Consciousness and Mind

Some of the oldest and deepest questions in philosophy fall under the umbrella of consciousness and mind: What is the mind and how is it related to the body? What provides our thoughts with content? How is consciousness related to the natural world? Do we have distinctive causal powers? Analytic philosophers have made significant progress on these and related problems in the last century. Given the high volume of work on such topics, this chapter is necessarily selective. It offers major touchstones but is slanted in favor of work that touches base with the sciences. The chapter starts by describing the progression of thought on the mind-body problem, from dualism and behaviorism to non-reductive materialism. It then describes the problem of intentionality, with a focus on partial solutions from Dretske and Millikan, ending with a brief discussion of 4E theories of mental content. The section on the problem of consciousness starts with the well-known knowledge and modal arguments before describing some attempted initial solutions, such as eliminative materialism and panpsychism. Finally, there is a brief section on agency and free will, which focuses on the link between free will and consciousness.

1. Mind and Body

The nature of the mind and its relation to the body is a fundamental issue in philosophy that informs many others. While most people are *dualists*, believing the mind to be distinct from the body in some way (perhaps allowing it to survive bodily death, for example), most philosophers are *physicalists*, believing the mind to be physical like the body (Bourget & Chalmers, n.d.; Slingerland & Chudek, 2011). The subjectivity of the mind problematizes this commitment to physicalism, leading to extensive discussion in the field about how to reconcile the apparent conflict (especially in the case of consciousness, discussed in section 3). Nagel famously argues, for instance, that we do not yet have the tools to reconcile the subjective with the objective, dramatized in his example of how we are unable to reconcile our external observations about bats with an understanding of "what it is like" to be a bat (Nagel, 1974). Yet, such reconciliation is thought by many to be an essential step in bringing philosophy into alignment with the sciences.

Dualism has at least two significant strands: *substance* and *property*. Substance dualism is usually attributed to Descartes, who famously argued for the separation of thought and extension—whereas extended substances take up space, thinking substances do not (Rodríguez Pereyra, 2008). Put another way, whereas a sparkling waterfall takes up space, imagining a sparkling waterfall does not (or does not seem to). Thus, a sparkling waterfall is a mode of an extended substance, but imagining a sparkling waterfall is a mode of a thinking substance. A substance dualist would say that the mind is separable from the body in that it is made up of different stuff or substance.

A famous difficulty with substance dualism is known as the *interaction problem*, a version of the *problem of mental causation*, introduced to Descartes by Elisabeth of Bohemia: "How can something immaterial and non-extended move something material and extended?" (L. Shapiro, 1999, p. 505). As Elisabeth pointed out, our typical understanding of causation is

through proximal interaction, but something that does not exist in space cannot be said to be proximal to something that does exist in space (it cannot be said to be proximal to anything at all). This problem is exacerbated by apparent regular interaction between the mind and body, as when imagining a sparkling waterfall causes the body to relax. How could such an occurrence in a non-spatial substance cause such a change in a spatial substance? Most philosophers reject substance dualism due to this problem.

Property dualism, on the other hand, is more widely accepted among philosophers, and allots two essentially distinct types of properties, mental and physical. It does not require the interaction of a spatially extended substance with a non-extended substance since both types of properties inhere in a single substance. The imagined sparkling waterfall may have both subjective and objective properties, both of which inhere in a single spatially extended substance (e.g. the brain). It is nonetheless mysterious how these different types of properties are related and how they might interact, if at all, which some have argued makes property dualism no better off than substance dualism (Lycan, 2013). Closely related to property dualism are *neutral monism* and *dual-aspect theory*, which both hold that what we understand as the mental and the physical are connected to a common substance that is neither mental nor physical (Stubenberg, 2018).

As is mentioned above, most philosophers reject dualism in favor of physicalism: the view that everything in the universe, including the mind, is physical. This view is said to go beyond the more traditional *materialism*, according to which everything is made up of matter, to include other ways of thinking about physical existence (e.g. energy). Physicalism is, instead, constrained to that which is or might be described by physics. Because what is described by physics continues to change, physicalism might be seen as an *attitudinal*, rather than an ontological claim—we ought to try to make our account of the mind consistent with our account of the rest of the universe, even if we don't yet have a final theory of just what the universe consists in (Ney, 2008).

In the early 20th century *behaviorism* espoused this attitudinal physicalism in seeking to explain the mind only through what can be externally observed, such as changes to behavior as a result of changes to a stimulus; in behaviorism "both sensation and perception may be analyzed as forms of stimulus control" (Skinner, 1963, p. 955). The imagined sparkling waterfall would thus be defined in terms of only input and output—the input of the text on the page and the output of bodily relaxation. In psychology, the resulting attempts at careful and systematic observation of human behavior led to substantial progress in terms of prediction and control of that behavior. Yet, not everything that makes a difference to behavior can be directly observed; whether or not someone is paying attention to the words will make a difference to whether they effectively imagine the sparkling waterfall and subsequently relax, but the impact of attention occurs between the stages of input and output (it is "internal"). Experimental findings on attention and other mental phenomena in the mid-20th century led to the so-called "cognitive revolution" and the rejection of behaviorism in favor of approaches that posited descriptions of the internal workings of the mind (Miller, 2003).

