

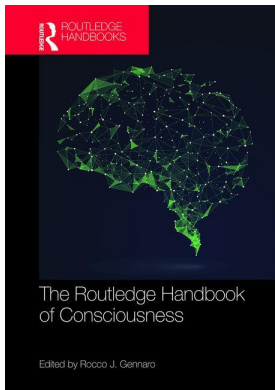
This article was downloaded by: 10.3.98.104

On: 04 Dec 2020

Access details: *subscription number*

Publisher: *Routledge*

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: 5 Howick Place, London SW1P 1WG, UK



The Routledge Handbook Of Consciousness

Rocco J. Gennaro

Materialism

Publication details

<https://www.routledgehandbooks.com/doi/10.4324/9781315676982-4>

Janet Levin

Published online on: 26 Mar 2018

How to cite :- Janet Levin. 26 Mar 2018, *Materialism from: The Routledge Handbook Of Consciousness* Routledge

Accessed on: 04 Dec 2020

<https://www.routledgehandbooks.com/doi/10.4324/9781315676982-4>

PLEASE SCROLL DOWN FOR DOCUMENT

Full terms and conditions of use: <https://www.routledgehandbooks.com/legal-notices/terms>

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

3

MATERIALISM

Janet Levin

1 Introduction

We humans have a great variety of conscious experiences: seeing the colors of the sunset, hearing thunder, feeling pain, tasting vegemite, hallucinating a dagger, or being in altered states of consciousness that are far from routine. It's hard to doubt, moreover, that many non-human animals have a variety of conscious experiences—some familiar, and some (e.g. the perceptual experiences of bats and octopuses) radically unlike any of our own. Nevertheless, there is a common feature, shared by all these states, that is essential to their being conscious experiences: they have a certain feel, or qualitative character; *there is something that it's like* to have them. Moreover, the distinctive *what it's like* to be in pain or hallucinate a dagger seems essential to their being conscious experiences of that type: one cannot be in pain, or hallucinate a dagger, unless one has an experience with a particular type of qualitative character, or feel.

Given this characterization of conscious experiences, the question naturally arises: what kinds of things could conscious experiences be, and what is their relation to the *physical* states and processes that occur in bodies and brains? One answer to this question, most closely associated with Descartes (1641), is that the locus of one's conscious experiences (and conscious thoughts) is an immaterial substance—a mind or (equivalently) a soul—that is distinct from, but able to interact with, bodies. A related view, held primarily by more contemporary theorists, is that while conscious mental states are states of the brain and body, their “feels” or qualitative features are special, non-physical, properties of those states. Both views are species of Dualism, the thesis that, in one way or another, the mental is distinct from the physical.

Dualism effectively captures the intuition that the qualitative features of conscious states and processes are radically unlike, and impossible to be explained by, any properties that occur elsewhere in the physical world, including neural processes such as the release of neurotransmitters, or the synchronized firing of certain neurons in the brain. As T.H. Huxley, a 19th-century Dualist, dramatically puts it (1881): “How it is that anything as remarkable as a state of consciousness comes about as a result of irritating nerve tissue, is just as unaccountable as the appearance of the Djinn, where Aladdin rubbed his lamp in the story.”

Almost as dramatically, G.W. Leibniz (1714) expresses a similar worry about any materialistic explanation of perception:

If we imagine that there is a machine whose structure makes it think, sense, and have perceptions, we could conceive it enlarged, keeping the same proportions, so that we could enter into it, as one enters a mill. Assuming that, when inspecting its interior, we will find only parts that push one another, and we will never find anything to explain a perception.

Nevertheless (and this is acknowledged even by its sympathizers), if Dualism were true, it would be hard to explain the occurrence of mental-physical *causation*. For example, I put my hand on the hot stove, I feel pain, I say “ouch”. This seems to involve a familiar causal sequence from physical to mental and then again to physical events, but it is hard to explain how a physical event could have effects on something non-physical—and even more seriously, how a non-physical state or event could have any sort of effect in the physical realm, given that we accept that every physical change that occurs in the world has a sufficient physical cause. There has been concern about mental-physical causation ever since Princess Elisabeth of Bohemia posed the question to Descartes in their (1643/1985) correspondence, and it has never been given a fully satisfactory answer.

Another serious question for Dualism concerns where, and how, consciousness arises on the phylogenetic spectrum in an otherwise physical world. Finally, Dualism raises epistemological worries: if conscious mental states or their qualitative properties are not physical, then they do not exist in space and cannot be perceived by anyone other than the subject who has them. But this means that we have no direct evidence that anyone other than ourselves ever sees the colors of the sunset, or feels pain, or for that matter has any conscious mental states at all—and, in addition, that scientists investigating the role of conscious mental states in the production of behavior have no way to determine which states are occurring in their subjects (if any) other than the introspective reports of those subjects themselves.

In contrast, the thesis of Materialism (often called “Physicalism”) maintains that there is nothing required for having conscious mental states and processes besides the occurrence of various types of physical states in the conscious creature’s brain and body (and perhaps in the world around it). It is easy to see why Materialism, in general, is an attractive view. If conscious mental states and processes can be fully characterized as various sorts of physical states and processes, then there is no need to explain how (or why) non-physical features arise in the natural world, and how they could be genuine causes of behavior. Materialism therefore seems to be a simpler and more economical theory than Dualism. In addition, if conscious mental states and their qualitative features are physical, then it is possible in principle for them to be observed by others.

