

PDF version of the entry
Functionalism
<https://plato.stanford.edu/archives/sum2023/entries/functionalism/>

from the SUMMER 2023 EDITION of the

STANFORD ENCYCLOPEDIA OF PHILOSOPHY



Co-Principal Editors: Edward N. Zalta & Uri Nodelman

Associate Editors: Colin Allen, Hannah Kim, & Paul Oppenheimer

Faculty Sponsors: R. Lanier Anderson & Thomas Icard

Editorial Board: <https://plato.stanford.edu/board.html>

Library of Congress ISSN: 1095-5054

Notice: This PDF version was distributed by request to members of the Friends of the SEP Society and by courtesy to SEP content contributors. It is solely for their fair use. Unauthorized distribution is prohibited. To learn how to join the Friends of the SEP Society and obtain authorized PDF versions of SEP entries, please visit <https://leibniz.stanford.edu/friends/>.

Stanford Encyclopedia of Philosophy
Copyright © 2023 by the publisher
The Metaphysics Research Lab
Department of Philosophy
Stanford University, Stanford, CA 94305

Functionalism
Copyright © 2023 by the author
Janet Levin

All rights reserved.

Copyright policy: <https://leibniz.stanford.edu/friends/info/copyright/>

Functionalism

First published Tue Aug 24, 2004; substantive revision Tue Apr 4, 2023

Functionalism in the philosophy of mind is the doctrine that what makes something a mental state of a particular type **does not depend on its internal constitution**, but rather on the way it functions, or **the role it plays, in the system of which it is a part**. This doctrine is rooted in Aristotle's conception of the soul, and has antecedents in Hobbes's conception of the mind as a "calculating machine", but it has become fully articulated (and popularly endorsed) only in the last third of the 20th century. Though the term 'functionalism' is used to designate a variety of positions in a variety of other disciplines, including psychology, sociology, economics, and architecture, this entry focuses exclusively on functionalism as a philosophical thesis about the nature of mental states.

The following sections will trace the intellectual antecedents of contemporary functionalism, sketch the different types of functionalist theories, and discuss the most serious objections to them.

- 1. What is Functionalism?
- 2. Antecedents of Functionalism
 - 2.1 Early Antecedents
 - 2.2 Thinking Machines and the "Turing Test"
 - 2.3 Behaviorism
- 3. Varieties of Functionalism
 - 3.1 Machine State Functionalism
 - 3.2 Functional Definitions and Ramsey-sentences
 - 3.3 Analytic Functionalism
 - 3.4 Psychofunctionalism
 - 3.5 Role-functionalism and Realizer-functionalism
- 4. Constructing Plausible Functional Theories

- 4.1 Characterizing Experiential States
- 4.2 Characterizing Intentional States
- 4.3 Characterizing the Inputs and Outputs of a System
- 5. Objections to Functionalism
 - 5.1 Functionalism and Holism
 - 5.2 Functionalism and Mental Causation
 - 5.3 Functionalism and Introspective Belief
 - 5.4 Functionalism and the Norms of Reason
 - 5.5 Functionalism and the Problem of Qualia
 - 5.5.1 Inverted and Absent Qualia
 - 5.5.2 Functionalism, Zombies, and the “Explanatory Gap”
 - 5.5.3 Functionalism and the Knowledge Argument
- 6. The Future of Functionalism
- Bibliography
- Academic Tools
- Other Internet Resources
- Related Entries

1. What is Functionalism?

Functionalism is the doctrine that what makes something a thought, desire, pain (or any other type of mental state) depends not on its internal constitution, but solely on its function, or the role it plays, in the cognitive system of which it is a part. More precisely, functionalist theories take the identity of a mental state to be determined by its causal relations to sensory stimulations, other mental states, and behavior.

For (an avowedly simplistic) example, a functionalist theory might characterize *pain* as 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. According to this theory, all and only creatures with internal states that can meet this condition, or play this role, are capable of being in pain, and an individual is in pain at time t if and only if they are in a state that is playing this role at t .

Suppose that, in humans, there is some distinctive kind of neural activity (C-fiber stimulation, for example) that plays this role. If so, then according to this functionalist theory, humans can be in pain simply by undergoing C-fiber stimulation. But the theory permits creatures with very different physical constitutions to have mental states as well: if there are silicon-based states of hypothetical Martians or inorganic states of hypothetical androids that also meet these conditions, then these creatures, too, can be in pain. As functionalists often put it, pain can be *realized* by different types of physical states in different kinds of creatures, or *multiply realized*. (See entry on multiple realizability.) Indeed, since descriptions that make explicit reference only to a state’s causal relations with stimulations, behavior, and one another are what have come to be known as “topic-neutral” (Smart 1959) – that is, as imposing no logical restrictions on the nature of the items that satisfy the descriptions – then it’s also logically possible for *non*-physical states to play the relevant roles, and thus realize mental states, in some systems as well. So functionalism is compatible with the sort of dualism that takes mental states to cause, and be caused by, physical states.

Still, though functionalism is officially neutral between materialism and dualism, it has been particularly attractive to materialists, since many materialists believe (or argue; see Lewis, 1966) that it is overwhelmingly likely that any states capable of playing the roles in question will be physical states. If so, then functionalism can stand as a materialistic alternative to the Psycho-Physical Identity Thesis (introduced in Place 1956, Feigl 1958, and Smart 1959, and defended more recently in Hill 1991, and Polger 2011), which holds that each type of mental state is

identical with a particular type of *neural* state. This thesis seems to entail that no creatures with brains unlike ours can share our sensations, beliefs, and desires, no matter how similar their behavior and internal organization may be to our own, and thus functionalism, with its claim that mental states can be multiply realized, has been regarded as providing a more inclusive, less “(species-) chauvinistic” (Block 1980b) – theory of the mind that is compatible with materialism. (More recently, however, some philosophers have contended that the identity thesis may be more inclusive than functionalists assume; see Section 6 for further discussion.)

Within this broad characterization of functionalism, however, a number of distinctions can be made. One of particular importance is the distinction between theories in which the functional characterizations of mental states purport to provide analyses of the meanings of our mental state terms (or otherwise restrict themselves to a priori information), and theories that permit functional characterizations of mental states to appeal to information deriving from scientific experimentation (or speculation). (See Shoemaker 1984c, and Rey 1997, for further discussion and more fine-grained distinctions.) There are other important differences among functionalist theories as well. These (sometimes orthogonal) differences, and the motivations for them, can best be appreciated by examining the origins of functionalism and tracing its evolution in response both to explicit criticisms of the thesis and changing views about the nature of psychological explanation.

2. Antecedents of Functionalism

Although functionalism attained its greatest prominence as a theory of mental states in the last third of the 20th century, it has antecedents in both modern and ancient philosophy, as well as in early theories of computation and artificial intelligence.

2.1 Early Antecedents

The earliest view in the Western canon that can be considered an ancestor of functionalism is Aristotle’s theory of the soul (350 BCE). In contrast to Plato’s claim that the soul can exist apart from the body, Aristotle argued (*De Anima* Bk. II, Ch. 1) that the (human) soul is the *form* of a natural, organized human body – the set of powers or capacities that enable it to express its “essential whatness”, which for Aristotle is a matter of fulfilling the function or purpose that defines it as the kind of thing it is. Just as the form of an axe is whatever enables it to cut, and the form of an eye is whatever enables it to see, the (human) soul is to be identified with whichever powers and capacities enable a natural, organized human body to fulfill its defining function, which, according to Aristotle, is to survive and flourish as a living, acting, perceiving, and reasoning being. So, Aristotle argues, the soul is inseparable from the body, and comprises whichever capacities are required for a body to live, perceive, reason, and act. (See Shields, 1990, and Nelson, 1990, for further debate about whether Aristotle’s view can be considered to be a version of functionalism.)

A second, relatively early, ancestor of contemporary functionalism is Hobbes’s (1651) account of reasoning as a kind of computation that proceeds by mechanistic principles comparable to the rules of arithmetic. Reasoning, he argues, is “nothing but *reckoning*, that is adding and subtracting, of the consequences of general names agreed upon for the *marking* and *signifying* of our thoughts.” (*Leviathan*, Ch. 5) In addition, Hobbes suggests that reasoning – along with imagining, sensing, and deliberating about action, all of which proceed according to mechanistic principles – can be performed by systems of various physical types. As he puts it in his Introduction to *Leviathan*, where he likens a commonwealth to an individual human, “why may we not say that all automata (engines that move themselves by springs and wheels...) have an artificial life? For

what is the heart but a spring; and the nerves but so many strings, and the joints but so many wheels...". It was not until the middle of the 20th century, however, that it became common to speculate that thinking may be nothing more than rule-governed computation that can be carried out by creatures of various physical types.

2.2 Thinking Machines and the "Turing Test"

In a seminal paper (Turing 1950), A.M. Turing proposed that the question, "Can machines think?" can be replaced by the question, "Is it theoretically possible for a finite state digital computer, provided with a large but finite table of instructions, or program, to provide responses to questions that would fool an unknowing interrogator into thinking it is a human being?" Now, in deference to its author, this question is most often expressed as "Is it theoretically possible for a Turing machine (appropriately programmed) to pass the Turing Test?" (See the entry on the Turing Test.)

In arguing that this question is a legitimate replacement for the original (and speculating that its answer is "yes"), Turing identifies thoughts with states of a system defined solely by their roles in producing further internal states and verbal outputs, given certain verbal inputs – a view on which, like Hobbes's and subsequent functionalist theories – many physically different systems could have internal states that play these roles. Indeed, Turing's work was explicitly invoked by many theorists during the beginning stages of 20th century functionalism, and was the avowed inspiration for a class of theories, the "machine state" theories most firmly associated with Hilary Putnam (1960, 1967) that had an important role in the early development of the doctrine.

2.3 Behaviorism

Other important recent antecedents of functionalism are the behaviorist theories that emerged in the early-to-mid twentieth century. These include both the "logical" or "analytical" behaviorism of philosophers such as Malcolm (1968) and Ryle (1949) (and, arguably, Wittgenstein 1953) and the empirical psychological theories associated primarily with Watson and Skinner.

Logical behaviorism is a thesis about the meanings of our mental state terms or concepts – in particular, that all statements about mental states and processes are equivalent in meaning to statements about behavioral dispositions. So, for (again, an overly simplified) example, "Henry has a toothache" would be equivalent in meaning to a statement such as "Henry is disposed (all things being equal) to cry out or moan and to rub his jaw". And "Amelia is thirsty" would be equivalent to a statement such as "If Amelia is offered some water, she will be disposed (all things being equal) to drink it." These candidate translations, like all behavioristic statements, eschew reference to any internal states of the organism, and thus do not threaten to denote, or otherwise induce commitment to, properties or processes (directly) observable only by introspection. In addition, logical behaviorists argued that if statements about mental states were equivalent in meaning to statements about behavioral dispositions, there could be an unproblematic account of how mental state terms could be applied both to oneself and others, and how they could be taught and learned.

In contrast, scientific behaviorism is an empirical theory that attempts to explain the behavior of humans (and other animals) by appealing solely to behavioral dispositions, that is, to the lawlike tendencies of organisms to behave in certain ways, given certain environmental stimulations. Stimulations and behavior, unlike thoughts, feelings, and other internal states that can be directly observed only by introspection, are objectively

observable, and are indisputably part of the natural world. Thus behavioral dispositions seemed to be fit entities to figure centrally in the emerging science of psychology, allowing for a science of human behavior as objective and explanatory as other “higher-level” sciences such as chemistry and biology. Also, behaviorist theories promised to avoid a potential regress that appeared to threaten psychological explanations invoking internal representations, namely, that to specify how such representations produce the behaviors in question, one must appeal to an internal intelligent agent (a “homunculus”) who interprets the representations, and whose skills would themselves have to be explained.