The middle of the 20th century thus saw the development of *identity theory* and *functionalism* within philosophy. Identity theory commits itself to an identity relation between the mind and physical states (typically brain states): "the identity theory says that experience-ascriptions have the same reference as certain neural-state-ascriptions" (Lewis, 1966, p. 19). The imagined sparkling waterfall might be identified with neural firing in a particular region of the brain, for example. A criticism of this theory from Putnam (1967) is that, plausibly, a mental state can be supported or "realized" by multiple brain states, and even non-brain states ("multiple realizability"). Two very different brains, with very different brain states, might imagine the same sparkling waterfall, for instance. Putnam suggested functionalism as an alternative approach.

Functionalism commits itself to an identity relation between the mind and functional states, allowing mental states to be realized in different physical mediums (Block, 1980). The imagined sparkling waterfall has the function of, for example, causing the input of the text to lead to the output of bodily relaxation. If a computer chip built out of silicon were able to replicate this function, then the imagined sparkling waterfall would be realized by this neuromorphic chip. Of course, whether functionalism allows for such multiple realization depends on the function in question; very fine-grained functions (e.g. the function of maintaining the blood-brain barrier) are more likely to depend on a specific physical realization (e.g. astrocytes, a type of cell in the brain), whereas very rough-grained functions (e.g. the function of detecting stimuli) are more likely to find expression in multiple physical substrates (Cao, forthcoming).

The latter half of the 20th century saw the rise of *non-reductive physicalism.* In non-reductive approaches, the mind has some degree of independence from its physical constituents. The imagined sparkling waterfall, while physical, would have some degree of independence from the neural firing that supports it. The type and degree of independence varies widely across theories. The weakest form is descriptive independence, which found early expression in Davidson's "anomalous monism." Anomalous monism holds that "all events are physical" (hence "monism") but rejects "that mental phenomena can be given purely physical explanations" due to a lack of law-like connections (hence "anomalous") between descriptively mental phenomena and physical phenomena (Davidson, 1980, p. 141). Thus, the mind is physical but mental descriptions cannot be reduced to physical descriptions.

Importantly, almost all forms of non-reductive physicalism, including anomalous monism, make use of the concept of *supervenience* to maintain some degree of dependence between the mind and its physical constituents. Supervenience is a relation between two sets such that a change in the first set is required to get a change in the second (the second set "supervenes" on the first), but not vice versa. If the imagined sparkling waterfall supervenes on its physical constituents, then it can't change (e.g. to include a rainbow) without a change in the physical constituents (e.g. different neural firing), but the physical constituents can change without having an impact on the waterfall. Non-reductive physicalists typically hold that the mind supervenes on the body, such that a change in the mind only occurs through a change in the body (but a change in the body can occur without a corresponding change in the mind).

DRAFT

Stronger forms of non-reductive physicalism have faced significant problems, in part due to this commitment to supervenience. Such views typically invoke some sort of mental power that goes beyond the physical constituents. This forces them to confront a new version of the problem of mental causation, raised by Kim: supervenient mental causation would either violate the causal closure of the physical or require overdetermination of the caused events. The causal closure of the physical is the idea that every physical event has a sufficient physical cause, leaving no room for additional mental causes. If we preserve causal closure then any instance of supervenient mental causation would count as a case of excessive overdetermination: an event with two separate causes, one mental and one physical (Kim, 2007).

Crucial to Kim's argument is the non-reductive physicalist's commitment to supervenience: if a change to the mental requires a change to its physical constituents, then any mental power will rely on the powers of its physical constituents, making that mental power redundant. Some have bypassed Kim's reasoning by rejecting local supervenience, such that mental powers can exceed the powers of local physical constituents. That is, mental powers might exceed the powers of the brain, even while they do not exceed the powers of the brain, body, and environment taken together over an extended period of time (the "global" physical constituents). Baker, for example, has pointed out that one can preserve causal closure through global supervenience while also leaving room for mental causation through the denial of local supervenience (Baker, 2009). Perhaps the mind emerges from its local physical constituents only in certain contexts, such that its causal power is not reducible to its local base (Jennings, 2020a). More recent discussions on mental powers have revived these and related issues (Grasso & Marmodoro, 2020).

While physicalist approaches grew in popularity over the 20th century, some have noted that they fail to adequately capture the subjectivity described by Nagel—the "what it's likeness" of experience. This basic idea led to what we might call the "consciousness revolution" of the late 20th century. In short, several prominent arguments challenged the idea that physicalism can capture all aspects of conscious experience, leading to a revival of many of the above issues under a new heading. These arguments are discussed in section 3.

2. Intentionality and Content

The "mind-body problem" introduced in section 1 concerns the general difficulty of aligning our understanding of the mind with that of the physical world, and includes within it two other problems: the *problem of intentionality* and the *problem of consciousness*. This section focuses on the former.