On the other hand, there are well-known arguments, both classical and contemporary, that purport to show that *no* materialistic theory could provide an adequate account of the qualitative character of conscious experience, of *what it’s like* to see red or feel pain. Thus, although Materialism may seem to have promise for integrating mental states into the physical world, and connecting the study of mental states to the sciences of chemistry, biology, and neurophysiology, many contend that this cannot be done.

The primary goal of this chapter is to explore the prospects for a materialistic theory of conscious mental states and processes—or, more precisely, the prospects for a number of different materialistic theories that began to be proposed during the beginning of the 20th century—in particular, Behaviorism, the Type-Identity Theory, Functionalism, and (in passing) other versions of what has come to be known as Non-Reductive Physicalism.

This chapter will focus on the strengths and weaknesses of each of these theories—while considering whether or not any of them could explain how, as Huxley puts it, “anything as remarkable as a state of consciousness comes about as a result of irritating nerve tissue.” It will

also explore the viability of Eliminativism, the thesis that despite popular belief and the deliverances of introspection, our bodies and brains have no real and robust qualitative features at all.

Contemporary Materialism has antecedents in both the Classical and Modern periods. Leucippus (5th century BCE) and his student Democritus—and later Epicurus (341–270 BCE) and Lucretius (d.c. 50 BCE) all contend that everything that exists in the world can be explained as configurations of, and interactions among, atoms in the void. In the Modern period, Descartes's contemporary, Hobbes (1668/1994), and later La Mettrie (1747/1994), articulate what can be identified as materialistic theories of mental states. However, because the current debates about the pros and cons of Materialism focus primarily on the more contemporary versions of the doctrine, they will be the topics of discussion here.

2 Behaviorism

Behaviorism achieved prominence in the early to mid-20th century, both as a scientific theory of behavior (associated primarily with Watson, 1930, and Skinner, 1953) and as a philosophical theory of the meanings of our mental state terms or concepts. According to scientific behaviorism, the best explanation of human (and animal) behavior appeals not to a subject's internal mental states, but rather to its *behavioral dispositions*—that is, its tendencies to behave in certain specified ways given certain environmental stimulations, which are shaped by the contingencies of its past interactions with the environment. A major attraction of scientific behaviorism is its promise to explain behavior by appeal to states and processes that are indisputably physical, and also intersubjectively observable, rather than accessible (by introspection) only to the subjects of those mental states themselves.

In contrast, philosophical (or logical) behaviorism, associated primarily with Malcolm (1968), Ryle (1949), and more contentiously, Wittgenstein (1953), is not a scientific thesis subject to empirical disconfirmation, but rather the product of conceptual analysis. According to logical behaviorism, reflection on our mental state terms or concepts suggests that our ordinary claims about mental states and processes can be translated, preserving meaning, into statements about behavioral dispositions. For (an overly simplified) example, “S believes that it is raining” would be equivalent to “If S were to leave the house, S would take an umbrella, and if S had been heading to the car wash, S would turn around,” and “R is thirsty” would be equivalent to “If R were offered some water, then R would drink it quickly.”

However, as many philosophers have argued (see Chisholm 1957, Putnam 1968), statements about behavioral dispositions are unlikely to provide adequate translations of our claims about mental states, since, intuitively, a subject can have the mental states in question without the relevant behavioral dispositions—and vice versa—if they have other mental states of various sorts. For example, S could believe that it's going to rain, and avoid taking an umbrella when leaving the house if S *enjoys getting wet*—and S may take an umbrella, even if she does *not* believe it will rain, if she superstitiously believes that carrying an umbrella will prevent it from raining (or wants to assure her mother that she is planning for all contingencies). In short, the arguments continue, it is impossible to specify a subject's mental states as pure behavioral dispositions; they can only be specified as dispositions to behave in certain ways, given the presence or absence of *other* mental states and processes.

Similar worries have been raised (perhaps most influentially by Chomsky 1959) about the explanatory prospects of scientific behaviorism. Although scientific behaviorism had (and continues to have) some success in explaining certain types of learning, these successes, arguably, depend on the implicit control of certain variables: experimenters implicitly assume, usually correctly, that (human) subjects *want* to cooperate with them, and *understand* and *know how* to

follow the instructions; in the absence of these controls, it is unclear that the subjects would be disposed to behave in the ways that they do. It seemed to the critics of behaviorism, therefore, that theories that explicitly take account of an organism's beliefs, desires, and other mental states, as well as stimulations and behavior, would provide a fuller and more accurate account of why organisms behave as they do.

In addition, it seems that both experimental practice and conceptual analysis suggest that mental states are genuine *causes* of behavior: when I put my hand on a hot stove, feel pain, and say "ouch", my saying "ouch" is not a manifestation of a behavioral disposition, but rather an event produced by my feeling pain. Therefore, despite its attractions, most philosophers and psychologists have abandoned behaviorism and attempt to provide other materialistic theories of conscious mental states and processes. One such theory is the Type-Identity Theory, another is Functionalism; these will be the topics of the next two sections.