Both varieties of behaviorism, however, faced a common problem.

As many philosophers have pointed out (e.g. Chisholm 1957; Geach 1957), logical behaviorism provides an implausible account of the meanings of our mental state terms, since, intuitively, a subject can have the mental states in question without the relevant behavioral dispositions – and vice versa. For example, Gene may believe that it’s going to rain even if he’s not disposed to wear a raincoat and take an umbrella when leaving the house (or to perform any other cluster of rain-avoiding behaviors), if Gene doesn’t mind, or actively enjoys, singing in the rain. And subjects with the requisite motivation can suppress their tendencies to pain behavior even in the presence of excruciating pain, while skilled actors can perfect the lawlike disposition to produce pain behavior under certain conditions, even if they don’t actually feel pain. (See e.g. Putnam 1965) The problem, these philosophers argued, is that no mental state, by itself, can plausibly be assumed to give rise to any particular behavior unless one also assumes that the subject possesses additional mental states of various types. And so, it seemed, it is not in fact possible to give meaning-preserving translations of statements invoking pains, beliefs, and desires in purely behavioristic terms; one needs to include reference to the subject’s other mental states as well. Nonetheless, the idea that our common sense

concepts of mental states reveal an essential tie between mental states and their typical behavioral expressions is retained, and elaborated, in contemporary “analytic” functionalist theories.

Scientific behaviorism faced similar challenges. The theories of Watson, Skinner, et al had some early successes, especially in the domain of animal learning, and its principles are still used, at least for heuristic purposes, in various areas of psychology. But as many psychologists (and others, e.g. Chomsky 1959) have argued, the successes of behaviorism seem to depend upon the experimenters’ implicit control of certain variables which, when made explicit, involve ineliminable reference to organisms’ other mental states. For example, rats are typically placed into an experimental situation at a certain fraction of their normal body weight – and thus can be assumed to *feel hunger* and to *want* the food rewards contingent upon behaving in certain ways. Similarly, it is assumed that humans, in analogous experimental situations, *want* to cooperate with the experimenters, and *understand* and know how to follow the instructions. It seemed to the critics of behaviorism, therefore, that theories that explicitly appeal to an organism’s beliefs, desires, and other mental states, as well as to stimulations and behavior, would provide a fuller and more accurate account of why organisms behave as they do. They could do so, moreover, without compromising the objectivity of psychology as long as the mental states to which these theories appeal are introduced as states that *together play a role* in the production of behavior, rather than states identifiable solely by introspection. Thus work was begun on a range of “cognitive” psychological theories which reflected these presumptions, and an important strain of contemporary functionalism, “psychofunctionalism” (Fodor 1968, Block and Fodor 1972) can be seen a philosophical endorsement of these new cognitive theories of mind.

3. Varieties of Functionalism

Given this history, it is helpful to think of functionalist theories as belonging to one of three major strains – “machine state functionalism”, “analytic functionalism”, and “psychofunctionalism” – and to see them as emerging, respectively, from early AI theories, empirical behaviorism, and logical behaviorism. It’s important to recognize, however, that there is at least some overlap in the bloodlines of these different strains of functionalism, and also that there are functionalist theories, both earlier and more recent, that fall somewhere in between. For example, Wilfrid Sellars’s (1956) account of mental states as “theoretical entities” is widely regarded as an important early version of functionalism, but it takes the proper characterization of thoughts and experiences to depend partially on their role in providing a scientific explanation of behavior, and partly on what he calls the “logic”, or the a priori interrelations, of the relevant concepts. Still, it is instructive to give separate treatment to the three major strains of the doctrine, as long as these caveats are kept in mind.

3.1 Machine State Functionalism

The early functionalist theories of Putnam (1960, 1967; see also Block and Fodor 1972) can be seen as a response to the difficulties facing behaviorism as a scientific psychological theory, and as an endorsement of the (new) computational theories of mind which were becoming increasingly significant rivals to it. (But see Putnam 1988, for subsequent doubts about machine functionalism, Chalmers 1996b, for a response, and Shagrir 2005, for a comprehensive account of the evolution of Putnam’s views on the subject)

According to *machine state functionalism*, any creature with a mind can be regarded as a Turing machine (an idealized finite state digital computer),

whose operation can be fully specified by a set of instructions (a “machine table” or program) having the form:

If the machine is in state S_i , and receives input I_j , it will go into state S_k and produce output O_l (for a finite number of states, inputs and outputs).

A machine table of this sort describes the operation of a *deterministic* automaton, but most machine state functionalists (e.g. Putnam 1967) take the proper model for the mind to be that of a *probabilistic* automaton: one in which the program specifies, for each state and set of inputs, the *probability* with which the machine will enter some subsequent state and produce some particular output.

On either model, however, the mental states of a creature are to be identified with such “machine table states” (S_1, \dots, S_n). These states are not mere behavioral dispositions, since they are specified in terms of their relations not only to inputs and outputs, but also to the state of the machine at the time. For example, if *believing it will rain* is regarded as a machine state, it will not be regarded as a disposition to take one’s umbrella after looking at the weather report, but rather as a disposition to take one’s umbrella if one looks at the weather report *and* is in the state of wanting to stay dry. So machine state functionalism can avoid what many have thought to be a fatal difficulty for behaviorism. In addition, machines of this sort provide at least a simple model of how internal states whose effects on output occur by means of mechanical processes can be viewed as *representations* (though the question of *what*, exactly, they represent has been an ongoing topic of discussion (see sections 4.4–5). Finally, machine table states are not tied to any particular physical (or other) realization; the same program, after all, can be run on different sorts of computer hardware.

It's easy to see, therefore, why Turing machines provided a fruitful model for early functionalist theories. And the idea that mental states are best regarded as computational states appears in many theories of the mind (see, for example, Rey 1997; but see Piccinini 2004 for dissent and the entry on the computational theory of mind for a comprehensive discussion of this question). Nonetheless, because machine table states are total states of a system, most contemporary functionalists – both analytic functionalists and psychofunctionalists – have adopted another way of characterizing mental states, namely, as states implicitly defined by the so-called Ramsey-sentence of a psychological theory – either one that derives from our commonly held beliefs about the causal roles of mental states in the production of behavior, or from the results of empirical psychological investigation. This will be the focus of the next section.

3.2 Functional Definitions and Ramsey-sentences

The key feature of this now-canonical method, presented initially by David Lewis (1972), building on a technique introduced by Frank Ramsey, is to treat mental state terms as being implicitly defined by the so-called *Ramsey sentence* of one or another psychological theory – common sense, scientific, or something in between. (Analogous steps, of course, can be taken to produce the Ramsey-sentence of *any* theory, psychological or otherwise). For (a still simplistic) example, consider the sort of generalizations about pain introduced before: pain tends to be caused by bodily injury; pain tends to produce the belief that something is wrong with the body and the desire to be out of that state; pain tends to produce anxiety; pain tends to produce wincing or moaning.

To construct the Ramsey-sentence of this “theory”, the first step is to conjoin these generalizations, then to replace all names of different types of mental states with different variables, and then to existentially quantify those variables, as follows:

$$\exists x \exists y \exists z \exists w (x \text{ tends to be caused by bodily injury} \ \& \ x \text{ tends to produce states } y, z, \text{ and } w \ \& \ x \text{ tends to produce wincing or moaning}).$$

Such a statement is free of any mental state terms. It includes only *quantifiers* that range over mental states, terms that denote stimulations and behavior, and terms that specify various causal relations among them. It can thus be regarded as providing implicit definitions of the mental state terms of the theory. An individual will have those mental states just in case it possesses a family of first-order states that interact in the ways specified by the theory. (Though functionalists of course acknowledge that the first-order states that satisfy the functional definitions may vary from species to species – or even from individual to individual – they specify that, for each individual, the functional definitions be *uniquely* satisfied.)

A helpful way to think of the Ramsey sentence of a psychological theory is to regard it as defining a system's mental states “all at once” as states that interact with stimulations in various ways to produce behavior (See Lewis 1972; also see Field 1980 for a more technical elaboration of Lewis's method, and an account of some crucial differences between this kind of characterization and the one Lewis initially proposed.) It is also helpful to view the differences between analytic and psychofunctionalism as differences in the Ramsey-sentences of our “commonsense theory of the mind” versus our empirical psychological theories of the roles of mental states in the production of other mental states and behavior.

3.3 Analytic Functionalism

Like the logical behaviorism from which it emerged, the goal of analytic functionalism is to provide dispositional, or other “topic-neutral”, translations or analyses of our ordinary mental state terms or concepts. Analytic functionalism, of course, has richer resources than logical behaviorism for such translations, since it permits reference to certain

causal and transitional relations that a mental state has to stimulations, behavior, *and other mental states*. So, for example, the statement “Blanca wants some coffee” need not be rendered, as logical behaviorism requires, in terms such as “Blanca is disposed to order coffee when it is offered”, but rather as “Blanca is disposed to order coffee when it is offered, *if* she has no stronger desire to avoid coffee”. But any theory – and its Ramsey-sentence – that is acceptable to analytic functionalists must include only generalizations about mental states, their environmental causes, and their joint effects on behavior that are so widely known and “platitudinous” to count as *analyzing our ordinary concepts* of the mental states in question. (See Smart 1959, Armstrong 1968, Shoemaker 1984a,b,c, Lewis 1972, and Braddon-Mitchell and Jackson 1996/2007.)

A major question, of course, is whether a theory that limits itself to such “platitudes” about the causal relations between stimulations, mental states, and behavior can make the right distinctions among mental states, or even worse, is so “liberal” that it can be realized by systems without any sort of mentality at all, such as the economy of Bolivia (Block, 1980b). However, commonsense psychology has more resources than it may initially seem. First, (at least arguably) it need not be restricted to the “platitudes” that can be accessed immediately; it may take a certain amount of Socratic questioning to prompt us to recognize certain similarities and differences among the causal-relational properties of our mental states. The information accessed by such questioning, however, was always available to us, at least in principle, and so can count as the deliverances of common sense, rather than empirical investigation. Moreover – and contrary to the claims of some theorists (e.g. Churchland, 1981) – commonsense psychology is not stagnant. It can, over time, absorb information acquired by exposure to empirical theories, while nonetheless retaining its platitudinous status. (Not that this is always a good thing: think, for example, of how Freudian theory and various now discredited theories

about the causes of autism, depression, and impulsive behavior once seemed to be the deliverances of common sense.)

Nonetheless, many functionalists argue that commonsense theories do not have sufficient resources to capture the causal roles of the internal states that differentiate us from other cognitive (and non-cognitive) systems. They look instead to a more empirically informed, and presumably more restrictive, theory of mental states and their effects on behavior – psychofunctionalism – that derives primarily from reflection upon the goals and methodology of the “cognitive” psychological theories that are the descendants of scientific behaviorism.

3.4 Psychofunctionalism

In contrast to the scientific behaviorists’ insistence that the laws of psychology appeal only to behavioral dispositions, cognitive psychologists argue that the best empirical theories of behavior take it to be the result of a complex of mental states and processes, introduced and individuated in terms of the roles they play in producing the behavior to be explained. For example (Fodor’s, in his 1968, Ch. 3), a psychologist may begin to construct a theory of memory by postulating the existence of “memory trace” decay, a process whose occurrence or absence is responsible for effects such as memory loss and retention, and which is affected by stress or emotion in certain distinctive ways.