Intentionality is often cast as "aboutness"—when we imagine a sparkling waterfall, our thoughts are *about* a sparkling waterfall. In the late 19th century Brentano introduced the concept: "Every mental phenomenon is characterized by…the intentional (or mental) inexistence of an object…reference to a content…Every mental phenomenon includes something as object within itself..." (Brentano as cited in Huemer, 2019). To say that every mental phenomenon is characterized by intentionality is to say that every mental

phenomenon is about something, its "content." Since the sparkling waterfall is only imagined, an "object within," at least some of that content is mental. Thus, the problem of intentionality is explaining how this mental content is consistent with the natural world.

Mental content is typically understood through the concept of *representation*. A representation is something that carries information. When I imagine a sparkling waterfall, I can be said to have a representation of the sparkling waterfall in my mind: something that carries the information consistent with a sparkling waterfall that enables me to say that it is a sparkling waterfall. But what is the nature of these representations, and how do they represent the world? The work of Turing and Shannon in the early 20th century was a first step in understanding information and representation in physical terms.

The cognitive revolution, introduced in section 1, happened at around the same time as the growth of computing (the "digital revolution"). An early model for the computer was the "Turing machine," a hypothetical machine posed by Turing that gathers input and then manipulates that input using algorithms to generate an output (van Leeuwen & Wiedermann, 2001). Similarly, we might see the mind as gathering input through the senses and then manipulating that input to provide behavioral outputs. Shannon's pioneering work, published at around the same time as Turing's, allows us to call that which is gathered, manipulated, and then expressed by the Turing machine "information" (Guizzo, 2003; see also Chapter 13). Thus, we might say that the mind, like the computer, is in the business of processing information (the mind is "computational"). This "mind as computer" metaphor led to numerous innovations in both the study of the mind (especially through the development of the new field of cognitive science) and computer science (especially through the development of new information technologies).

Shannon information is a first step to naturalizing intentionality, but it doesn't fully capture the mental content discussed by Brentano. It seems unlikely, for example, that the manipulation of input would allow the computer to represent something that is absent, as with the sparkling waterfall. Whereas philosophers work to bridge the conceptual gap between Shannon information and Brentano intentionality, scientists for the most part assume this gap will be bridged, and wonder simply what form representations will take: "no one but the antiquated behaviorist doubts that the brain does represent. The interesting questions about representation are, from the scientist's perspective, ones like the following: are mental images represented in a quasi-pictorial format, or are they represented sententially?" (Shapiro, 1997).

Along these lines, the late 20th century saw the rise of naturalistic theories of representation, especially those of Dretske and Millikan. These philosophers wanted to show how it is possible for biological organisms to represent features of their environments even when those features are absent. A start is to look at features that are present but misrepresented. Dretske famously considers the case of bacteria that can only live "in the absence of oxygen" and so use internal magnets ("magnetosomes") to infer the absence of oxygen: "in the bacteria's normal habitat, the internal orientation of their magnetosomes means that there is relatively little oxygen in *that* direction…[but] in the presence of the bar magnet…the organism's sensory state misrepresents the location of

oxygen-free water" (Dretske, 1993, p. 27). That is, the bacteria have developed a way to track a feature of their environment in accordance with their needs (i.e. low oxygen), but indeterminacy in their tracking system (magnetic orientation can indicate oxygen levels but also other environmental features) leads to misrepresentation when the environment changes in unexpected ways. Thus, Dretske argued that indeterminacy is at the heart of misrepresentation in biological creatures.

Millikan further explains the fact of indeterminacy through the concept of representation "consumers": "representation consumers are devices that have been designed…by a selection process to cooperate with a certain representation producer. The producer, likewise, has been designed to match the consumer" (Millikan, 1990, p. 153). Millikan's account is intended to be consistent with the biological sciences, and especially evolutionary theory. In her account, the oxygen-hating bacteria (the "consumers") have evolved in a specific environment (the "producer"), and can thus be tricked in an abnormal environment, such as in the presence of a bar magnet: "What the magnetosome represents is only what its consumers require that it correspond to in order to perform their tasks"(Millikan, 1989, p. 290). In order to show how sophisticated these evolved representations can be, she famously uses the example of how bees "dance" to represent the location of pollen to one another: "Variations in the tempo of the dance and in the angle of its long axis vary with the distance and direction of the nectar…So, the dances are representations of the location of nectar" (Millikan, 1989, p. 288). Just as the bacteria can incorrectly represent the absence of oxygen, bee dances can incorrectly represent the location of pollen, a start to understanding mental representations in the absence of a stimulus.