3 The Type-Identity Theory

The Type-Identity Theory, first articulated by U.T. Place (1956), H. Feigl (1958), and J.J.C. Smart (1959; also see his 2007), contends that for each type of mental state or process M, there is a type of brain state or process B, such that M is identical with B. For example, *pain* is identical with *C-fiber stimulation*. These claims are to be understood as property identities: *being a state of Type M is just being a state of Type B*—which entails that every instance of an M is a B, and vice versa. Thus, for the Type-Identity Theory to be true, there must be (at minimum) a *correlation* between instances of mental Type M (determined by the introspective reports of the individuals who are in them)—and physical Type B (determined by instruments such as brain scans).

Place, Feigl, Smart, and other early Type-Identity theorists recognized that the science of the time was nowhere near discovering any such universal correlations, but they were most concerned to establish, against intuitions (and arguments) to the contrary, that mental state–brain state identities are *possible*; that there are no logical or conceptual reasons to think that they could not be true. If these identities are possible, they argued, and if there are in fact correlations between instances of mental and physical states, then identity theorists could argue that the simplest and most economical explanation of these correlations—and the one that avoids the other difficulties of Dualism—is that the correlated mental and physical properties are identical.

Early identity theorists suggested that many objections to the possibility of mental–physical identities arise from the mistaken assumption that if mental–physical identity statements are true, then they should be knowable *a priori*; that is, solely by reason and reflection, without need for empirical investigation. They went on, however, to challenge this assumption, and to liken statements such as "Pain is C-fiber stimulation" to scientific identity statements such as "Lightning is electrical discharge" or "Water is H₂O"—statements that we believe to be true, but that can be known only *a posteriori*; only by appeal to observations of the world as it is.

However, early identity theorists also faced another important objection, the "Distinct Property Objection", articulated by Smart (1959), namely, that the only way that an *a posteriori* identity statement $A = B$ can be true is for both A and B to pick out their common referent by means of logically distinct (that is, conceptually unconnected) properties, or "modes of presentation," of that object that entail, respectively, its being an A and its being a B. For example, "water" picks out its referent as *the colorless odorless liquid that comes out of our faucets*; "H₂O" picks out its referent as *the compound of two hydrogen atoms and one oxygen atom*—and if, in fact, it turns out that the colorless odorless stuff that comes out of our faucets is composed of that compound of hydrogen and oxygen atoms, then we have an explanation of how "water is H₂O," though *a posteriori*, can be true.

However, the objection continues, in the case of mental–physical identities, the only sorts of properties that could entail being a *conscious mental* state of the relevant type (e.g. a pain, or an experience of a sunset) are *qualitative* properties (e.g. feeling a certain distinctive way, or being qualitatively reddish–orange). But then one can establish the identity of mental and physical states or processes only by attributing an irreducibly qualitative property to that state or process—and so one has not established a purely materialistic theory of conscious mental states.

Smart’s solution is to argue that mental state terms can be translated, preserving meaning, into “topic–neutral” terms, that is, terms that describe certain properties or relations that can be satisfied by *either* mental *or* physical states, processes, or events. He suggests, for example, that “I see a yellowish–orange after–image” can be translated into “There is something going on [in me] which is like what goes on when I have my eyes open, am awake, and there is an [unripe] orange illuminated in good light in front of me.” This term picks out a relational property that is “logically distinct” from any physical (or mental) property, and—if there really is a meaning equivalence between mental and topic–neutral terms—a state’s having that topic–neutral property will indeed entail its being a mental state of the relevant sort.

This particular suggestion for a topic–neutral translation, however, is generally regarded as unsatisfactory, since such topic–neutral terms are not sufficiently specific to serve as translations of our ordinary mental state terms. After all, many different mental states can be *like*, in some way or another, what goes on in me when I’m looking at an unripe orange; I could be having an after–image of a banana, or a perception of a faded basketball—or the thought that the orange juice I’m about to make for breakfast will be sour. One needs to say more about the *way* in which my having an experience is like what goes on when I’m seeing an unripe orange, and—as many have argued—it’s unclear that the relevant sort of resemblance can be specified in topic–neutral terms.

However, other Type–Identity theorists have attempted, with greater success, to provide topic–neutral equivalents of our ordinary mental state vocabulary; for example, David Armstrong (1981) attempts to characterize mental states in terms of their “aptness” to cause certain sorts of behavior. The most developed account of this sort is David Lewis’s (1966) suggestion that topic–neutral translations of our mental state terms can be extracted from our “common sense theory” of the mind, which can be understood to define mental states “all at once” by specifying (what we commonly believe to be) their causal interactions with environmental stimulations, behavior, and one another. For (an overly simplified) example:

Pain is the state that tends to be caused by bodily injury, to produce the belief that something is wrong with the body and the desire to be out of that state, to produce anxiety, and, in the absence of any stronger, conflicting desires, to cause wincing or moaning.

This way of characterizing mental states and processes is often called a *functional specification*, since it specifies the way these states, together, function to produce behavior. If this specification indeed provides a translation (or close enough) of “pain,” and if it is uniquely satisfied by C–fiber stimulation, then “pain = C–fiber stimulation” is true—and so on for other mental–physical identity statements. Moreover, Lewis explicitly argues, it would thereby be unnecessary to invoke simplicity or economy to establish the Type–Identity Theory: if these causal–relational descriptions indeed capture the meanings of our mental state terms, then any brain states that (uniquely) satisfy those descriptions will automatically be instances of those mental states.