On a theory of this sort, what makes some neural process an instance of memory trace decay is a matter of how it functions, or the role it plays, in a cognitive system; its neural or chemical properties are relevant only insofar as they enable that process to do what trace decay is hypothesized to do. And similarly for all mental states and processes invoked by cognitive psychological theories. Cognitive psychology, that is, is intended by its proponents to be a “higher-level” science like biology, and thus to

have autonomy from lower-level sciences such as neurophysiology: just as, in biology, physically disparate entities can all be hearts as long as they function to circulate blood in a living organism, and physically disparate entities can all be eyes as long as they enable an organism to see, disparate physical structures or processes can be instances of memory trace decay – or more familiar phenomena such as thoughts, sensations, and desires – as long as they play the roles described by the relevant cognitive theory.

Psychofunctionalism, therefore, can be seen as straightforwardly adopting the methodology of cognitive psychology in its characterization of mental states and processes as entities defined by their role in a cognitive psychological theory. What distinguishes it from analytic functionalism is that the information used in the functional characterization of mental states and processes needn't be restricted to what is considered common knowledge or common sense, but can include information available only by careful empirical observation and experimentation. For example, a psychofunctional theory might be able to distinguish phenomena such as depression from sadness or listlessness even though the distinctive causes and effects of these syndromes are difficult to untangle solely by consulting intuitions or appealing to common sense. And psychofunctional theories will not include characterizations of mental states for which there is no scientific evidence, such as buyer's regret or hysteria, even if the existence and efficacy of such states is something that common sense affirms.

This may seem to be an unmitigated advantage, since psychofunctional theories can avail themselves of all the tools of inquiry available to scientific psychology, and will presumably make all, and only, the distinctions that are scientifically sound. This methodology, however, leaves psychofunctionalism open to the charge that it, like the Psycho-Physical Identity Thesis, may be overly "chauvinistic" (Block 1980b), since creatures whose internal states share the rough, but not fine-grained,

causal patterns of ours wouldn't count as sharing our mental states. Many psychofunctionalists, however, may not regard this as an unhappy consequence, and argue that it's appropriate to treat only those who are psychologically similar as having the same mental states.

But there is another serious worry about the thesis, namely, that if the laws of the best empirical psychological theories diverge from even the broad contours of our "folk psychology" – that is, our common sense beliefs about the causal roles of our thoughts, sensations, and perceptions – it will be hard to take psychofunctional theories as providing an account of our mental states, rather than merely changing the subject (Loar 1981, Stich 1983, Greenwood 1991, Braddon-Mitchell and Jackson 1996/2007). Many theorists, however (Horgan and Woodward 1985), argue that it's likely that future psychological theories will be recognizably close to "folk psychology", though this question has been the subject of debate (Churchland 1981).

Yet another reason that some prefer analytic functionalism derives from a debate that occurred in the early days of the Psycho-Physical Identity Theory, the thesis that each type of mental state can be identified with some type of brain state or neural activity. For example, early identity theorists (e.g. Smart 1959) argued that it makes perfect sense (and may well be true) to identify *pain* with *C-fiber stimulation*. The terms 'pain' and 'C-fiber stimulation', they acknowledged, do not have the same *meaning*, but nonetheless they can denote the same state; the fact that an identity statement is not a priori, they argued, does not mean that it is not true. And just because I need not consult some sort of brain scanner when reporting that I'm in pain doesn't mean that the pain I report is not a neural state that a brain scanner could (in principle) detect.

An important – and enduring – objection to this argument, however, was raised early on by Max Black (reported in Smart 1959). Black argued,

following Frege (1892), that the only way that terms with different meanings can denote the same state is to express (or connote) different properties, or “modes of presentation”, of that state. But this implies, he argued, that if terms like ‘pain’, ‘thought’ and ‘desire’ are not equivalent in meaning to any physicalistic descriptions, they can denote physical states only by expressing *irreducibly mental properties* of them. This argument has come to be known as the “Distinct Property Argument”, and was taken by its proponents to undermine a thorough-going materialistic theory of the mind.

In response, Smart (1959) and later, Armstrong (1968) countered that there could be relational, “topic-neutral” terms that are equivalent in meaning to terms such as ‘pain’, ‘thought’, and ‘desire’, and if so, then there are properties distinct from physical properties by virtue of which mental state terms can denote brain states, yet are not irreducibly mental. However, Smart’s and Armstrong’s suggestions were widely regarded as inadequate. The appeal of meaning-preserving functional characterizations, therefore, derives from their promise to provide more plausible topic-neutral equivalents of our mental state terms and concepts and thereby blunt the anti-materialistic force of the Distinct Property Argument. (See Lewis 1966).

Still, though many regard functional characterizations as an improvement on the earlier attempts by Smart and Armstrong, there is a dispute about whether any relational characterizations of our mental states, especially sensations, could fully preserve the meanings of our mental state terms. On the other hand, there is a dispute about how seriously to take the Distinct Property objection. (See White 1986 and 2007, for more recent versions of this objection, and Block 2007, and Levin, 2020, for a response; these questions will be addressed more fully in Section 5.5.)

There is yet another distinction between kinds of functional theory – one that crosscuts the distinctions described so far – that is important to note. This is the distinction between what has come to be known as “role” functionalism and “realizer” (or “filler”) functionalism (McLaughlin 2006).

3.5 Role-functionalism and Realizer-functionalism

To see the difference between role-functionalism and realizer-functionalism, consider – once again – the (avowedly simplistic) example of a functional theory of pain introduced in the first section.

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.

As noted earlier, if in humans this functional role is played by C-fiber stimulation, then, according to this functionalist theory, humans can be in pain simply by undergoing C-fiber stimulation. But there is a further question to be answered, namely, what is the property of pain itself? Is it the higher-level relational property of being in some state or other that plays the “pain role” in the theory, or the C-fiber stimulation that actually plays this role?

Role functionalists identify pain with that higher-level relational property. Realizer functionalists, however, take a functional theory merely to provide definite descriptions of whichever *lower-level* properties satisfy the functional characterizations. On these views (also called “functional specification” theories), if the property that occupies the causal role of pain in human beings is C-fiber stimulation, then pain (or at least pain-in-humans) would *be* C-fiber stimulation, rather than the higher-level

property of having some lower-level state that plays the relevant role. (This is not to suggest that there is a difference in kind between higher-level “role” properties and the lower-level “realizations” of those roles, since it may be that, relative to even lower-level descriptions, those realizations can be characterized as functional states themselves (Lycan 1987)).

Some of the earliest versions of analytic functionalism (Lewis 1966, Armstrong 1968 – but see Lewis 1980, for a modification) were presented as functional specification theories, as topic-neutral “translations” of mental state terms that could pave the way for a psycho-physical identity theory by defusing the Distinct Property Argument (see section 3.3). However, if there are differences in the physical states that satisfy the functional definitions in different (actual or hypothetical) creatures, such theories – like most versions of the identity theory – would violate a key motivation for functionalism, namely, that creatures with states that play the same role in the production of other mental states and behavior possess, literally, the same mental states.

It may be that there are some important, more general, physical similarities between the neural states of seemingly disparate creatures that satisfy a given functional characterization. (This issue will be discussed further in Section 6.) However, even if this is so, it is unlikely that these similarities hold of *all* the creatures, including Martians and other hypothetical beings, who could share our functional organization, and thus our theory of mental states would remain, in Block’s (1980) terms, overly “chauvinistic”. One could counter the charge of chauvinism, of course, by suggesting that all creatures with lower-level states that satisfy a given functional characterization possess a common (lower-level) *disjunctive* state or property. Or one could suggest that, even if all creatures possessing states that occupy (for example) the pain role are not literally in the same mental state, they nonetheless share a closely related higher-level

property (call it, following Lewis 1966 (note 6), “the attribute of having pain”). But neither alternative, for many functionalists, goes far enough to preserve the basic functionalist intuition that functional commonality outweighs physical diversity in determining whether creatures can possess the same mental states. Thus many functionalists – both analytic and empirical – advocate role functionalism, which, in addition to avoiding chauvinism, permits mental state terms to be rigid designators (Kripke 1972), denoting the same items – those higher-level “role” properties – in all possible worlds.

On the other hand, some functionalists – here, too, both analytic and empirical – consider realizer functionalism to be in a better position than role functionalism to explain the causal efficacy of the mental. If I stub my toe and wince, we believe that my toe stubbing causes my pain, which in turn causes my wincing. But, some have argued (Malcolm 1968; Kim 1989, 1998) that if pain is realized in me by some neural event-type, then insofar as there are purely physical law-like generalizations linking events of that type with wincings, one can give a complete causal explanation of my wincing by citing the occurrence of that neural event (and the properties by virtue of which it figures in those laws). And thus it seems that the higher-level role properties of that event are causally irrelevant. This is known as the “causal exclusion problem”, which is claimed to arise not just for functional role properties, but for dispositional properties in general (Prior, Pargetter, and Jackson 1982) – and indeed for any sort of mental states or properties not type-identical to those invoked in physical laws. This problem will be discussed further in Section 5.2.

4. Constructing Plausible Functional Theories

So far, the discussion of how to provide functional characterizations of individual mental states has been vague, and the examples avowedly simplistic. Is it possible to do better, and, if so, which version of

functionalism is likely to have the greatest success? These questions will be the focus of this section, and separate treatment will be given to experiential (often called ‘qualitative’ or ‘phenomenal’) states, such as perceptual experiences and bodily sensations, which have a distinctive qualitative character or “feel”, and intentional states, such as thoughts, beliefs, and desires, which purport to represent the world in various ways. To be sure, there is increasing consensus that experiential states have representational contents and intentional states have qualitative character and thus that these two groups may not be mutually exclusive (see Horgan and Tienson, 2002). Nonetheless I will discuss them separately to focus on what all agree to be the distinctive features of each.

4.1 Characterizing Experiential States

The common strategy in the most successful treatments of perceptual experiences and bodily sensations (Shoemaker 1984a, Clark 1993; adumbrated in Sellars 1956) is to individuate experiences of various general types (color experiences, experiences of sounds, feelings of temperature) in part by appeal to their positions in the “quality spaces” associated with the relevant sense modalities – that is, the (perhaps multidimensional) matrices determined by judgments about the relative similarities and differences among the experiences in question. So, for example, the experience of a very reddish-orange could be (partially) characterized as the state produced by the viewing of a color swatch within some particular range, which tends to produce the judgment or belief that the state just experienced is more similar to the experience of red than of orange. (Analogous characterizations, of course, will have to be given of these other color experiences.) The judgments or beliefs in question will themselves be (partially) characterized in terms of their tendencies to produce sorting or categorization behavior of certain specified kinds.

This strategy may seem fatal to analytic functionalism, which restricts itself to the use of a priori (or platitudinous) information to distinguish among mental states, since it’s not clear that the information needed to distinguish among experiences such as color perceptions is available to commonsense. However, this problem may not be as dire as it seems. For example, if sensations and perceptual experiences are characterized in terms of their places in a “quality space” determined by a person’s pre-theoretical judgments of similarity and dissimilarity (and perhaps also in terms of their tendencies to produce various emotional effects), then these characterizations may qualify as platitudinous, even though they would have to be elicited by a kind of “Socratic questioning”.

There are limits to this strategy, however (see Section 5.5.1 on the “inverted spectrum” problem), which seem to leave two options for analytic functionalists: fight – that is, deny that it’s coherent to suppose that there exist the distinctions that the critics suggest, or switch – that is, embrace another version of functionalism in which the characterizations of mental states, though not platitudinous, can provide information rich enough to individuate the states in question. To switch, however, would be to give up the benefits (if any) of a theory that offers meaning-preserving translations of our mental state terms.

There has been significant skepticism, however, about whether *any* functionalist theory – analytic or empirical – can capture what seems to be the distinctive qualitative character of experiential states such as color perceptions, pains, and other bodily sensations; these questions will be addressed in section 5.5 below.

