International Dharmakīrti Conference Heidelberg*, August 26 to 30, 2014, 21–40, Wien: Verlag der Österreichischen Akademie der Wissenschaften.
Coseru, C. (2021), "Whose Consciousness: Reflexivity and the Problem of Self-Knowledge," in M.Siderits, C. Keng and J. Spackman (eds), *Buddhist Philosophy of Consciousness: Tradition and Dialogue*, 121–53, Leiden: Brill.
Cox, C. (1995), *Disputed Dharmas—Early Buddhist Theories on Existence—An Annotated Translation of the Section on Factors Dissociated from Thought from Saṅghabhadra's Nyāyānusāra*, Tokyo: Studia Philologica Buddhica.
Flanagan, O. (1992), *Consciousness Reconsidered*, Cambridge, MA: MIT Press.
Flanagan, O. (2007), *The Really Hard Problem: Meaning in a Material World*, Cambridge, MA: MIT Press.
Flanagan, O. (2011), *The Bodhisattva's Brain: Buddhism Naturalized*, Cambridge, MA: The MIT Press.
Goff, P. (2015), "Against Constitutive Russellian Monism," in T. Alter and Y. Nagasawa (eds), *Consciousness in the Physical World: Perspectives on Russellian Monism*, 370–400, New York: Oxford University Press.
Hayes, R. (1993), "Dharmakīrti on *punarbhava*," in Egaku Maeda (ed.), *Studies in Original Buddhism and Mahāyāna Buddhism,* Vol. I, 111–30, Kyōto: Nagata Bunshodo.
Hohwy, J. (2013), *The Predictive Mind*, Oxford: Oxford University Press.
Hohwy, J. (2016), "The Self-evidencing Brain," *Nous,* 50: 259–85.
Huxley, T. H. (1874), "On the Hypothesis That Animals Are Automata, and Its History," *Fortnightly Review*, 16: 555–80.

Kim, J. (1998), *Mind in a Physical World*, Cambridge, MA: The MIT Press.
Kim, J. (1999), "Making Sense of Emergence," *Philosophical Studies*, 95: 3–36.
Koch, C. (2004), *The Quest for Consciousness: A Neurobiological Approach*, Englewood, CO: Roberts.
Koch, C. (2009), "Free Will, Physics, Biology, and the Brain," in M. Nancey, E. George, F. R. O'Connor and Timothy (eds), *Downward Causation and the Neurobiology of Free Will*, 31–52, Springer Verlag.
Koch, C. (2012), *Consciousness: Confessions of a Romantic Reductionist*, Cambridge, MA: The MIT Press.
Kriegel, U. (2005), "Naturalizing Subjective Character," *Philosophy and Phenomenological Research*, 71 (1): 23–57.
Laureys, S., F. Pellas, P. Van Eeckhout, S. Ghorbel, C. Schnakers, F. Perrin, J. Berre, M. E. Faymonville, K. H. Pantke, F. Damas, M. Lamy, G. Moonen and S. Goldman (2005), "The Locked-in Syndrome: What Is It Like to Be Conscious but Paralyzed and Voiceless?" *Prog. Brain Res*, 150: 495–511.
Laurey, S., F. Perrin and S. Brédart (2007), "Self-consciousness in Non-Communicative Patients," *Consciousness and Cognition*, 16: 722–41.
Lewis, D. (1972), "Psychophysical and Theoretical Identifications," *Australasian Journal of Philosophy*, 50: 249–58.
Lloyd M. C. (1923), *Emergent Evolution*, London: Williams & Norgate.
Lowe, E. J. (2010), "Causal Closure Principles and Emergentism," *Philosophy*, 75 (294): 571–86.
Lowe, E. J. (2013), "The Probable Simplicity of Personal Identity," in G. Gasser and M. Stefan (eds), *Personal Identity: Complex or Simple*, 137–55, Cambridge: Cambridge University Press.
Meixner, U. (2009), "States of Affairs—The Full Picture," in M. E. Reicher (ed.), *States of Affairs*, 51–70, Frankfurt: Ontos Verlag.
Mill, J. S. (1943), *A System of Logic*, London: John W. Parker, West Strand.
Naci, L., L. Sinai and A. M. Owen (2017), "Detecting and Interpreting Conscious Experiences in Behaviorally Non-responsive Patients," *NeuroImage*, 145: 304–13.
Nida-Rümelin, M. (2013), "The Non-descriptive Individual Nature of Conscious Beings," in G. Gasser and M. Stefan (eds), *Personal Identity: Complex or Simple*, 157–76, Cambridge: Cambridge University Press.