Not surprisingly, there is skepticism about whether these sorts of “common sense” functional specifications can provide logically necessary and sufficient conditions for the occurrence of

conscious mental states. Isn't it possible, many ask, for a creature to satisfy such a specification, but not feel pain—or indeed not have any conscious mental states at all—or, conversely, for a creature to be in pain without satisfying the common sense specification? These questions are similar to the classic objections to logical behaviorism, and will be discussed further in Section 5. However, there is another worry about the Type-Identity Theory put forward by materialists themselves that needs to be addressed, namely, that it is too restrictive, or “chauvinistic,” in that it restricts the range of those who can possess mental states to humans, or at least mammals with neural structures similar to our own.

After all, it seems that there could be creatures that respond to the environment much like humans—who cry out when they're injured, and report feeling pain or hearing thunder in the same circumstances as we do, and whose other mental states interact with one another and with environmental stimulations like our own—but whose internal states are physically quite different from ours. Presumably, some argue, certain non-human animals (perhaps dolphins or octopuses) are like this, and certainly we can imagine silicon-based life forms with different types of physical states that satisfy the same functional specification as ours (think of androids, familiar from science fiction). But if some sort of experiential-neural identity thesis is true, then we could not consider these creatures to share our conscious mental states.

This worry has motivated some materialists to propose a related theory of what it is for someone to be in a particular type of mental state: Role Functionalism, or the Functional State Identity Theory. This theory will be addressed in the next section.

4 Role Functionalism

Consider (a fragment of) the functional specification presented earlier as a topic-neutral characterization of pain, namely, “*Pain* is the state that tends to be caused by bodily injury... and, in the absence of any stronger, conflicting desires, to cause wincing or moaning.” This specification depicts the causal role of pain in our so-called “common sense theory” of the mind, and may be satisfied, in humans, by C-fiber stimulation, and by different types of physical states in other, non-human, creatures. However, an alternative to maintaining that these other creatures are not in the same type of state as we are—or that pain is the *disjunctive* property that comprises whichever states satisfy the functional specification in different creatures—is to contend that pain is not to be identified with any particular type (or disjunction of types) of physical states that satisfy that description (or occupy that causal role), but rather with that causal role property itself.

Role Functionalism, that is, maintains that S is in pain just in case S is in the (higher-order) state of being in *one or another first-order state* that plays the causal role specified by the relevant functional description. Pain itself is to be identified with that higher-order state; those first-order states that occupy that role in some creature (e.g. C-fiber stimulation) are said to *realize* that state, and if different types of states can occupy the “pain role” in different creatures, pain is said to be *multiply realized*.

A major attraction of Role Functionalism, in contrast to the Type-Identity Theory, is that it permits humans, octopuses, silicon-based creatures—and even the non-biological but human-like androids familiar from science fiction—to count, literally, as being in the same mental state, as long as their first-order internal states occupy the same causal roles. Role Functionalism would thereby avoid the (alleged) human chauvinism of the Type-Identity Theory, although it would be compatible with a “token” identity theory, in which each *instance* (or token) of a mental state of some type (e.g. pain) is identical with an instance (token) of some type of physical state or other.

Role Functionalism, it should be noted, comes in two versions: one that derives from our “common sense” theory of the causal roles of mental states, and another (often called *Psychofunctionalism*; see Block 1980) that derives from empirical theories, developed by experimental psychologists and cognitive scientists, that include generalizations that may depart from the deliverances of common sense. Psychofunctionalist theories can provide more precise and detailed characterizations of mental states than our commonsense theories, which makes them less likely to be satisfied by systems (such as the economy of Bolivia; see Block 1980) that do not seem to have mental states at all. On the other hand, while psychofunctional characterizations can be topic-neutral, if specified solely in causal and relational language, they may not provide *translations*, however loose, of our mental state terms. Therefore the resulting identity statements linking mental and functional states will have no claim to being *a priori*, and thus may be subject to the “Distinct Property Objection.” Whether or not these identity statements—or any mental-physical identity statements—need to be *a priori* to avoid Dualism will be discussed later (in Section 5), but there is a further worry about Role Functionalism that threatens both versions of the view.

The worry is that Role Functionalism (like property Dualism) cannot account for the causal efficacy of mental states. Once again, it seems that if I put my hand on a hot stove, feel pain, and then say “ouch,” my feeling pain causes my saying “ouch.” However, if every physical event has a complete, sufficient physical cause, then my saying “ouch” will be caused by the physical, presumably neural, state that satisfies the functional specification of (or “realizes”) pain. But then my being in pain, if this is identified with a higher-order functional state, seems causally irrelevant. This is regarded as a problem not only for Role Functionalism (and property Dualism), but also for any materialistic view that treats the relation between mental and physical states as anything other than identity—for example, the view (Pereboom 2011) that mental states are *constituted by* physical states (in just the way that, as some suggest, a statue is constituted by, but not identical with, the material from which it is made).

Many Role Functionalists, in response, argue that this worry arises from the assumption that a genuine cause must “generate” or “produce” its effect, where this involves some sort of transfer of energy. However, they continue, this is not the only way to think about causation. Instead, causation should be regarded as a special sort of counterfactual dependence between effects and their causes (Loewer 2002), or as a special sort of regularity that holds between them (Melnyk 2003). If this is correct, then functional role properties and the physical events or states that realize them could *both* count as causally efficacious.