4.2 Characterizing Intentional States

On the other hand, intentional states such as beliefs, thoughts, and desires (sometimes called “propositional attitudes”) have often been thought to be easier to characterize functionally than experiential states such as pains and color experiences (but not always: see Searle 1992, G. Strawson 1994, Horgan and Tienson 2002, Kriegel 2003, Pitt 2008, and Mendelovici, 2018) who suggest that intentional states have qualitative character as well). We can begin by characterizing beliefs as (among other things) states produced in certain ways by sense-perception or inference from other beliefs, and desires as states with certain causal or counterfactual relations to the system’s goals and needs, and specify further how (according to the relevant commonsense or empirical theory) beliefs and desires tend to interact with one another, and other mental states, to produce behavior.

Once again, this characterization is crude, and needs more detail. Moreover, there are some further questions about characterizing intentional states – particularly belief – that have emerged in recent discussions. One is whether a subject should be regarded as believing that *p* if there is a mismatch between her avowals that *p* and the characteristic behaviors associated with believing that *p* in standard circumstances: do avowals outweigh behaviors, or vice versa – or are there pragmatic factors that determine what the answer should be in different contexts? (See Gendler, 2008, and Schwitzgebel, 2010). Another question is whether the states that interact with desires (and other mental states) to produce behavior are best regarded as “full on” or “outright” beliefs, or rather as representations of the world for which individuals have varying degrees of confidence. (See Staffel, 2013, 2019, and the many contributions to Huber and Schmidt-Petri, 2009, and Ebert and Smith, 2012, for further discussion. Functionalism, at least arguably, can accommodate a number

of different answers to these questions, but the project of characterizing beliefs may not be straightforward.

Independently of these questions, functionalists need to say more (outright or not) about what makes a state a particular belief (outright or not) or desire, for example, the belief – or desire – that it will snow tomorrow. Most functional theories describe such states as different relations (or “attitudes”) toward the same state of affairs or proposition (and to describe the belief that it will snow tomorrow and the belief that it will rain tomorrow as the same attitude toward different propositions). This permits differences and similarities in the contents of intentional states to be construed as differences and similarities in the propositions to which these states are related. But what makes a mental state a relation to, or attitude toward, some proposition *P*? And can these relations be captured solely by appeal to the functional roles of the states in question?

The development of conceptual role semantics may seem to provide an answer to these questions: what it is for Julian to believe that *P* is for Julian to be in a state that has causal and counterfactual relations to other beliefs and desires that mirror certain inferential, evidential, and practical (action-directed) relations among propositions with those formal structures (Field 1980; Loar 1981; Block 1986). This proposal raises a number of important questions. One is whether states capable of entering into such interrelations can (must?) be construed as being, or including elements of, a “language of thought” (Fodor 1975; Harman 1973; Field 1980; Loar 1981). Another is whether idiosyncrasies in the inferential or practical proclivities of different individuals make for differences in (or incommensurabilities between) their intentional states. (This question springs from a more general worry about the *holism* of functional specification, which will be discussed more generally in Section 5.1.)

Yet another challenge for functionalism are the widespread intuitions that support “externalism”, the thesis that what mental states represent, or are about, cannot be characterized without appeal to certain features of the environments in which those individuals are embedded. Thus, if one individual’s environment differs from another’s, they may count as having different intentional states, even though they reason in the same ways, and have exactly the same “take” on those environments from their own points of view.

The “Twin Earth” scenarios introduced by Putnam (1975) are often invoked to support an externalist individuation of beliefs about natural kinds such as water, gold, or tigers. Twin Earth, as Putnam presents it, is a (hypothetical) planet on which things look, taste, smell, and feel exactly the way they do on Earth, but which have different underlying microscopic structures; for example, the stuff that fills the streams and comes out of the faucets, though it looks and tastes like water, has molecular structure XYZ rather than H₂O. Many theorists find it intuitive to think that we thereby *mean* something different by our term ‘water’ than our Twin Earth counterparts mean by theirs, and thus that the beliefs we describe as beliefs about water are different from those that our Twin Earth counterparts would describe in the same way. Similar conclusions, they contend, can be drawn for all cases of belief (and other intentional states) regarding natural kinds.

The same problem, moreover, appears to arise for other sorts of belief as well. Tyler Burge (1979) presents cases in which it seems intuitive that a person, Oscar, and his functionally equivalent counterpart have different beliefs about various syndromes (such as arthritis) and artifacts (such as sofas) because the usage of these terms by their linguistic communities differ. For example, in Oscar’s community, the term ‘arthritis’ is used as we use it, whereas in his counterpart’s community ‘arthritis’ denotes inflammation of the joints and also various maladies of the thigh. Burge’s

contention is that even if Oscar and his counterpart both complain about the ‘arthritis’ in their thighs and make exactly the same inferences involving ‘arthritis’, they mean different things by their terms and must be regarded as having different beliefs. If these cases are convincing, then there are differences among types of intentional states that can only be captured by characterizations of these states that make reference to the practices of an individual’s linguistic community. These, along with the Twin Earth cases, suggest that if functionalist theories cannot make reference to an individual’s environment, then capturing the representational content of (at least some) intentional states is beyond the scope of functionalism. (See Searle, 1980, for related arguments against “computational” theories of intentional states.)

On the other hand, the externalist individuation of intentional states may fail to capture some important psychological commonalities between ourselves and our counterparts that are relevant to the explanation of behavior. If my Twin Earth counterpart and I have both come in from a long hike, declare that we’re thirsty, say “I want some water” and head to the kitchen, it seems that our behavior can be explained by citing a common desire and belief. Some theorists, therefore, have suggested that functional theories should attempt merely to capture what has been called the “narrow content” of beliefs and desires – that is, whichever representational features individuals share with their various Twin Earth counterparts. There is no consensus, however, about just how functionalist theories should treat these “narrow” representational features (Block 1986; Loar 1987, Yli-Vakkuri and Hawthorne, 2018), and some philosophers have expressed skepticism about whether such features should be construed as representations at all (Fodor 1994; also see the entry on narrow mental content). Even if a generally acceptable account of narrow representational content can be developed, however, if the intuitions inspired by “Twin Earth” scenarios remain stable, then one must conclude that the full representational content of intentional states cannot be

captured by “narrow” functional characterizations alone (and this will be true as well for certain sorts of qualitative states, e.g. color experiences, if they too have representational content).

4.3 Characterizing the Inputs and Outputs of a System

Considerations about whether certain sorts of beliefs are to be externally individuated raise the related question about how best to characterize the stimulations and behaviors that serve as *inputs and outputs* to a system. Should they be construed as events involving objects in a system’s environment (such as fire trucks, water and lemons), or rather as events in that system’s sensory and motor systems? Theories of the first type are often called “long-arm” functional theories (Block 1990), since they characterize inputs and outputs – and consequently the states they are produced by and produce – by reaching out into the world. Adopting a “long-arm” theory would prevent our Twin Earth counterparts from sharing our beliefs and desires, and may thus honor intuitions that support an externalist individuation of intentional states (though further questions may remain about what Quine has called the “inscrutability of reference”; see Putnam 1988).

If functional characterizations of intentional states are intended to capture their “narrow contents”, however, then the inputs and outputs of the system will have to be specified in a way that permits individuals in different environments to be in the same intentional state. On this view inputs and outputs may be better characterized as activity in specific sensory receptors and motor neurons. But this (“short-arm”) option also restricts the range of individuals that can share our beliefs and desires, since creatures with different neural structures will be prevented from sharing our mental states, even if they share all our behavioral and inferential dispositions. (In addition, this option would not be open to analytic functionalist theories, since generalizations that link mental states

to neurally specified inputs and outputs would not, presumably, have the status of platitudes.)

Perhaps there is a way to specify sensory stimulations that abstracts from the specifics of human neural structure enough to include any possible creature that intuitively seems to share our mental states, but is sufficiently concrete to rule out entities that are clearly not cognitive systems (such as the economy of Bolivia; see Block 1980b). If there is no such formulation, however, then functionalists will either have to dispel intuitions to the effect that certain systems can’t have beliefs and desires, or concede that their theories may be more “chauvinistic” than initially hoped.

Clearly, the issues here mirror the issues regarding the individuation of intentional states discussed in the previous section. More work is needed to develop the “long-arm” and “short-arm” alternatives, and to assess the merits and deficiencies of both.

5. Objections to Functionalism

The previous sections were by and large devoted to the presentation of the different varieties of functionalism and the evaluation of their relative strengths and weaknesses. There have been many objections to functionalism, however, that apply to all versions of the theory. Some of these have already been introduced in earlier discussions, but they, and many others, will be addressed in more detail here.

5.1 Functionalism and Holism

One difficulty for every version of the theory is that functional characterization is *holistic*. Functionalists hold that mental states are to be characterized in terms of their roles in a psychological theory – be it common sense, scientific, or something in between – but all such theories

incorporate information about a large number and variety of mental states. Thus if pain is interdefined with certain highly articulated beliefs and desires, then animals who don't have internal states that play the roles of our articulated beliefs and desires can't share our pains, and humans without the capacity to feel pain can't share certain (or perhaps any) of our beliefs and desires. In addition, differences in the ways people reason, the ways their beliefs are fixed, or the ways their desires affect their beliefs – due either to cultural or individual idiosyncrasies – might make it impossible for them to share the same mental states. These are regarded as serious worries for all versions of functionalism (see Stich 1983, Putnam 1988).

Some functionalists, however (e.g. Lewis, 1972; Shoemaker 1984c), have suggested that if a creature has states that *approximately* realize our functional theories, or realize some more specific *defining* subset of the theory particularly relevant to the specification of those states, then they can qualify as being mental states of the same types as our own. The problem, of course, is to specify more precisely what it is to be an approximate realization of a theory, or what exactly a “defining” subset of a theory is intended to include, and these are not easy questions. (They have particular bite, moreover, for *analytic* functionalist theories, since specifying what belongs inside and outside the “defining” subset of a functional characterization raises the question of what are the conceptually essential, and what the merely collateral, features of a mental state, and thus raise serious questions about the feasibility of (something like) an analytic-synthetic distinction. (Quine 1953, Rey 1997)).

5.2 Functionalism and Mental Causation

Another worry for functionalism is the “causal exclusion problem”, introduced in Section 3.4: the worry about whether role functionalism can account for what we take to be the causal efficacy of our mental states

(Malcolm 1968, Kim 1989, 1998). For example, if pain is realized in me by some neural state-type, then insofar as there are purely physical law-like generalizations linking states of that type with pain behavior, one can give a complete causal explanation of my behavior by citing the occurrence of that neural state (and the properties by virtue of which it figures in those laws). And thus, some have argued, the higher-level role properties of that state – its being a pain – are causally irrelevant.

There have been a number of different responses to this problem. Some (e.g. Loewer 2002, 2007, Antony and Levine 1997, Burge 1997, Baker 1997) suggest that it arises from an overly restrictive account of causation, in which a cause must “generate” or “produce” its effect, a view which would count the macroscopic properties of other special sciences as causally irrelevant as well. Instead, some argue, causation should be regarded as a special sort of counterfactual dependence between states of certain types (Loewer 2002, 2007, Fodor 1990, Block 1997), or as a special sort of regularity that holds between them (Melnyk 2003). If this is correct, then functional role properties (along with the other macroscopic properties of the special sciences) could count as causally efficacious (but see Ney 2012 for dissent). However, the plausibility of these accounts of causation depends on their prospects for distinguishing bona-fide causal relations from those that are clearly epiphenomenal, and some have expressed skepticism about whether they can do the job, among them Crane 1995, Kim 2007, Jackson 1995, Ludwig 1998, and McLaughlin 2006, 2015. (On the other hand, see Lyons (2006) for an argument that if functional properties are causally inefficacious, this can be viewed as a *benefit* of the theory.)