Millikan sees the presence of a consumer as a necessary part of any theory of information in order to separate simple correlations in nature from a rich sense of information—something carried by a representation. The sparkles in a waterfall correspond with light hitting the water at a certain angle, but neither the sparkles nor the angled light contain information about the other; they simply co-occur. How does the simple correspondence of "natural information" differ from the information contained in minds? For Millikan, this comes down to fact that the mind is an information consumer: "Suppose, for example, that there were abundant 'natural information'… This information could still not serve the system as information, unless the signs were understood by the system, and, furthermore, understood as bearers of whatever specific information they, in fact, do bear" (Millikan, 1989, p. 286). In this way of thinking about information the gap between Shannon information (simple correspondence) and Brentano intentionality (mental content) comes down to the presence or absence of an information consumer, which is subject to evolutionary pressures.

Discussions concerning intentionality and content are very rich in philosophy, and go far beyond what is mentioned here, but I want to explore just one more thread for the sake of this chapter: 4E theories of mind. In general, 4E theories draw attention to that which was traditionally seen as "external" to the mind. The four "Es" are *embodied*, *extended*, *embedded*, and *enactive*: "embodied" theories take the body to be constitutive of mental life, "extended" theories take the mind to extend into the surrounding environment,

DRAFT

"embedded" theories see the mind as embedded within the environment, and "enactive" theories see action as at the heart of all that is mental. The bee dance serves as a good illustration of an externalized representation: as Millikan puts it, "Any representation the time or place of which is a significant variable obviously cannot be stored away, carried about with the organism for use on future occasions" (Millikan, 1989, p. 295). By externalizing mental content, 4E theories aim to close the gap with the natural world.

A more recent example of externalized mental content is provided by Orlandi, who denies that perceptual content must always involve internal representations. Instead, they argue that a visual system can rely on stable features of the environment, using the example of fire alarms: "Fire alarms seem to assume that smoke is typically caused by fire. This is why, when they detect smoke, they signal the presence of fire. But it is implausible to describe fire alarms as actually knowing anything about such regularity" (Orlandi, 2012, p. 561). In their view, the visual system can be built from visual feature detectors that act like fire alarms, signaling to other cognitive areas the presence of a visual feature without actually containing any information about that visual feature. This view may appear to be inconsistent with the computational approach, since it denies that information is stored in the brain's early visual system. Yet, Orlandi sees the views as compatible: "the embedded view I favor, and ecological views more generally, are compatible with computationalism. The development of new kinds of computational systems is precisely what inspires non-cognitive understandings of the visual act" (Orlandi, 2014, p. 5).

4E perspectives have revolutionized the ways that philosophers think about mental content. Yet, even those who argue for more radical 4E approaches note that "it is actually very difficult to reject internal representations" (Chemero, 2009, p. x). Our ability to imagine a sparkling waterfall without the presence of such a waterfall, for instance, seems to depend on information we carry about the waterfall, and thus some sort of internal representation. Memory, imagination, hallucinations, and dreams all challenge a fully externalized picture of mental content. As Chemero puts it, "it is still an open question how far beyond minimally cognitive behaviors radical embodied cognitive science can get"(Chemero, 2009, p. 43).

3. Consciousness

Alongside the problem of intentionality, the 20th century saw the rise of the problem of consciousness. The definition of "consciousness" is fraught, but we can follow Velmans by starting with an "ostensive" definition, in which we point to the phenomenon without describing all of its qualities: "We know what it is like to be conscious when we are awake as opposed to not being conscious when in dreamless sleep" (Velmans, 2009). Normally, this "what it's like" is taken to include the subjectivity of experience—the difference between being awake and in dreamless sleep is a difference that means the most to the person (or "subject") in question. But being characterized by subjectivity seems to set consciousness apart from the rest of the natural world. The problem of consciousness is thus to naturalize subjectivity, to explain how consciousness fits within a scientific worldview. This problem is separable from the mind-body problem for those who accept

the existence of the unconscious mind, and is separable from the problem of intentionality for those who believe in consciousness without content.

An oft-cited paper on this topic is Nagel's, mentioned in section 1, which introduces the language of "what it's like." Yet, work on the topic is varied, voluminous, and of course precedes Nagel. As early as 1904 James argued that we should do away with the concept of consciousness, seeing it as unscientific: "it is the name of a nonentity…those who still cling to it are clinging to a mere echo, the faint rumor left behind by the disappearing 'soul' upon the air of philosophy" (James, 1904, p. 477). Yet, at around the same time, at the first meeting of the American Association for the Advancement of Science, Ribot argued that consciousness is the "oldest problem of philosophy and one of the youngest problems of science," hypothesizing that "there are two fundamentally different things in the universe, force and consciousness," both of which should be explored (Minot, 1902, p. 12). One might see these as early versions of *eliminativism* and *panpsychism*, respectively. For the sake of this chapter, I will focus on these and just a few other approaches to this difficult problem.

A few arguments serve as background to contemporary accounts of consciousness. Some of these are known as "knowledge arguments," since they contrast the type of knowledge we have of consciousness with the type of knowledge we have of the objective physical world. Nagel's argument is one of the most famous examples, and posits a gap in our understanding of subjective experience and objective facts: "if the subjective character of experience is fully comprehensible only from one point of view, then any shift to greater objectivity—that is, less attachment to a specific viewpoint—does not take us nearer to the real nature of the phenomenon: it takes us farther away from it" (Nagel, 1974, p. 445). He frames it as a conceptual problem; we cannot conceive what it would mean for the subjective, and especially consciousness, to be identical to something objective, as physicalism requires.