To be sure, property dualists could avail themselves of this defense as well. However, there is a further worry about causation (articulated by Kim 1989, 1998) that may differentiate the views, namely, that if mental and physical events (or properties) are both causally sufficient for producing behavior, then any behavior that has a mental cause would be causally *overdetermined*; that is, there would be more than one event that could have caused it by itself. But overdetermination occurs elsewhere in the world only rarely—for example, when two individuals simultaneously hit a window with a hardball, each with enough force to break it (or when more than one member of a firing squad hits the victim with lethal force)—and so it is counterintuitive to suggest that this is a routine and widespread occurrence in the causation of behavior.

One response to this worry (developed in different ways by Yablo 1992 and Bennett 2008) is to argue that the causation of behavior by a lower-level neural state and a functional role state does not fit the profile of classic overdetermination because lower-level neural states *necessitate* the functional states they realize; that is, if N is a realization of R, then necessarily, if some individual were to be in state N, then that individual would be in state R. If this is so, there is an explanation for the ubiquity of the production of behavior by both a mental and physical cause.

This response is available to Role Functionalists and other non-reductive physicalists such as those who maintain that mental states are constituted by physical states of various types. But this response would not be available to property dualists, who (usually) maintain that there is no necessary connection between mental and physical properties. Nevertheless, this response remains controversial—and thus the question of whether mental causation provides an insurmountable problem for Role Functionalism (or any materialistic theory other than the Type-Identity Theory) remains a matter of debate.

There are other recent theorists (Bechtel and Mundale 1999; Polger and Shapiro 2016) who attempt to “split the difference” between Type-Identity and Functionalism by arguing that Type-Identity Theory can achieve nearly as much universality as Role Functionalism, at least in its characterization of the mental states of actual existing creatures. These theorists argue, first, that a closer look at the functional organization of humans and other species such as dolphins and octopuses reveals that there is less functional similarity between these species and ourselves as philosophers once assumed. In addition, they continue, a closer look at the way neural states and processes are individuated in practice by neuroscientists shows that the neural states of different species that initially may seem to be quite different have certain properties in common that are more abstract or general—but are still decidedly physical, rather than functional. If this is so, then the Type-Identity Theory would allow for a greater range of creatures that could share the same mental states—but it still would not (presumably) include silicon-based life forms, or non-biological androids, as creatures capable of having mental states like our own. It remains a controversial issue among materialists whether an adequate theory needs to account for such creatures—and thus there is no consensus about which theory is most promising.

Moreover, as noted in Section 1, there are some well-known arguments directed against *all* materialistic theories of conscious mental states that must be considered in evaluating the pros and cons of Materialism. These arguments purport to show that no materialistic theory, no matter how detailed and comprehensive in specifying the internal structure of our physical states *and* their causal and other topic-neutral relations, can provide an adequate account of the qualitative character of conscious experience, of *what it's like* to see red, feel pain, or be in any other kind of conscious mental state. The best-known contemporary arguments against all forms of Materialism are the so-called Zombie Argument, presented by David Chalmers (1996, 2010), and the Knowledge Argument, presented by Frank Jackson (1982). (See Kripke 1980, Block 1980, and Searle 1980 for arguments similar to the Zombie Argument, and Nagel 1974 for an argument similar to the Knowledge Argument.) These arguments will be addressed in the next section.

5 General Arguments against Materialism

In the Zombie Argument, Chalmers defines a zombie as a molecule for molecule duplicate of a conscious human being—that is, a creature that is exactly like us both physically and functionally—but which has no conscious mental states whatever: there is *nothing that it's like* to be a zombie. He then argues as follows:

- 1 Zombies are conceivable.
 - 2 If zombies are conceivable, then zombies are genuinely possible.
 - 3 If zombies are genuinely possible, then Materialism is false.
- (C) *Therefore*, Materialism is false.

The general idea behind Premise (1) is that we can think of a body in all its physical (and functional) detail—and think about what it's like to be in a conscious state in all its *qualitative*

detail—and see no connection whatsoever between the two. The general idea behind Premise (2) is that such a radical disconnect between our conceptions of the physical and the qualitative is evidence that physical (including functional) and qualitative states and properties must be radically different types of things—and this is because what we can (carefully) conceive to be possible or impossible is our only source of knowledge about possibility and necessity; about what can, or cannot, be.

The Knowledge Argument, although superficially different, relies on similar ideas. Jackson describes a brilliant neuroscientist, Mary, who has been born and raised in a black-and-white room, but has nevertheless managed to learn all the physical and functional facts about human color experience via achromatic textbooks and videos. However, Jackson continues, it seems clear that if she were released from her room and presented with a ripe strawberry, she would be surprised by her experience and consider herself to have learned something new, namely, *what it's like* to see red. Jackson then argues as follows:

- 1 Mary knows all the physical and functional facts about human color experience while still in her black-and-white room, but does not know what it's like to see red (since she learns this only when she actually experiences red).
 - 2 If Mary knows all the physical and functional facts about human color experience before leaving the black-and-white room, but does not know what it's like to see red, then there is a fact about human color experience that is not a physical or functional fact.
 - 3 If there is a fact about human color experience that is not a physical or functional fact, then Materialism is false.
- (C) *Therefore*, Materialism is false.