Yet other philosophers argue that causation is best regarded as a relation between types of events that must be invoked to provide sufficiently general explanations of behavior (Antony and Levine 1997, Burge 1997, Baker 1997). Though many who are moved by the exclusion problem (e.g.

Kim, Jackson) maintain that there is a difference between generalizations that are truly causal and those that contribute in some other (merely epistemic) way to our understanding of the world, theorists who advocate this response to the problem charge that this objection, once again, depends on a restrictive view of causation that would rule out too much.

Another problem with views like the ones sketched above, some argue (Kim 1989, 1998), is that mental and physical causes would thereby *overdetermine* their effects, since each would be causally sufficient for their production. And, though some theorists argue that overdetermination is widespread and unproblematic (see Loewer 2002, and also Shaffer, 2003, and Sider 2003, for a more general discussion of overdetermination), others contend that there is a special relation between role and realizer that provides an intuitive explanation of how both can be causally efficacious *without* counting as overdetermining causes. For example, Yablo (1992), suggests that mental and physical properties stand in the relation of determinable and determinate (just as red stands to scarlet), and argues that our conviction that a cause should be *commensurate* with its effects permits us to take the determinable, rather than the determinate, property to count as causally efficacious in psychological explanation. Bennett (2003) suggests, alternatively, that the realizer properties *metaphysically necessitate* the role properties in a way that prevents them from satisfying the conditions for overdetermination. Yet another suggestion (Wilson, 1999, 2011, and Shoemaker, 2001) is that the causal powers of mental properties are included among (or are proper subsets of) the causal powers of the physical properties that realize them. (See also Macdonald and Macdonald 1995, Witmer, 2003, Yates, 2012, and Strevens, 2012, for related views.)

There has been substantial recent work on the causal exclusion problem, which, as noted earlier, arises for any non-reductive theory of mental states. (See the entry on mental causation, as well as Bennett 2007, and

Funkhouser 2007, for further discussion and extensive bibliographies.) But it is worth discussing a related worry about causation that arises exclusively for role-functionalism (and other dispositional theories), namely, the problem of “metaphysically necessitated effects” (Rupert 2006, Bennett 2007). If *pain* is functionally defined (either by an analytic or an empirical theory) as the state of being in some lower-level state or other that, in certain circumstances causes wincing, then it seems that the generalization that pain causes wincing (in those circumstances) is at best uninformative, since the state in question would not *be* pain if it didn’t. And, on the Humean view of causation as a contingent relation, the causal claim would be false. Davidson (1980b) once responded to a similar argument by noting that even if a mental state *M* is defined in terms of its production of an action *A*, it can often be redescribed in other terms *P* such that ‘*P* caused *A*’ is not a logical truth. But it’s unclear whether any such redescriptions are available to role (vs. realizer) functionalists.

Some theorists (e.g. Antony and Levine 1997) have responded by suggesting that, though mental states may be defined in terms of some of their effects, they have *other* effects that do not follow from those definitions which can figure into causal generalizations that are contingent, informative, and true. For example, even if it follows from a functional definition that pain causes wincing (and thus that the relation between pain and wincing cannot be truly causal), psychologists may discover, say, that pain produces resilience (or submissive behavior) in human beings. One might worry, nonetheless, that functional definitions threaten to leave too many commonly cited generalizations outside the realm of contingency, and thus causal explanation: surely, we may think, we want to affirm claims such as “pain causes wincing”. Such claims could be affirmed, however, if (as seems likely) the most plausible functional theories define sensations such as pain in terms of a small subset of their distinctive *psychological*, rather than behavioral, effects (see section 4.2).

A different line of response to this worry (Shoemaker 1984d, 2001) is to deny the Humean account of causation altogether, and contend that causal relations are themselves metaphysically necessary, but this remains a minority view. (See also Bird, 2002, and Latham, 2011, for further discussion.)

5.3 Functionalism and Introspective Belief

Another important question concerns the beliefs that we have about our own “occurrent” (as opposed to dispositional) mental states such as thoughts, sensations, and perceptions. We seem to have immediately available, non-inferential beliefs about these states, and the question is how this is to be explained if mental states are identical with functional properties.

The answer depends on what one takes these introspective beliefs to involve. Broadly speaking, there are two dominant views of the matter (but see Peacocke 1999, Ch. 5 for further alternatives). One popular account of introspection – the “inner sense” model on which introspection is taken to be a kind of “internal scanning” of the contents of one’s mind (Armstrong 1968) – has been taken to be unfriendly to functionalism, on the grounds that it’s hard to see how the objects of such scanning could be second-order relational properties of one’s neural states (Goldman 1993). Some theorists, however, have maintained that functionalism can accommodate the special features of introspective belief on the “inner sense” model, since it would be only one of many domains in which it’s plausible to think that we have immediate, non-inferential knowledge of causal or dispositional properties (Armstrong 1993; Kobes 1993; Sterelney 1993). A full discussion of these questions goes beyond the scope of this entry, but the articles cited above are just three among many helpful pieces in the Open Peer Commentary following Goldman (1993), which provides a good introduction to the debate about this issue.

Another account of introspection, identified most closely with Shoemaker (1996a,b,c,d), is that the immediacy of introspective belief follows from the fact that occurrent mental states and our introspective beliefs about them are functionally interdefined. For example, one satisfies the definition of being in pain only if one is in a state that tends to cause (in creatures with the requisite concepts who are considering the question) the belief that one is in pain, and one believes that one is in pain only if one is in a state that plays the belief role, and is caused directly by the pain itself. On this account of introspection, the immediacy and non-inferential nature of introspective belief is not merely compatible with functionalism, but required by it.

But there is an objection, most recently expressed by George Bealer (1997; see also Hill 1993), that, on this model an introspective belief can only be defined in one of two unsatisfactory ways: either as a belief produced by a (second-order) functional state specified (in part) by its tendency to produce that very type of belief – which would be circular – or as a belief about the first-order *realization* of the functional state, rather than that state itself. Functionalists have suggested, however (Shoemaker 2001, McCullagh 2000, Tooley 2001), that there is a way of understanding the conditions under which beliefs can be caused by, and thus be about, one’s second-order functional states that permits mental states and introspective beliefs about them to be non-circularly defined (but see Bealer 2001, for a skeptical response). A full treatment of this objection involves the more general question of whether second-order properties can have causal efficacy, and is thus beyond the scope of this discussion (see section 5.2 and the mental causation entry). But even if this objection can ultimately be deflected, it suggests that special attention must be paid to the functional characterizations of “self-directed” mental states.

5.4 Functionalism and the Norms of Reason

Yet another objection to functionalist theories of any sort is that they do not capture the interrelations that we take to be definitive of beliefs, desires, and other intentional states. Whereas even analytic functionalists hold that mental states – and also their contents – are implicitly defined in terms of their (causal or probabilistic) roles in *producing* behavior, these critics understand intentional states to be implicitly defined in terms of their roles in *rationalizing*, or *making sense of*, behavior. This is a different enterprise, they claim, since rationalization, unlike causal explanation, requires showing how an individual's beliefs, desires, and behavior conform, or at least approximate, to certain a priori *norms* or *ideals* of theoretical and practical reasoning – prescriptions about which beliefs and desires we *should* have, how we *should* reason, or what, given our beliefs and desires, we *ought* to do. (See Davidson 1980c, Dennett 1978, and McDowell 1985 for classic expressions of this view.) Thus the defining (“constitutive”) normative or rational relations among intentional states expressed by these principles cannot be expected to correspond to causal and probabilistic relations among our internal states, sensory stimulations and behavior, since they constitute a kind of explanation that has sources of evidence and standards for correctness that are different from those of empirical theories (Davidson 1980c). One can't, that is, extract facts from values.

Thus, although attributions of mental states can in some sense explain behavior by permitting an observer to “interpret” it as making sense, they should not be expected to denote entities that figure in empirical generalizations, either common sense or scientific. (This is not to say, these theorists stress, that there are no causes, or empirical laws of, behavior. These, however, will be expressible only in the vocabularies of the neurosciences, or other lower-level sciences, and not as relations among beliefs, desires and behavior.)

Functionalists have replied to these worries in different ways. Many just deny the intuition behind the objection, and maintain that even the strictest conceptual analyses of our intentional terms and concepts purport to define them in terms of their bona-fide causal roles, and that any norms they reflect are explanatory rather than prescriptive. They argue, that is, that if these generalizations are idealizations, they are the sort of idealizations that occur in any scientific theory: just as Boyle's Law depicts the relations between the temperature, pressure, and volume of a gas under certain ideal experimental conditions, our a priori theory of the mind consists of descriptions of what normal humans *would* do under (physically specifiable) ideal conditions, not prescriptions as to what they *should*, or are *rationally required*, to do.

Other functionalists agree that we may advert to various norms of inference and action in attributing beliefs and desires to others, but deny that there is any in principle incompatibility between normative and empirical explanations. They argue that if there are causal relations among beliefs, desires, and behavior that even approximately mirror the norms of rationality, then the attributions of intentional states can be empirically confirmed (Fodor 1990; Rey 1997). In addition, many who hold this view suggest that the principles of rationality that intentional states must meet are quite minimal, and comprise at most a weak set of constraints on the contours of our theory of mind, such as that people can't, in general, hold (obviously) contradictory beliefs, or act against their (sincerely avowed) strongest desires (Loar 1981). Still others suggest that the intuition that we attribute beliefs and desires to others according to rational norms is based on a fundamental mistake; these states are attributed not on the basis of whether they rationalize the behavior in question, but whether those subjects can be seen as using principles of inference and action sufficiently like our own – be they rational, like Modus Ponens, or irrational, like the Gambler's Fallacy or the now familiar instances of “predictable irrationality” documented in Kahneman, 2011. (See Stich 1981, and Levin

1988, for discussion of this question, and for a more general debate about the compatibility of normative and psychological principles, see Rey, 2007, and Wedgwood, 2007. See also Glüer and Wikforss, 2009, 2013, and for further discussion, the entry on the normativity of meaning and content.)

Nonetheless, although many functionalists argue that the considerations discussed above show that there is no in principle bar to a functionalist theory that has empirical force, these worries about the normativity of intentional ascription continue to fuel skepticism about functionalism (and, for that matter, any scientific theory of the mind that uses intentional notions).

In addition to these general worries about functionalism, there are particular questions that arise for functional characterizations of experiential or phenomenal states. These questions will be discussed in the following section.

5.5 Functionalism and the Problem of Qualia

Functionalist theories of all varieties – whether analytic or empirical, role or realizer – attempt to characterize mental states exclusively in relational, specifically causal, terms. A common and persistent objection, however, is that no such characterizations can capture the phenomenal character, or “qualia”, of experiential states such as perceptions, emotions, and bodily sensations, since they would leave out certain essential properties of those experiential states, namely, “what it’s like” (Nagel 1974) to have them. The next three sections will present the most serious worries about the ability of functionalist theories to give an adequate characterization of these states. (These worries, of course, will extend to intentional states, if, as some philosophers argue, “what it’s like” to have them is among their essential properties as well. (See Searle 1992, G. Strawson, 1986, Horgan

and Tienson, 2002, Kriegel 2003, Pitt 2008, Mendelovici, 2018, for presentations of this view, and see Bayne and Montague, 2011, and Smithies, 2013a and 2013b for more general discussions of whether intentional states possess phenomenal character – often called “cognitive phenomenology” – and if so, what, more precisely, it is.)