Just 8 years later Jackson framed another well-known argument in this vein, which took the point further: physicalism isn't just difficult to conceive, but false. Jackson reasons about two cases, Fred and Mary, the latter of which is now widely discussed. Fred can see more colors than we can and can sort what we see as identically red things into two groups, red_1 and red_2. "What kind of experience does Fred have when he sees red_1 and red_2? What is the new colour or colours like? We would dearly like to know but do not; and it seems that no amount of physical information about Fred's brain and optical system tells us" (Jackson, 1982, p. 129). Mary, on the other hand, is a color expert who knows all the physical facts about color but has never seen color, since she is confined to a black and room. Jackson asserts that Mary will learn something new when she experiences color for the first time, "but she had *all* the physical information. Ergo there is more to have than that, and Physicalism is false" (Jackson, 1982, p. 130). While Fred's case seems similar to that of Nagel's bats, Jackson sees both cases as demonstrating that physical facts cannot capture experiential ones, and thus the falsity of physicalism.

Another form of argument used against physicalism is known as the "modal argument." While this form of argument has been around at least as long as Nagel's bats, it was made most famous in the work of Chalmers, and is now known by many as the "zombie

argument."[1] The argument is essentially that we can imagine a world physically identical to ours but in which our physical duplicates are not conscious, thus in at least one respect consciousness is not physical (the respect that allows it to be conceivably absent in a physically identical world). Chalmers calls our non-conscious physical duplicates "phenomenal zombies" (hence the "zombie argument"). He uses this argument and others to coin a famous distinction between "easy" problems about consciousness and the "hard problem": "How does the brain process environmental stimulation? How does it integrate information? How do we produce reports on internal states? These are important questions, but to answer them is not to solve the hard problem: Why is all this processing accompanied by an experienced inner life?" (Chalmers, 1996, p. xii). In other words, given the possibility of phenomenal zombies, we are missing an explanation of why there is consciousness at all.

As mentioned above, one approach to this problem is to call for major revision or even elimination of the concept. The movement of "eliminative materialism" is most associated with Paul and Patricia Churchland, the former of whom argued "that our commonsense conception of psychological phenomena constitutes a radically false theory…that theory will eventually be displaced, rather than smoothly reduced, by completed neuroscience" (P. M. Churchland, 1981). While consciousness is not included as one of the psychological phenomena to be replaced in this early work on the topic, Patricia later writes that we should treat consciousness like other psychological functions and aim to understand it through reductionist neuroscience: "it is practical to earmark even our fondest intuitions about mind/brain function as revisable hypotheses rather than as transcendental absolutes or introspectively given certainties" (P. S. Churchland, 1994). Another, related approach is to treat the subjectivity of consciousness as an illusion, a view that is most strongly associated with Dennett: "conscious experience has *no* properties that are special in *any* of the ways qualia have been supposed to be special" (Dennett, 1988). Frankish has dubbed this view "illusionism" (Frankish, 2016).

In contrast with this perspective is one in which the subjectivity of consciousness is described as a fundamental feature of the physical world. One such approach links consciousness with life: "subjectivity and consciousness have to be explicated in relation to the autonomy and intentionality of life" (Thompson, 2010, p. 15). Thompson does this by demonstrating that the autonomy of all living beings is a physical phenomenon, dubbed "autopoiesis": "every molecular reaction in the system is generated by the very same system that those molecular reactions produce" (Thompson, 2010, p. 92). The interiority created by autopoiesis is a start to understanding consciousness, Thompson argues, that

[1] Jackson, for example, describes it this way: "there is a possible world with organisms exactly like us in every physical respect (and remember that includes functional states, physical history, et al.) but which differ from us profoundly in that they have no conscious mental life at all. But then what is it that we have and they lack? Not anything physical ex hypothesi. In all physical regards we and they are exactly alike. Consequently there is more to us than the purely physical" (Jackson, 1982, p. 130) He sets this argument aside as less convincing than Fred or Mary, since some report being unable to conceive of this world.

DRAFT

allows us to bypass the hard problem and replace it with the "body-body problem": "trying to understand a lived body as a special kind of autonomous system, one whose sense-making brings forth, enacts, or constitutes a phenomenal world" (Thompson, 2010, p. 237).

Another approach takes consciousness to be present even in the absence of life. This view has been labeled "panpsychism" and linked with scientists as early as Darwin: "subjectivity, albeit of an inconceivably primitive variety, is a feature of all matter—organic and inorganic" (Smith, 1978). This view has been more recently popularized by philosophers like Goff, who find it to be an elegant solution to the hard problem: "rather than trying to account for consciousness in terms of utterly nonconscious elements, one may try to explain the complex consciousness of humans and other animals in terms of simpler forms of consciousness which are postulated to exist in simpler forms of matter" (Goff, 2017). Yet, it faces the combination problem: how do individual bits of consciousness combine to create a larger unified consciousness? Without something like autopoiesis, it isn't obvious why bits of consciousness would form a complex, rather than remain individuated bits.