Here too (Premise 1) the contention is that no amount of knowledge of the physical (and functional) features of the brains of those who are seeing colors could provide knowledge about the qualitative features of color experiences (and by analogy any type of state that there is *something it is like* to be in) and (Premise 2) that this lack of connection entails that there is something about these qualitative features that is different from anything physical (or functional).

To challenge these arguments, some materialists (e.g. Dennett 1988; Van Gulick 1993) and later Jackson himself, who (2004) eventually rejects the Knowledge Argument and its relatives, challenge Premise (1) of these arguments. They argue that although it may initially seem plausible that we can conceive of a zombie, on second thought this should seem implausible, since doing so would require that we have in mind, and be able to attend to, all the details of the physical structure and functional organization of our molecular duplicates, which is exceedingly hard to do. If we could do this, however, then we would recognize that such creatures were indeed having conscious mental states with qualitative properties just like our own. Similarly, they suggest, if Mary could internalize and concentrate sufficiently on all her physical knowledge about color experiences while still in her black-and-white room, then she would be able to know what it's like to have those experiences before she actually sees colors. These views maintain that there is an *a priori* link between our concepts of the qualitative and the physical (or functional), even though it may be difficult to discern. Chalmers (2002b) calls this *Type A Materialism*. He also discusses a related view—called *Type C Materialism*—which maintains that there are *a priori* connections between the qualitative and physical-functional features of our experiences, but that we haven't yet, or (McGinn, 1989) because of certain inescapable conceptual limitations *cannot*, form the concepts that are required to see that this is so.

However, many theorists—both dualist and materialist—(e.g. Chalmers 2002b; Stoljar 2001; Alter 2016) remain skeptical, and contend that learning, internalizing, and attending to more

physical and functional information about our brains and bodies could not possibly provide knowledge of what it's like to feel pain, see red or have any other sort of conscious mental state. The reason, they argue, is that physical and functional descriptions provide information solely about the "structure and dynamics" of what goes on in our brains and bodies, and these are all *relational* properties, whereas the distinctive qualitative features of our conscious mental states—as we can tell from introspection—are *intrinsic* properties.

Some Type A materialists question whether introspection reveals that the distinctive qualitative properties of conscious mental states are exclusively intrinsic—after all, they ask, would we really count an experience as *pain* if we didn't experience it as something we want to get rid of? And would we really count an experience as a *yellow-orange after-image* if we didn't experience its qualitative features as fading in certain ways over time, and being similar to and different from other color experiences? In short, they argue that the claim that the qualitative properties of experience are intrinsic is itself a product of inattentive (or biased) introspection.

However, there are other materialists—in Chalmers's locution, *Type B Materialists* (e.g. Loar 1997; Hill and McLaughlin 1999; Papineau 2002; Levin 2007; Balog 2012)—who accept Premise (1) of both the Zombie and the Knowledge Arguments, and challenge Premise (2) instead. They argue that our ability to *conceive* of a zombie does not show that zombies are genuinely possible, but only that our qualitative or phenomenal concepts of experience, derived by "pointing" introspectively at some feature of an experience one is currently having, are radically different from any physical-functional characterizations of what is going on in the brain. Similarly, they argue that when Mary first sees colors, she does not gain access to any new, non-physical, facts about human color experience, but only (via introspection) to new qualitative or phenomenal *concepts* of the neurophysiological processes that she learned about in her black-and-white room.

These views, in short, concede that there is no *a priori* link between our introspection-derived and physical-functional concepts of our conscious experiences—but deny that this shows that they cannot be concepts of the very same things. In addition, these materialists respond to the "Distinct Property Objection" to the Type-Identity Theory discussed by Smart (see Section 3) by contending that the concepts of our conscious mental states acquired by introspection can pick out those states *directly*, by demonstration, without need for any modes of presentation that entail that what has been picked out is a mental state of a particular qualitative type.

There are a number of different versions of Type B Materialism, but all face a common objection, namely, that while scientific identity statements such as "Water = H₂O" or "Heat = Mean molecular kinetic energy" seem perfectly intelligible after we learn more about the composition of the items in the world around us, it remains mysterious how, in Huxley's terms, "anything as remarkable as a state of consciousness comes about as a result of irritating nerve tissue"—even as we come to know more and more about the brain and nervous system. Type B materialists, in response, argue that the fact that our qualitative or phenomenal concepts derive from introspection, and are therefore radically disconnected from our physical and functional concepts, provides a compelling explanation of why there remains a hint of mystery in these cases, and these cases alone. In addition, some theorists (e.g. Nagel 1965; Brown 2010) argue that if there were a developed theory of *immaterial* substances and properties, then dualists would face a similar problem.

Type B Materialism nevertheless remains a controversial view. However, there is yet another way for materialists to avoid any unsatisfying consequences of the materialistic alternatives presented so far—namely, to embrace *Eliminativism* about conscious mental states. This view (which also, to be sure, has counterintuitive consequences) will be discussed in the next section.

6 Eliminativism

To embrace Eliminativism about some category of things is to deny that those things exist. One of the best-known eliminativists about mental states is Paul Churchland (1981) who argues that our common sense views about the role played by beliefs and desires in explaining behavior and other psychological phenomena are radically false, and, moreover, that they do not mirror, even approximately, the empirically established generalizations of a truly explanatory psychological theory. Thus, he concludes, it is reasonable to deny the existence of beliefs and desires, and take our routine attributions of such states no more literally than our talk of the sun's rising and setting.