5.5.1 Inverted and Absent Qualia

The first to be considered are the “absent” and “inverted” qualia objections most closely associated with Ned Block (1980b; see also Block and Fodor 1972). The “inverted qualia” objection to functionalism maintains that there could be an individual who (for example) is in a state that satisfies the functional definition of our experience of red, but is experiencing green instead – and similarly for all the colors in the spectrum. It is a descendant of the claim, discussed by philosophers from Locke to Wittgenstein, that there could be an individual with an “inverted spectrum” who is behaviorally indistinguishable from someone with normal color vision; both objections trade on the contention that the purely relational characterizations in question cannot make distinctions among distinct experiences with isomorphic causal patterns. (Nida-Rümelin, 1996, argues that the science of color vision leaves open the possibility that there could be functionally equivalent red-green “inverts”, but even if inverted qualia are not really an empirical possibility for human beings, given certain asymmetries in our “quality space” for color, and differences in the relations of color experiences to other mental states such as emotions (Hardin 1988), it seems possible that there are creatures with perfectly symmetrical color quality spaces for whom a purely functional characterization of color experience would fail.)

A related objection, the “absent qualia” objection, maintains that there could be creatures functionally equivalent to normal humans whose mental states have no qualitative character at all. In his well-known

“Chinese nation” thought-experiment, Block (1980b) invites us to imagine that the population of China (chosen because its size approximates the number of neurons in a typical human brain) is recruited to duplicate his functional organization for a period of time, receiving the equivalent of sensory input from an artificial body and passing messages back and forth via satellite. Block argues that such a “homunculi-headed” system would not have mental states with any qualitative character (other than the qualia possessed by the individuals themselves), and thus that there could be states functionally equivalent to sensations or perceptions that lack their characteristic “feels”. Conversely, some argue that functional role is not *necessary* for qualitative character: for example, it seems that one could have mild, but distinctive, twinges that have no typical causes or characteristic effects.

All these objections purport to have characterized a creature with the functional organization of normal human beings, but without any, or the right sort, of qualia (or vice versa), and thus to have produced a counterexample to functional theories of experiential states. One line of response, initially advanced by Sydney Shoemaker (1994b), is that although functional duplicates of ourselves with inverted qualia may be possible, duplicates with absent qualia are not, since their possibility leads to untenable skepticism about the qualitative character of one’s own mental states. This argument has been challenged, however (Block 1980b; but see Shoemaker’s response in 1994d, and Balog, 1999, for a related view), and the more common response to these objections – particularly to the absent qualia objection – is to question whether scenarios involving creatures such as Block’s Chinese nation provide genuine counterexamples to functionalist theories of experiential states.

For example, some theorists (Dennett 1978; Levin 1985; Van Gulick 1989) argue that these scenarios provide clear-cut counterexamples only to *crude* functional theories, and that attention to the subtleties of more

sophisticated characterizations will undermine the intuition that functional duplicates of ourselves with absent qualia are possible (or, conversely, that there are qualitative states without distinctive functional roles). The plausibility of this line of defense is often questioned, however, since there is tension between the goal of increasing the sophistication (and thus the individuating powers) of the functional definitions, and the goal (for analytic functionalists) of keeping these definitions within the bounds of the platitudinous (though see Section 4.2), or (for psychofunctionalists) broad enough to be instantiable by creatures other than human beings. A related suggestion is that absent qualia seem possible only because of our imaginative shortcomings, in particular, that it is hard for us to attend, at any one time, to all the relevant features of even the simplest functional characterization of experiential states; another is that the intuition that these creatures lack qualia is based on prejudice – against creatures with unfamiliar shapes and extended reaction times (Dennett 1978), or creatures with parts widely distributed in space (Lycan, 1981, Schwitzgebel 2015 and commentary).

There are other responses to analogous absent qualia arguments that are prominent in the literature, but the target of those arguments is broader. Block’s argument was initially presented as a challenge exclusively to functionalist theories, both analytic and empirical, and not generally to physicalistic theories of experiential states; the main concern was that the purely relational resources of functional description were incapable of capturing the intrinsic qualitative character of states such as feeling pain, or seeing red. (Indeed, in Block’s 1980b, p. 291, he suggests that qualitative states may best be construed as “composite state[s] whose components are a *quale* and a [functional state],” and adds, in an endnote (note 22) that the *quale* “might be identified with a physico-chemical state”.) But there are similar objections that have been raised against all physicalistic theories of experiential states that are important to consider in

evaluating the prospects for functionalism. These will be discussed in the next two sections.

5.5.2 Functionalism, Zombies, and the “Explanatory Gap”

One important objection, advanced by (among others) Kripke (1972) and Chalmers (1996a), derives from Descartes’s well-known argument in the *Sixth Meditation* (1641) that since he can clearly and distinctly conceive of himself existing apart from his body (and vice versa), and since the ability to clearly and distinctly conceive of things as existing apart guarantees that they are in fact distinct, he is in fact distinct from his body.

Chalmers’s version of the argument (1996a, 2002), known as the “Zombie Argument”, has been particularly influential. The first premise of this argument is that it is *conceivable*, in a special, robust, “positive” sense, that there are molecule-for-molecule duplicates of oneself with no qualia (call them “zombies”, following Chalmers 1996a). The second premise is that scenarios that are “positively” conceivable in this way represent real, metaphysical, possibilities. Thus, he concludes, zombies are possible, and functionalism – or, more broadly, physicalism – is false. The force of the Zombie Argument is due in large part to the way Chalmers defends its two premises; he provides a detailed account of just what is required for zombies to be conceivable, and also an argument as to why the conceivability of zombies entails their possibility (see also Chalmers 2002, 2006, 2010, Ch. 6, and Chalmers and Jackson 2002). This account, based on a more comprehensive theory of how we can evaluate claims about possibility and necessity known as “two-dimensional semantics”, reflects an increasingly popular way of thinking about these matters, but remains controversial. (For alternative ways of explaining conceivability, see Kripke (1986), Hart (1988); for criticism of the argument from two-dimensional semantics, see Yablo 2000, 2002, Bealer 2002, Stalnaker 2002, Soames 2004, Byrne and Prior 2006; but see also Chalmers 2006.)

In a related challenge, Joseph Levine (1983, 1993, 2001) argues that, even if the conceivability of zombies doesn’t entail that functionalism (or more broadly, physicalism) is false, it opens an “explanatory gap” not encountered in other cases of inter-theoretical reduction, since the qualitative character of an experience cannot be deduced from any physical or functional description of it. Such attempts thus pose, at very least, a unique epistemological problem for functionalist (or physicalist) reductions of qualitative states.

In response to these objections, analytic functionalists contend, as they did with the inverted and absent qualia objections, that sufficient attention to what is required for a creature to duplicate our functional organization would reveal that zombies are not really conceivable, and thus there is no threat to functionalism and no explanatory gap. A related suggestion is that, while zombies may *now* seem conceivable, we will eventually find them inconceivable, given the growth of empirical knowledge, just as we now find it inconceivable that there could be H₂O without water (Yablo 1999). Alternatively, some suggest that the inconceivability of zombies awaits the development of new *concepts* that can provide a link between our current phenomenal and physical concepts (Nagel 1975, 2000), while others (McGinn 1989) agree, but deny that humans are capable of forming such concepts.

None of these responses, however, would be an effective defense of Psychofunctionalism, which does not attempt to provide analyses of experiential concepts (or suggest that there would, or could, be any to come). But there is an increasingly popular strategy for defending physicalism against these objections that could be used to defend Psychofunctionalism, namely, to concede that there can be no conceptual analyses of qualitative concepts (such as *what it’s like to see red* or *what it’s like to feel pain*) in purely functional terms, and focus instead on

developing arguments to show that the conceivability of zombies neither implies that such creatures are possible nor opens up an explanatory gap.

One line of argument (Block and Stalnaker 1999; Yablo 2000) contends that the conceivability of (alleged) counterexamples to psycho-physical or psycho-functional identity statements, such as zombies, has analogues in other cases of successful inter-theoretical reduction, in which the lack of conceptual analyses of the terms to be reduced makes it conceivable, though not possible, that the identities are false. However, the argument continues, if these cases routinely occur in what are generally regarded as successful reductions in the sciences, then it's reasonable to conclude that the conceivability of a situation does not entail its possibility.

A different line of argument (Horgan 1984; Loar 1990; Lycan 1990; Hill 1997, Hill and McLaughlin 1999, Balog 1999, Levin 2018) maintains that, while generally the conceivability of a scenario entails its possibility, scenarios involving zombies stand as important exceptions. The difference is that the phenomenal, or "what it's like", concepts used to describe the properties of experience that we conceive zombies to lack are significantly different from the third-personal, discursive concepts of our common sense and scientific theories such as *mass*, *force* or *charge*; they comprise a special class of non-discursive, first-personal, perspectival representations of those properties. Whereas conceptually independent third-personal concepts *x* and *y* may reasonably be taken to express metaphysically independent properties, or modes of presentation, no such metaphysical conclusions can be drawn when one of the concepts in question is third-personal and the other is *phenomenal*, since these concepts may merely be picking out the same properties in different ways. Thus, the conceivability of zombies, dependent as it is on our use of phenomenal concepts, provides no evidence of their metaphysical possibility.

Key to this line of defense is the claim that these special phenomenal concepts can denote functional (or physical) properties without expressing some irreducibly qualitative modes of presentation of them, for otherwise it couldn't be held that these concepts do in fact apply to our functional (or physical) duplicates, even though it's conceivable that they don't. This, not surprisingly, has been disputed, and there is currently much discussion in the literature about the plausibility of this claim. If this line of defense is successful, however, it can also provide a response to the "Distinct Property Argument", discussed in section 3.3. (See Chalmers 1999, Holman 2013 for criticism of this view, but see the responses of Loar 1999, Hill & McLaughlin 1999, Balog 2012, Levin 2008, 2018, Diaz-Leon, 2010, 2014; see also see Levin 2002, 2008, and Schroer 2010, for the presentation, if not endorsement, of a hybrid view.)

5.5.3 Functionalism and the Knowledge Argument

In another important, related, challenge to functionalism (and, more generally, physicalism), Thomas Nagel (1974) and Frank Jackson (1982) argue that a person could know all the physical and functional facts about a certain type of experience and still not "know what it's like" to have it. This is known as the "Knowledge Argument", and its conclusion is that there are certain properties of experiences – the "what it's like" to see red, feel pain, or sense the world through echolocation – which cannot be identified with functional (or physical) properties. Though neither Nagel (2000) nor Jackson (1998) now endorse this argument, many philosophers contend that it raises special problems for any physicalistic view (see Alter 2007, and, in response, Jackson 2007).

An early line of defense against these arguments, endorsed primarily but not exclusively by a priori functionalists, is known as the "Ability Hypothesis". (Nemirow 1990, 2007, Lewis 1990, Levin 1986). "Ability" theorists suggest that knowing what it's like to see red or feel pain is

merely a sort of *practical* knowledge, a “knowing how” (to imagine, remember, or re-identify, a certain type of experience) rather than a knowledge of propositions or facts. (See Tye 2000, for a summary of the pros and cons of this position; for further discussion, see the essays in Ludlow, Nagasawa, and Stoljar 2004.) An alternative view among contemporary functionalists is that coming to know what it’s like to see red or feel pain is indeed to acquire propositional knowledge uniquely afforded by experience, expressed in terms of first-personal concepts of those experiences. But, the argument continues, this provides no problem for functionalism (or physicalism), since these special first-personal concepts need not denote, or introduce as “modes of presentation”, any irreducibly qualitative properties. This view, of course, shares the strengths and weaknesses of the analogous response to the conceivability arguments discussed above. If it is plausible, however, it can also challenge the argument of some philosophers (e.g., Chalmers, 2002, Stoljar, 2001, and Alter, 2016) that maintains that no physicalistic theory, not even fundamental physics, can provide anything but a relational characterization of the items in their domains – their structure and dynamics – and concludes that no physicalistic theory can capture what seems, from the inside, to be the intrinsic, non-relational properties of our experiential states. (See the essays in Alter and Nagasawa, 2015, Part III, for further discussion.)