While there are many more stops on the tour of consciousness, I will end this section with just one more: epiphenomenalism. This is the view that consciousness does not make a causal difference in the physical world. That is, conscious states are caused by physical states but do not have the power to causally influence physical states in return. Epiphenomenalism is consistent with but distinct from panpsychism, since an epiphenomenalist need not hold consciousness to be present in all matter, and since a panpsychist might take the conscious elements present in all matter to have a causal role (Popper, 1977). Yet, the two are naturally paired, since panpsychism as a solution avoids placing consciousness within the physical world as it is known, which tends to be seen as causally closed (see section 1). This perspective is distinct from one in which consciousness has causal power, which is at the heart of the problem of free will, the topic of the next and final section.

4. Agency

Since it is covered in another chapter, the following discussion of free will is brief and focused on attempts to explain the experience of agency. In one way of looking at it, the problem of free will is that of accounting for the experience of agency (a "free will") in a determined universe. In a determined universe every event is necessitated by prior events. In other words, given past states of the universe, the present state of the universe could not have been otherwise. This is the doctrine of determinism, and it includes mental states and events. Yet, there is a sense in which mental events seem undetermined, especially in the case of choices: our choices seem up to us, and not determined by prior events. It feels up to us whether or not we imagine a sparkling waterfall, for example. Yet, if determinism is right then we could not have done otherwise. Because this problem is driven by a conflict between experiential evidence and a scientific worldview, it shares some overlap with the problem of consciousness. Yet, it is a separable problem requiring a separate solution.

The most significant progress on the problem of free will has come in the form of *compatibilism*, the stance most philosophers hold on the problem. Compatibilists aim to

resolve the apparent conflict between free will and determinism by reconsidering the meaning of these concepts. Hume was an early compatibilist who argued that freedom is nothing other than the ability to act on our desires (Hume, 1902, p. 95). If you want to imagine a sparkling waterfall, freedom is the ability to do so, whereas a lack of freedom prevents you from doing so. In this notion of freedom there is no conflict between freedom and determinism, and determinism is instead seen as supporting freedom: someone is freer if their desires determine their actions, and less free if their desires do not determine their actions. Yet, it is left unsaid in Hume's account how our desires are determined, and by whom or what. Arguably, we are less free if our desires are determined by others, so we want an account of freedom that includes not only freedom of action on the basis of our desires, but freedom in terms of the desires themselves (see, e.g., the discussion of the "classical compatibilism" interpretation of Hume in Russell, 2021).

Contemporary compatibilists have thus gone further, aiming to support the idea that our desires are up to us in some way. Frankfurt, for example, argues that we act freely when we are able to act on the desires we endorse, or the desires we wish to have (desires that are in line with our "higher-order" desires; Frankfurt, 1988). You might, for example, both want to imagine a sparkling waterfall and wish you didn't want this. Perhaps imagining sparkling waterfalls has become compulsive at this point in the chapter—perhaps you find the exercise silly but can't help yourself. In that case, on this account, you are less free than if you embraced the exercise, since while you are acting on your desires in imagining the sparkling waterfall, you are not acting on desires that you endorse. Of course, to be consistent with compatibilism we would want this endorsement to be itself determined—whether or not you endorse your desire to imagine the sparkling waterfall is not a real choice that you have. Explaining how our endorsements can be determined and yet up to us is a problem left unresolved in these accounts (see, e.g., two problems for a hierarchical theory in McKenna & Coates, 2021).

While compatibilists eschew the idea of the mind having independent causal power, libertarians embrace it, primarily due to evidence from experience. The experience of having causal power is nicely explicated in Kane's example of a businesswoman faced with a difficult choice: "She is on the way to a meeting important to her career when she observes an assault in an alley. An inner struggle ensues between her moral conscience, to stop and call for help, and her career ambitions, which tell her she cannot miss this meeting—a struggle she eventually resolves by turning back to help the victim" (Kane, 1999, p. 225). From the businesswoman's perspective it was not already determined that she would help the victim. Instead, she made it the case through her choice that she would help the victim. She experienced this choice as requiring effort on her behalf, solidifying her perspective that the choice was up to her. For Kane, reflection on this type of case provides evidence in favor of a libertarian approach, which rejects determinism in favor of a free will.

One way to reconcile libertarian and compatibilist approaches is through non-reductive physicalism, introduced in section 1. Determinism need not be committed to reductionism; it is possible to believe in determinism without also believing that all events are determined at the micro-physical level. List, for example, argues that "determinism at the

physical level does not rule out the possibility of doing otherwise at the agential level," since the agential level is not reducible to the physical level (List, 2014). An agent can freely choose between options, as required by libertarians, while the agent is made up of micro-physical processes that are determined. As was briefly discussed in section 1, one can hold onto causal closure of the physical and global supervenience while denying local supervenience, providing space for causal power on behalf of the emergent agents (see also Jennings, 2020b). This approach allows for reconciliation between the conscious experience of agency and our scientific worldview. While the problems associated with understanding and explaining agency are numerous, this work is yet another instance of how analytic philosophy has made progress on difficult problems in this domain.