Churchland's contention is highly controversial, but—regardless of its plausibility—he does not extend it to conscious mental states such as after-images, perceptual experiences, and sensations. There are a few radical eliminativists about such states; for example, Georges Rey (1983) denies outright that there are any properties that have the features that we ascribe to our conscious experiences. But most materialists who consider themselves eliminativists endorse what we may call Partial Eliminativism. Dennett (2002) argues that our common sense conception of conscious experience includes elements that further reflection will reveal to be incompatible—and argues that those theses that conflict with a broadly functionalist account of conscious experiences should be rejected. More recently, Par Sundstrom (2008) argues that we may be more willing than we think to be eliminativists: we start by being willing to deny that our color experiences possess qualities like (what seems to be) the yellow-orangeness of a yellow-orange after-image—and go on to recognize that it's far from clear, even by means of introspection, what the qualitative properties of our sensations and perceptual experiences are supposed to be. (See also Schwitzgebel 2008, for more general skepticism about the deliverances of introspection.)

In the end, both materialists and dualists may have to concede that there are, and always will be, some unsatisfying consequences of the views they endorse, and leave things at that. Indeed, Eric Schwitzgebel (2014) argues that all (well-developed) metaphysical theories of the nature of mental states, be they dualist or materialist, are “crazy,” in the sense that they include at least some important (“core”) theses that conflict with common sense—which we are given no compelling evidence to believe. Whether or not further reflection (or acculturation) will alleviate the bizarreness of some of these theses—or, alternatively, provide a compelling explanation of why they may always seem bizarre)—this view needs to be taken seriously.

7 Conclusion

However, even if all extant theories of the nature of conscious experience are crazy, in Schwitzgebel's sense, materialists can argue that adopting Dualism has, overall, too high a price: one has to accept two types of fundamental entities in the world, with little explanation of how non-physical properties arise in humans and certain non-human animals, and how they can have causal efficacy. Surely, materialists (or at least Type B materialists) argue, it is reasonable to accept that qualitative-physical identity statements may retain a hint of “mystery”—as long as there is an explanation for why such mystery may arise in these, and only these, cases.

But even if the pros of Materialism outweigh the cons, the materialists' work is far from done, since it is far from settled *which* materialist view is most promising. Does the greater universality of Functionalism (or Psychofunctionalism) outweigh its potential problems with mental causation, or are Type-Identity theories superior, even if they may not seem sufficiently universal? If Functionalism is superior, just what are the relations among mental states, stimulations, and behavior that make them *conscious* states: must these states be somehow “scanned” by the individual who is in them, or be the objects of that individual's thoughts (see Lycan 1996;

Rosenthal 1986; Gennaro 2004)? And which relations make mental states conscious states of particular types, e.g. experiences of red versus experiences of green? Moreover, perceptual experiences seem to *represent* items in the world: is this to be taken at face value, and if so, can there be an adequate materialistic account of what it is for a mental state to represent some object or property that allows for illusion and hallucination? These are just some of the questions that need to be answered to provide an adequate theory of conscious mental states, and therefore, even for those who believe that there are good grounds for embracing Materialism, there is still a lot of work to be done.