There is one final strategy for defending a functionalist account of qualitative states against all of these objections, namely, eliminativism (Dennett 1988; Rey 1997, Pereboom 2011, Frankish 2016). One can, that is, deny that there are any such things as irreducible *qualia*, and maintain that the conviction that such things do, or perhaps even *could*, exist is due to illusion – or confusion.

6. The Future of Functionalism

In the last part of the 20th century, functionalism stood as the dominant theory of mental states. Like behaviorism, functionalism takes mental states out of the realm of the “private” or subjective, and gives them status as entities open to scientific investigation. But, in contrast to behaviorism, functionalism’s characterization of mental states in terms of their roles in the production of behavior grants them the causal efficacy that common sense takes them to have. Moreover, functionalism, in contrast to the psychophysical identity thesis, seems to offer an account of mental states that is friendly to materialism without limiting the class of those with minds to creatures with brains like ours.

More recently, however, there has been a resurgence of interest in the psycho-physical (type-) identity thesis. This is fueled in part by the observation that in the actual practice of neuroscience, neural states are type-individuated more coarsely than early identity theorists such as Place, Feigl, and Smart assumed, and the contention that, contrary to the claims of early defenders of functionalism (e.g. Putnam), there are relatively few extant creatures that are physically unlike humans but share our functional organization. If this is so, then it may well be that many of our genuine functional equivalents that differ from us in their fine-grained neurophysiological make-up can nonetheless share our neural, as well as mental, states, and thus that the psycho-physical identity thesis can claim some of the scope once thought to be exclusive to functionalism. (See Bechtel 2012, Bickle 2012, McCauley 2012, Shapiro & Polger 2012, and Polger & Shapiro 2016.)

The plausibility of this thesis, of course, depends on whether or not these underlying similarities would in fact be (coarse grained) *neural* similarities, and not (fine-grained) psycho-functional similarities. In addition, functionalists can argue that there are *possible* creatures, both

biological and (especially) non-biological, that are functionally equivalent to us but do not possess even our coarse-grained neural properties. If so, and if these creatures can plausibly be regarded as sharing our mental states – to be sure, a controversial thesis – then even if neural states can be individuated more coarsely, some variety of functionalism will retain its claim to greater universality than the identity thesis.

Another source of skepticism about the relative universality of functionalist (versus type-identity) theories derives from the suggestion that – at least in humans – the properties that could play the relevant roles in our functional architecture are not exclusively electrical impulses of our neurons, but a combination of these impulses *plus* other features of the brain and body, including the endocrine system. (See Cao 2022 and Damasio 1999.) If this is so, then there may be fewer extant examples of creatures that are functional, but not physical, duplicates of humans, and it may be less likely that such creatures are even nomologically possible. They may, of course, be metaphysically possible, existing in fairly distant possible worlds, but there is debate about whether this is plausible, and if so, whether it would seriously undermine the psycho-physical identity theory.

On the other hand, some critics of the psycho-physical identity theory contend that only a theory that characterizes mental states as higher-order properties, with causal powers not possessed by the states or properties that realize them, can preserve the explanatory hierarchy modeled on the relations between biology and chemistry, and chemistry and physics. The identification of mental states with functional roles would do the trick, whether or not our mental states are in fact multiply realized. (See Gillett 2002, Aizawa and Gillett 2007, 2009, and the essays in DeJoong and Shouten 2007 for further discussion; see also the entry on multiple realizability.) These questions remain the subject of active debate.

Yet another question is whether Role Functionalism, by itself, is as friendly to materialism as its initial advocates had suggested – even if the functional roles to be identified with mental states are realized exclusively by physical states. It is generally agreed that the mere supervenience of the mental on the physical – that is, the impossibility of there being a mental difference without some physical difference – leaves open the possibility that mental properties are emergent properties with additional causal powers, whereas the realization of mental states by physical states eliminates this possibility. However, some say, while others deny, that materialism requires something more, namely, an explanation of how (or why) those physical properties can occupy the relevant functional roles. This too is a contentious question. (See, for further discussion, Endicott 2016 and Shaffer 2021).

There remain other substantive questions about functionalism. One is whether it can provide an adequate account of all mental states. Most discussions of the prospects for functionalism focus on its adequacy as an account of familiar experiential states such as sensational and perceptual experiences, and familiar intentional states such as beliefs and desires. But what about emotions – and moods? Although there is some discussion of these states in the functionalism literature, (e.g. Rey 1990, Nussbaum 2003, deHooze, et al. 2011) there is increasing interest in these questions, and more work needs to be done. In addition, there is increasing interest in determining whether there can be plausible functional characterizations of non-standard perceptual experiences, such as synesthesia, and various sorts of altered states of consciousness that can arise from the ingestion of drugs, natural hallucinogenic substances, or focused meditation. (See Gray et al. 2002, 2004, and Deroy 2017, for discussions of synesthesia. For general discussions of altered states of consciousness, see Velmans and Schneider (eds.) 2007, and Thompson 2015.)

Another question is whether functional theories can accommodate non-standard views about the location of mental states, such as the hypothesis of extended cognition, which maintains that certain mental states such as memories – and not just their representational contents – can reside outside the head. This question has implications not only for the viability of a functionalist characterization of memory, but also of beliefs, emotions, hallucinations, and moods. (See Clark and Chalmers 2002, Clark, 2008, Adams and Aizawa 2008, Rupert 2009, Sprehack 2009, Byrne and Manzotti 2022, and the essays in Menary 2010, for further discussion.)

Yet another question of increasing interest to philosophers is whether groups of individuals, or entire communities, can possess mental states (e.g. beliefs, desires, intentions, moods, and emotions) that cannot be reduced to the mental states of the individuals in those communities. (See Gilbert 2013.) Is it merely metaphorical to say that a community (and not just the individuals in it) is optimistic, or conservative, or believes that human activities cause climate change, or remembers the 60s? Or can claims like this be literally true – and if so, how; does it matter how many members of the community in question individually possess the belief, intention, or emotion in question? It is easy to see, moreover, that answers to these questions would have implications not only for theories of the nature of mental states, but also for theories of self-knowledge and knowledge of other minds. They would also have implications for theories of moral evaluation: Desires and intentions lead to actions, and if a community intends to help resettle new immigrants or block access to a public beach, then whom (or what) can we legitimately praise and blame for these actions, and to whom (or what) do we have moral obligations?

In general, the sophistication of functionalist theories has increased since their introduction, but so has the sophistication of the objections to functionalism, especially to functionalist accounts of mental causation

(section 5.2), introspective knowledge (Section 5.3), and the qualitative character of experiential states (Section 5.5). For those unconvinced of the plausibility of dualism, however, and unwilling to restrict mental states to creatures physically like ourselves, the initial attractions of functionalism remain. The primary challenge for future functionalists, therefore, will be to meet these objections to the doctrine, either by articulating a functionalist theory in increasingly convincing detail, or by showing how the intuitions that fuel these objections can be explained away.

Bibliography

- Adams, F. and K. Aizawa, 2008, *The Bounds of Cognition*, Oxford: Wiley-Blackwell.
- Aizawa, K., 2008, “Neuroscience and Multiple realization: A Reply to Bechtel and Mundale”, *Synthese*, 167: 495–510.
- Aizawa, K. and Gillett, C., 2007, “Multiple realization and Methodology in Neuroscience and Psychology”, *British Journal of the Philosophy of Science*.
- Aizawa, K. and Gillett, C., 2009, “The Multiple realization of Psychological and Other Properties in the Sciences”, *Mind and Language*, 24(2): 181–208.
- Alter, T., 2007, “Does Representationalism Undermine the Knowledge Argument?”, in Alter and Walter (2007), 65–76.
- Alter, T., 2016, “The Structure and Dynamics Argument Against Materialism”, *Noûs*, 50(4): 794–815.
- Alter, T. and Y. Nagasawa, 2015, *Consciousness in the Physical World: Perspectives on Russellian Monism*, New York: Oxford University Press.
- Alter, T. and S. Walter, 2007, *Phenomenal Concepts and Phenomenal Knowledge*, New York: Oxford University Press.
- Antony, L. and J. Levine, 1997, “Reduction With Autonomy”, *Philosophical Perspectives*, 2: 83–105.

- Armstrong, D., 1968, *A Materialistic Theory of the Mind*, London: RKP.
- , 1981, *The Nature of Mind*, Brisbane: University of Queensland Press.
- , 1993, “Causes are perceived and introspected”, in *Behavioral and Brain Sciences*, 16(1): 29–29.
- Baker, L. R., 1995, “Metaphysics and mental Causation”, in Heil and Mele 1995, 75–96.
- Balog, K., 1999, “Conceivability, Possibility, and the Mind-Body Problem”, *Philosophical Review*, 108(4): 497–528.
- Balog, K., 2012, “In Defense of the Phenomenal Concepts Strategy”, *Philosophy and Phenomenological Research*, 84(1): 1–23.
- Bayne, T. and Montague, M. (eds.), 2011, *Cognitive Phenomenology*, Oxford: Oxford University Press.
- Bealer, G., 1997, “Self-Consciousness”, *Philosophical Review*, 106: 69–117.
- , 2001, “The self-consciousness argument: Why Tooley’s criticisms fail”, *Philosophical Studies*, 105(3): 281–307.
- , 2002, “Modal Epistemology and the Rationalist Renaissance”, in Gendler, T. and Hawthorne, J. (eds.), *Conceivability and Possibility*, Oxford: Oxford University Press, 71–126.
- Bechtel, W., 2012, “Identity, reduction, and conserved mechanisms: perspectives from circadian rhythm research”, in Gozzano and Hill 2012, 88–110.
- Bennett, K., 2003, “Why the Exclusion Problem Seems Intractable, and How, Just Maybe, to Tract It”, *Noûs*, 37(3): 471–497.
- , 2007, “Mental Causation”, *Philosophy Compass*, 2(2): 316–337.
- Bickle, J., 2012, “A brief history of neuroscience’s actual influences on mind-brain reductionism”, in Gozzano and Hill 2012, 43–65.
- Bird, A., 2002, “On Whether Some Laws are Contingent”, *Analysis*, 63(2): 257–270.
- Block, N., 1980a, *Readings in the Philosophy of Psychology, Volumes 1 and 2*, Cambridge, MA: Harvard University Press.
- , 1980b, “Troubles With Functionalism”, in Block 1980a, 268–305.
- , 1980c, “Are Absent Qualia Impossible?”, *Philosophical Review*, 89: 257–274.
- , 1986, “Advertisement for a Semantics for Psychology”, in French, Euling, and Wettstein (eds.), *Midwest Studies in Philosophy*, 10: 615–678.
- , 1990, “Inverted Earth”, in J. Tomberlin (ed.), *Philosophical Perspectives*, 4, Atascadero, CA: Ridgeview Press, 52–79.
- , 1997, “Anti-Reductionism Slaps Back”, *Philosophical Perspectives*, 11, Atascadero, CA: Ridgeview Press, 107–132.
- Block, N., and O. Flanagan and G. Guzeldere (eds.), 1997, *The Nature of Consciousness*, Cambridge, MA: MIT Press.
- Block, N. and J. Fodor, 1972, “What Psychological States Are Not”, *Philosophical Review*, 81: 159–181.
- Block, N. and R. Stalnaker, 1999, “Conceptual Analysis, Dualism, and the Explanatory Gap”, *Philosophical Review*, 108(1): 1–46.
- Braddon-Mitchell, D. and F. Jackson, 1996/2007, *Philosophy of Mind and Cognition*, Malden, MA: Blackwell Publishing.
- Burge, T., 1979, “Individualism and the Mental”, *Midwest Studies in Philosophy*, 4: 73–121.
- , 1995, “Mind-Body Causation and Explanatory Practice”, in Heil and Mele 1995, 97–120.
- Byrne, A., 2018, *Transparency and Self Knowledge*. New York: Oxford University Press.
- Byrne, A. and Manzotti, R., 2022, “Hallucination and its Objects”, *Philosophical Review*, 131(3): 327–359.
- Byrne, A. and Prior, J., 2006, “Bad Intensions”, in Garcia-Carpintero, M. and Macia, J. (eds.), *Two-Dimensional Semantics*, Oxford: Oxford University Press, 38.
- Cao, R., 2022, “Multiple realizability and the spirit of functionalism”, *Synthese*, 200(6): 1–31.