5. Conclusion and Summary

This chapter has covered the four most significant strands in philosophy of mind: the mind-body problem, the problem of intentionality, the problem of consciousness, and the problem of free will. The tools of analytic philosophy—systematic analysis of concepts and careful argument—together with advancement in the sciences have allowed us to make progress on all four strands.

- On the mind-body problem, philosophers have moved on from the unworkable positions of substance dualism and behaviorism toward functionalism and non-reductive physicalism.
- On the problem of intentionality, explaining our ability to represent what is absent made substantial headway with the insight that biological consumers use indeterminate representations that can misrepresent their environments.
- The problem of consciousness, while somewhat recalcitrant, has had two significant developments, eliminativism and panpsychism, both of which have led to considerable re-evaluation of the problem.
- And, finally, compatibilists have helped to explain some of the intuitions behind the problem of free will, while non-reductive physicalists promise to close the gap even further.

In each of these cases it is clear that the work of philosophers is to both identify fail points and to find new solutions, moving the problem forward to its next stage.

Philosophy of mind also touches on numerous problems not covered here, and there are doubtless new problems on the horizon. The 21st century so far looks to bring even greater connection between philosophy of mind and the sciences, with more exploration of focused topics, such as artificial intelligence, attention, emotions, memory, perception, the unconscious, and social cognition (see, e.g., Young & Jennings, 2022). These topics come with their own problems and solutions while also inviting us to revisit our understanding of the "core" problems described in this chapter.

Bibliography

DRAFT

Baker, L. R. (2009, January 15). *Non-Reductive Materialism*. The Oxford Handbook of Philosophy of Mind. https://doi.org/10.1093/oxfordhb/9780199262618.003.0007

Block, N. (1980). Troubles with functionalism. *Readings in Philosophy of Psychology*, *1*, 268–305.

Bourget, D., & Chalmers, D. (n.d.). *Philosophers on Philosophy: The 2020 PhilPapers Survey*. 46.

Cao, R. (forthcoming). *Multiple Realizability and the Spirit of Functionalism (Synthese, forthcoming)*. Dropbox. https://www.dropbox.com/s/pkozc51oqxzszg4/MR%20for%20Synthese%20Nov%202021.pdf?dl=0

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

Chemero, A. (2009). *Radical Embodied Cognitive Science*. MIT Press. http://ebookcentral.proquest.com/lib/ucm/detail.action?docID=3339068

Churchland, P. M. (1981). Eliminative Materialism and the Propositional Attitudes. *The Journal of Philosophy*, *78*(2), 67–90. https://doi.org/10.2307/2025900

Churchland, P. S. (1994). Can neurobiology teach us anything about consciousness? *Proceedings and Addresses of the American Philosophical Association*, *67*(4), 23–40.

Davidson, D. (1980). *Mental Events,[in:] Essays on Actions and Events*. Clarendon Press, Oxford.

Dennett, D. C. (1988). Quining Qualia. In A. Marcel & E. Bisiach (Eds.), *Consciousness in Modern Science*. Oxford University Press. http://cogprints.org/254/

Dretske, F. (1993). Misrepresentation. *Readings in Philosophy and Cognitive Science*, 297–314.

Frankfurt, H. G. (1988). Freedom of the Will and the Concept of a Person. In M. F. Goodman (Ed.), *What Is a Person?* (pp. 127–144). Humana Press. https://doi.org/10.1007/978-1-4612-3950-5_6

Frankish, K. (2016). Illusionism as a Theory of Consciousness. *Journal of Consciousness Studies*, *23*(11–12), 11–39.

Goff, P. (2017). Panpsychism. *The Blackwell Companion to Consciousness*, 106–124.

Grasso, M., & Marmodoro, A. (2020). Introduction: Mental Powers. *Topoi*, *39*(5), 1017–1020. https://doi.org/10.1007/s11245-019-09680-3

Guizzo, E. M. (2003). *The essential message: Claude Shannon and the making of information theory* [Thesis, Massachusetts Institute of Technology]. https://dspace.mit.edu/handle/1721.1/39429

Huemer, W. (2019). Franz Brentano. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2019). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2019/entries/brentano/

Hume, D. (1902). *Enquiries Concerning the Human Understanding: And Concerning the Principles of Morals*. Clarendon Press.

Jackson, F. (1982). Epiphenomenal qualia. *The Philosophical Quarterly (1950-)*, *32*(127), 127–136.