Noirhomme, Q., R. Brecheisen, D. Lesenfants, G. Antonopoulos and S. Laureys (2017), "'Look at My Classifier's Result': Disentangling Unresponsive from (minimally) Conscious Patients," *NeuroImage*, 145: 288–303.
Olson, K. R. (1987), *An Essay on Facts*, Stanford, CA: CSLI Publications.
Princess Elizabeth of Bohemia and R. Descartes (2007), *The Correspondence between Princess Elizabeth of Bohemia and René Descartes*, ed. and trans. L. Shapiro, Chicago: University of Chicago Press.
Quine, W. Van O. (1960), *Word and Object*, Cambridge, MA: The MIT Press.
Reicher, M. E. (2009), "Introduction," in M. E. Reicher (ed.), *States of Affairs*, 7–38, Frankfurt: Ontos Verlag.
Roy, J.-M., J. Petitot, B. Pachoud and F. Varela (1999), "Beyond the Gap: An Introduction to Naturalizing Phenomenology," in J. Petitot, F. Varela, B. Pachoud and J.-M. Roy (eds), *Naturalizing Phenomenology: Issues in Contemporary Phenomenology and Cognitive Science*, 1–82, Stanford, CA: Stanford University Press.
Robertson, K. R. (2004), "Translator's Introduction," in trans. K. R. Peterson, *First Outline of a System of the Philosophy of Nature*, xi–xxxv, Albany, NY: SUNY.

~~Schelling, F. W. J. (1976ff), *Werke: Historisch-kritische Ausgabe*, ed. I. M. Baumgartner, W. G. Jacobs and H. Krings, Stuttgart: Fromann-Holzboog.~~
Schelling, F. W. J. (2004), *First Outline of a System of the Philosophy of Nature*, trans. K. R. Peterson, Albany, NY: SUNY.
Sellars, W. (1956), "Empiricism and the Philosophy of Mind," in H. Feigl and M. Scriven (eds), *The Foundations of Science and the Concepts of Psychology and Psychoanalysis: Minnesota Studies in the Philosophy of Science*, Vol. 1, 253–329, Minneapolis: University of Minnesota Press.
Shani, I. (2021), "Eden Benumbed: A Critique of Panqualityism and the Disclosure View of Consciousness," *Philosophia*. Available online: https://doi.org/10.1007/s11406-021-00377-9.
Siderits, M. (2003), *Personal Identity and Buddhist Philosophy*, Aldershot: Ashgate.
Siderits, M. (2011), "Buddhas as Zombies: A Buddhist Reduction of Subjectivity," in M. Siderits, E. Thompson and D. Zahavi (eds), *Self, No Self: Perspectives from Analytical, Phenomenological, and Indian Traditions*, 308–332, Oxford: Oxford University Press.
Smart, J. J. C. (1959), "Sensations and Brain Processes," *Philosophical Review*, 68: 141–56.
Stapp, H. (1993), "A Quantum Theory of the Mind-Brain Interface," in H. Stapp (ed.), *Mind, Matter, and Quantum Mechanics*, 145–71, Berlin: Springer.
Stapp, H. (2007), *Mindful Universe*, Berlin: Springer.
Strawson, G. (2006), "Realistic Monism: Why Physicalism Entails Panpsychism," in A. Freedman (ed.), *Consciousness and Its Place in Nature*, 3–31, Exeter: Imprint Academic.
Swinburne, R. (2013), "How to Determine Which Is the True Theory of Personal Identity," in G. Gasser and M. Stefan (eds), *Personal Identity: Complex or Simple*, 105–22, Cambridge: Cambridge University Press.
Swinburne, R. (2018), "Mental Causation Is Really Mental Causation," in M. P. Guta (ed.), *Consciousness and the Ontology of Properties*, 174–85, London: Routledge.
Swinburne, R. (2019), "The Implausibility of the Causal Closure of the Physical," *Organon F*, 26 (1): 25–39.
Taber, J. (2003), "Dharmakīrti against Physicalism," *Journal of Indian Philosophy*, 31: 479–502.
Vetter, T. (1964), *Erkenntnisproblemebei Dharmakīrti*, Wien: Verlag der Österreichischen Akademie der Wissenschaften.
Voltolini, A. (2006), *How Ficta Follow Fiction. A Syncretistic Account of Fictional Entities*, Dordrecht: Springer.
von Neumann, J. (1955), *Mathematical Foundations of Quantum Mechanics*, trans. R. T. Beyer, Princeton, NJ: Princeton University Press.
Waldron, W. S. (2003), *The Buddhist Unconscious: The Ālaya-vijñāna in the Context of Indian Buddhist Thought*, London: Routledge.
Woodard, B. (2019), *Schelling's Naturalism Space, Motion and the Volition of Thought*, Edinburgh: Edinburgh University Press.