References

- Alter, T. (2016) "The Structure and Dynamics Argument," *Noûs* 50: 794–815.
- Balog, K. (2012) "In Defense of the Phenomenal Concept Strategy," *Philosophy and Phenomenological Research* 84: 1–23.
- Bechtel, W. and Mundale, J. (1999) "Multiple Realizability Revisited: Linking Neural and Cognitive States," *Philosophy of Science* 66: 175–207.
- Block, N. (1980) "Troubles with Functionalism," in N. Block (ed.) *Readings in Philosophy of Psychology, Volume One*, Cambridge, MA: Harvard University Press.
- Chalmers, D. (1996) *The Conscious Mind: In Search of a Fundamental Theory*, New York: Oxford University Press.
- Chalmers, D. (ed.) (2002a) *Philosophy of Mind*, New York: Oxford University Press.
- Chalmers, D. (2002b) "Consciousness and Its Place in Nature," in Chalmers 2002a.
- Chisholm, R. (1957) *Perceiving*, Ithaca, NY: Cornell University Press.
- Chomsky, N. (1959) "Review of Skinner's *Verbal Behavior*," *Language* 35: 26–58.
- Dennett, D. (1988) "Quining Qualia," Reprinted in Chalmers (2002a).
- Descartes, R. (1641) *Meditations on First Philosophy*. Reprinted in Cottingham, Stoothoff, Murdoch (tr.), (1985). *The Philosophical Writings of Descartes*, Vol. 2, Cambridge: Cambridge University Press.
- Descartes, R., Elisabeth, Princess of Bohemia (1643/1985) Correspondence with Descartes. In Cottingham, Stoothoff, Murdoch, Kenny (tr.), *The Philosophical Writings of Descartes*, Vol. 3. Cambridge: Cambridge University Press.
- Feigl, H. (1958) "The 'Mental' and the 'Physical,'" in H. Feigl, M. Scriven and G. Maxwell (eds.) *Concepts, Theories and the Mind-Body Problem* (Minnesota Studies in the Philosophy of Science, Volume 2), Minneapolis, MN: University of Minnesota Press.
- Gennaro, R. (2004) "Higher-Order Theories of Consciousness: An Overview," in R. Gennaro (ed.) *Higher-Order Theories of Consciousness*, Amsterdam and Philadelphia: John Benjamins Publishers.
- Hill, C. and McLaughlin, B. (1999) "There Are Fewer Things in Reality than Are Dreamt of in Chalmers' Philosophy," *Philosophy and Phenomenological Research* 59: 445–454.
- Hobbes, T. (1668/1994) *Leviathan*, Indianapolis, IN: Hackett Publishing Company; Underlined, Notations edition.
- Huxley, T.H. (1881) *Lessons in Elementary Physiology*, Macmillan and Co.
- Jackson, F. (1982) "Epiphenomenal Qualia," *Philosophical Quarterly* 32: 127–136.
- Jackson, F. (2004) "Mind and Illusion," In P. Ludlow, Y. Nagasawa, and D. Stoljar (eds.) *There's Something About Mary*, Cambridge, MA: MIT Press.
- Kim, J. (1998) *Mind in a Physical World*, Cambridge, MA: MIT Press.
- Kripke, S. (1980) *Naming and Necessity*, Cambridge, MA: Harvard University Press.
- La Mettrie, J. (1747/1994) *Man, a Machine*, Indianapolis, IN: Hackett Publishing.
- Levin, J. (2007) "What Is a Phenomenal Concept?," in T. Alter and S. Walter (eds.) *Phenomenal Concepts and Phenomenal Knowledge*, Oxford: Oxford University Press.
- Leibniz, G.W. (1714) Reprinted in D. Garber (tr.) and R. Ariew (ed.) (1991). *Discourse on Metaphysics and Other Essays*, Indianapolis, IN: Hackett Publishing.
- Loar, B. (1997) "Phenomenal States (revised)," in Chalmers 2002a.
- Lycan, W.G. (1996) *Consciousness and Experience*, Cambridge, MA: Bradford Books, MIT Press.
- Malcolm, N. (1968) "The Conceivability of Mechanism," *Philosophical Review* 77: 45–72.
- McGinn, C. (1989) "Can We Solve the Mind-Body Problem?" *Mind* 98: 349–366.
- Nagel, T. (1965) "Physicalism," *The Philosophical Review* 74: 339–356.

- Nagel, T. (1974) "What Is It Like to Be a Bat?" *The Philosophical Review* 83: 435–450.
- Papineau, D. (2002) *Thinking About Consciousness*, Oxford: Clarendon Press.
- Pereboom, D. (2011) *Consciousness and the Prospects of Physicalism*, New York: Oxford University Press.
- Polger, T. and Shapiro, L. (2016) *The Multiple Realization Book*, Oxford: Oxford University Press.
- Place, U.T. (1956) "Is Consciousness a Brain Process?" *British Journal of Psychology* 47: 44–50.
- Putnam, H. (1968) "Brains and Behavior," in R.J. Butler (ed.) *Analytical Philosophy, Second Series*. Blackwell: 1–19.
- Rey, G. (1983) "A Reason for Doubting the Existence of Consciousness," in R. Davidson, G. Schwartz, and D. Shapiro (eds.) *Consciousness and Self-Regulation*, Vol. 3, New York: Plenum.
- Rosenthal, D. (1986) "Two Concepts of Consciousness," *Philosophical Studies* 49: 329–359.
- Ryle, G. (1949) *The Concept of Mind*, London: Hutcheson.
- Schwitzgebel, E. (2008) "The Unreliability of Naïve Introspection," *Philosophical Review* 117: 245–273.
- Schwitzgebel, E. (2014) "The Crazyist Metaphysics of Mind," *Australasian Journal of Philosophy* 92: 665–682.
- Searle, J. (1980) "Minds, Brains, and Programs," *The Behavioral and Brain Sciences* 3: 417–457.
- Skinner, B.F. (1953) *Science and Human Behavior*, New York: Macmillan.
- Smart, J.J.C. (1959) "Sensations and Brain Processes," *Philosophical Review* 68: 141–156
- Smart, J.J.C. (2007) "The Mind/Brain Identity Theory," *The Stanford Encyclopedia of Philosophy* (Winter 2014 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2014/entries/mind-identity/>.
- Stoljar, D. (2001) "Two Conceptions of the Physical," *Philosophical and Phenomenological Research* LXII: 253–281.
- Sundstrom, P. (2008) "A Somewhat Eliminativist Proposal about Phenomenal Consciousness," in A. Hieke and H. Leitgeb (eds.) *Reduction and Elimination in Philosophy and the Sciences: Papers of the 31st International Wittgenstein Symposium*, Kirchberg am Wechsel: The Austrian Ludwig Wittgenstein Society.
- Van Gulick, R. (1993) "Understanding the Phenomenal Mind: Are We All Just Armadillos?" In M. Davies and G. Humphreys (eds.) *Consciousness: Psychological and Philosophical Essays*. Oxford: Blackwell.
- Watson, J. (1930) *Behaviorism*, Norton: New York.
- Wittgenstein, L. (1953) *Philosophical Investigations*, New York: Macmillan.
- Yablo, S. (1992) "Mental Causation," *Philosophical Review* 101: 245–280.

Related Topics

Consciousness in Western Philosophy
Dualism
Biological Naturalism and Biological Realism
The Neural Correlates of Consciousness