James, W. (1904). Does "Consciousness" Exist? *The Journal of Philosophy, Psychology and Scientific Methods*, *1*(18), 477–491. https://doi.org/10.2307/2011942

DRAFT

Jennings, C. D. (2020a). Practical Realism About the Self. In L. R. G. Oliveira & K. Corcoran (Eds.), *Common Sense Metaphysics: Themes From the Philosophy of Lynne Rudder Baker*. Routledge. https://philarchive.org/rec/JENPRA-2

Jennings, C. D. (2020b). *The Attending Mind*. Cambridge University Press. https://doi.org/10.1017/9781108164238

Kane, R. (1999). Responsibility, luck, and chance: Reflections on free will and indeterminism. *The Journal of Philosophy*, *96*(5), 217–240.

Kim, J. (2007). *Physicalism, or Something Near Enough*. Princeton University Press.

Lewis, D. K. (1966). An Argument for the Identity Theory. *The Journal of Philosophy*, *63*(1), 17–25. https://doi.org/10.2307/2024524

List, C. (2014). Free Will, Determinism, and the Possibility of Doing Otherwise. *Noûs*, *48*(1), 156–178. https://doi.org/10.1111/nous.12019

Lycan, W. G. (2013). Is property dualism better off than substance dualism? *Philosophical Studies*, *164*(2), 533–542. https://doi.org/10.1007/s11098-012-9867-x

McKenna, M., & Coates, D. J. (2021). Compatibilism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2021). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2021/entries/compatibilism/

Miller, G. A. (2003). The cognitive revolution: A historical perspective. *Trends in Cognitive Sciences*, *7*(3), 141–144. https://doi.org/10.1016/S1364-6613(03)00029-9

Millikan, R. G. (1989). Biosemantics. *The Journal of Philosophy*, *86*(6), 281–297. https://doi.org/10.2307/2027123

Millikan, R. G. (1990). Compare and Contrast Dretske, Fodor, and Millikan on Teleosemantics. *Philosophical Topics*, *18*(2), 151–161.

Minot, C. S. (1902). The Problem of Consciousness in Its Biological Aspects. *Science*, *16*(392), 1–12.

Nagel, T. (1974). What is it like to be a bat. *The Philosophical Review*, *83*, 435–450.

Ney, A. (2008). Defining Physicalism. *Philosophy Compass*, *3*(5), 1033–1048. https://doi.org/10.1111/j.1747-9991.2008.00163.x

Orlandi, N. (2012). Embedded seeing-as: Multi-stable visual perception without interpretation. *Philosophical Psychology*, *25*(4), 555–573. https://doi.org/10.1080/09515089.2011.579425

Orlandi, N. (2014). *The Innocent Eye: Why Vision is Not a Cognitive Process*. Oxford University Press.

Popper, K. R. (1977). Some Remarks on Panpsychism and Epiphenomenalism. *Dialectica*, *31*(1/2), 177–186.

Putnam, H. (1967). Psychological predicates. In W. H. Capitan & D. D. Merrill (Eds.), *Art, Mind, and Religion*. University of Pittsburgh Press. pp. 37--48

Rodríguez Pereyra, G. (2008). Descartes's Substance Dualism and His Independence Conception of Substance. *Journal of the History of Philosophy*, *46*(1), 69–89. https://doi.org/10.1353/hph.2008.1827

Russell, P. (2021). Hume on Free Will. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2021). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2021/entries/hume-freewill/

Shapiro, L. (1999). Princess Elizabeth and Descartes: The union of soul and body and the practice of philosophy. *British Journal for the History of Philosophy*, *7*(3), 503–520. https://doi.org/10.1080/09608789908571042

DRAFT

Shapiro, L. A. (1997). The nature of nature: Rethinking naturalistic theories of intentionality. *Philosophical Psychology*, *10*(3), 309–322. https://doi.org/10.1080/09515089708573222

Skinner, B. F. (1963). Behaviorism at Fifty. *Science*, *140*(3570), 951–958.

Slingerland, E., & Chudek, M. (2011). The Prevalence of Mind-Body Dualism in Early China. *Cognitive Science*, *35*(5), 997–1007. https://doi.org/10.1111/j.1551-6709.2011.01186.x

Smith, C. U. M. (1978). Charles Darwin, the Origin of Consciousness, and Panpsychism. *Journal of the History of Biology*, *11*(2), 245–267.

Stubenberg, L. (2018). Neutral Monism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2018). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2018/entries/neutral-monism/

Thompson, E. (2010). *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Harvard University Press. https://www.hup.harvard.edu/catalog.php?isbn=9780674057517

van Leeuwen, J., & Wiedermann, J. (2001). The Turing Machine Paradigm in Contemporary Computing. In B. Engquist & W. Schmid (Eds.), *Mathematics Unlimited—2001 and Beyond* (pp. 1139–1155). Springer. https://doi.org/10.1007/978-3-642-56478-9_59

Velmans, M. (2009). How to define consciousness. *Journal of Consciousness Studies*, *16*(5), 139–156.

Young, B. D., & Jennings, C. D. (Eds.). (2022). *Mind, Cognition, and Neuroscience: A Philosophical Introduction*. Routledge. https://www.routledge.com/Mind-Cognition-and-Neuroscience-A-Philosophical-Introduction/Young-Jennings/p/book/9781138392366

DRAFT