- Chalmers, D., 1996a, *The Conscious Mind*, Oxford: Oxford University Press.
- , 1996b, “Does a Rock Implement Every Finite State Automaton?”, *Synthese*, 108: 309–333.
- , 1999, “Materialism and the Metaphysics of Modality”, in *Philosophy and Phenomenological Research*, 59(2): 473–496.
- , 2002, “Does Conceivability Entail Possibility?”, in Gendler and Hawthorne 2002, 145–200.
- , 2006, “The Foundation of Two-Dimensional Semantics”, in M. Garcia-Carpintero, M. and J. Macia (eds.), *Two-Dimensional Semantics*, Oxford: Oxford University Press, 55–140.
- , 2010, *The Character of Consciousness*, New York: Oxford University Press.
- Chisholm, R., 1957, *Perceiving*, Ithaca: Cornell University Press.
- Chomsky, N., 1959, “Review of Skinner’s *Verbal Behavior*”, *Language*, 35: 26–58; reprinted in Block 1980, 48–63.
- Churchland, P., 1981, “Eliminative Materialism and Propositional Attitudes”, *Journal of Philosophy*, 78: 67–90.
- , 2005, “Functionalism at Forty: A Critical Retrospective”, *Journal of Philosophy*, 102: 33–50.
- Clark, A., 2008, *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*, Oxford: Oxford University Press.
- Clark, A. and Chalmers, D., 2002, “The Extended Mind”, in Chalmers (2002). *Philosophy of Mind*, New York: Oxford University Press, 643–651.
- Crane, T., 1995, “The Mental Causation Debate”, *Proceedings of the Aristotelian Society*, 69 (Supplement): 211–236.
- Damasio, A., 1999, *The Feeling of What Happens*. New York: Harcourt.
- Davidson, D., 1980a, *Essays on Actions and Events*, New York: Oxford University Press.
- , 1980b, “Actions, Reasons, and Causes”, in Davidson 1980a.
- , 1980c, “Mental Events”, in Davidson 1980a.
- deHooze, I., Zeelenberg, M., and Breugelmans, S., 2011, “A Functionalist Account of Shame-Induced Behavior”, *Cognition and Emotion*, 25(5): 939–46.
- De Joong, H.L. and M. Shouten (eds.), 2012, *Rethinking Reduction*, Oxford: Blackwell Publishers.
- Dennett, D., 1978a. “Intentional Systems”, in Dennett 1978c, 3–22.
- , 1978b. “Towards a Cognitive Theory of Consciousness”, in Dennett 1978c, 149–173.
- , 1978c. *Brainstorms*, Montgomery, VT: Bradford Books.
- , 1988, “Quining Qualia”, in A. Marcel and E. Bisiach (eds.), *Consciousness in Contemporary Science*, New York: Oxford University Press, 43–77.
- Deroy, O. (ed.), 2017, *Sensory Blending: On Synaesthesia and Related Phenomena*, Oxford: Oxford University Press.
- Diaz-Leon, E., 2010, “Can Phenomenal Concepts Explain the Epistemic Gap?”, *Mind*, 119(476): 533:51.
- , 2014, “Do A Posteriori Physicalists Get our Phenomenal Concepts Wrong?”, *Ratio*, 27(1): 1–16.
- Ebert, P., and Smith, M., 2012, “Introduction: Outright Belief and Degrees of Belief”, *Dialectica*, 66(3): 305–308.
- Endicott, R., 2016, “Functionalism, Superduperfunctionalism, and Physicalism: Lessons From Supervenience”, *Synthese*, 193(7): 2205–2235.
- Feigl, H., 1958, “The ‘Mental’ and the ‘Physical’”, in Feigl, H., Scriven, M. and Maxwell, G., *Minnesota Studies in the Philosophy of Science: Concepts, Theories, and the Mind-Body Problem*, Minneapolis: Minnesota University Press, 370–409.
- Field, H., 1980, “Mental Representation”, in Block 1980, 78–114.
- Fodor, J., 1968, *Psychological Explanation*, New York: Random House.
- , 1975, *The Language of Thought*, New York: Crowell.

- , 1990a. “Fodor’s Guide to Mental Representation”, in J. Fodor, *A Theory of Content and Other Essays*, Cambridge, MA: MIT Press, 3–29.
- , 1990b. “Making Mind Matter More”, in J. Fodor, *A Theory of Content and Other Essays*, Cambridge, MA: MIT Press, 137–159.
- , 1994, *The Elm and the Expert*, Cambridge, MA: MIT Press.
- Frankish, K., 2004, *Mind and Supermind*, Cambridge: Cambridge University Press.
- , 2016, “Illusionism as a Theory of Consciousness”, *Journal of Consciousness Studies*, 23(11–12): 11–39.
- Frege, G., 1892/1952, “On Sense and Reference”, in P. Geach and M. Black (eds.), *Translations from the Work of Gottlob Frege*, Oxford: Blackwell.
- Funkhouser, E., 2007, “Multiple Realizability”, *Philosophy Compass*, 2(2): 303–315.
- Geach, P., 1957, *Mental Acts*, London: RKP.
- Gendler, T., 2008, “Belief and Alief”, *Journal of Philosophy*, 105(10): 634–663.
- Gendler, T. and Hawthorne, J (eds.), 2002, *Conceivability and Possibility*, Oxford: Oxford University Press.
- Gertler, B., 2007, “Overextending the Mind”, in B. Gertler and L. Shapiro (eds.), *Arguing About the Mind*, London: Routledge, 192–206.
- Gilbert, M., 2013, *Joint Commitment: How We Make the Social World*, New York: Oxford University Press.
- Gillet, C., 2002, “The dimensions of realization. A critique of the Standard view”, *Analysis*, 62(4): 316–323.
- Glüer, K. and Wikforss, Å., 2009, “Against Content Normativity”, *Mind*, 118: 31–70.
- , 2013, “Against Belief Normativity”, in T. Chan (ed.), *The Aim of Belief*, Oxford: Oxford University Press, 121–146.

- Goldman, A., 1993, “The psychology of folk psychology”, in *Behavioral and Brain Sciences*, 16(1): 15–28.
- Gozzano, S. and C. Hill, 2012, *New Perspectives on Type Identity*, New York: Cambridge University Press.
- Gray, J.A., and S. Chopping, J. Nunn, D. Parslow, L. Gregory, S. Williams, M.J. Brammer, and S. Baron-Cohen, 2002, “Implications of Synaesthesia for Functionalism”, *Journal of Consciousness Studies*, 9(12): 5–31.
- Hardin, C.L., 1988, *Color for Philosophers: Unweaving the Rainbow*, Indianapolis: Hackett.
- Harman, G., 1973, *Thought*, Princeton, NJ: Princeton University Press.
- Hart, W., 1988, *The Engines of the Soul*, Cambridge: Cambridge University Press.
- Hill, C., 1991, *Sensations*, Cambridge: Cambridge University Press.
- , 1993, “Qualitative characteristics, type materialism, and the circularity of analytic functionalism”, in *Behavioral and Brain Sciences*, 16(1): 50–51.
- , 1997, “Imaginability, Conceivability, Possibility and the Mind-Body Problem”, *Philosophical Studies*, 87: 61–85.
- Hill, C. and B. McLaughlin, 1999, “There are Fewer Things in Reality than Dreamt of in Chalmers’ Philosophy”, in *Philosophy and Phenomenological Research*, 59(2): 445–454.
- Holman, E. L., 2013, “Phenomenal concepts as bare recognitional concepts: harder to debunk than you thought...but still possible”, *Philosophical Studies*, 164: 807–827.
- Horgan, T., 1984, “Jackson on Physical Information and Qualia”, *Philosophical Quarterly*, 34: 147–183.
- , 1993, “From supervenience to superdupervenience: meeting the demands of a material world”, *Mind*, 102(408): 555–86.
- Horgan, T. and J. Tienson, 2002, “The Intentionality of Phenomenology and the Phenomenology of Intentionality”, in Chalmers, D. (ed.)

- (2002). *Philosophy of Mind*, New York. Oxford University Press, 520–533.
- Horgan, T. and J. Woodward, 1985, “Folk Psychology is Here to Stay”, *Philosophical Review*, 94: 197–226.
- Huber, F. and Schmidt-Peri, C. (eds.), 2009, *Degrees of Belief*. Dordrecht: Springer, 75–93.
- Jackson, F., 1982, “Epiphenomenal Qualia”, *Philosophical Quarterly*, 32: 127–136.
- , 1996, “Mental Causation”, *Mind*, 105(419): 377–413.
- , 1998, “Postscript on Qualia”, in F. Jackson, *Mind, Method, and Conditionals: Selected Essays*, London: Routledge, 76–79.
- , 2007, “The Knowledge Argument, Diaphanousness, Representationalism”, in Alter and Walter 2007, 52–64.
- Kahneman, D., 2011, *Thinking, Fast and Slow*, New York: Farrar, Strauss and Giroux.
- Kim, J., 1989, “Mechanism, Purpose, and Explanatory Exclusion”, in J. Kim, *Supervenience and Mind*, Cambridge, Cambridge University Press.
- , 1998, *Mind in a Physical World*, Cambridge, MA: Bradford.
- , 2007, “Causation and Mental Causation”, in McLaughlin and Cohen 2007, 227–242.
- Kobes, B., 1993, “Self-attributions help constitute mental types”, in *Behavioral and Brain Sciences*, 16(1): 54–56.
- Kriegel, U., 2003, “Is intentionality dependent on consciousness?”, *Philosophical Studies*, 116(3): 271–307.
- Kripke, S., 1980, *Naming and Necessity*, Cambridge, MA: Harvard University Press.
- Latham, Noa, 2011, “Are Fundamental Laws Necessary or Contingent?”, in J. Campbell, M. O’Rourke, and M. Slater (eds.), *Carving Nature at its Joints: Natural Kinds in Metaphysics and Science*, Cambridge, MA: MIT Press, 97–112.

- Levin, J., 1985, “Functionalism and the Argument from Conceivability”, *Canadian Journal of Philosophy*, 11 (Supplement): 85–104.
- , 1986, “Could Love be Like a Heatwave?”, *Philosophical Studies*, 49: 245–261.
- , 1998, “Must Reasons be Rational?”, in *Philosophy of Science*, 55: 199–217.
- , 2002, “Is Conceptual Analysis Needed for the Reduction of Qualitative States?”, *Philosophy and Phenomenological Research*, 64(3): 571–591.
- , 2008, “Taking Type-B Materialism Seriously”, *Mind and Language*, 23(4): 402–425.
- , 2018, “Once More Unto the Breach: Type B Physicalism, Phenomenal Concepts, and the Epistemic Gap”, *Australasian Journal of Philosophy*.
- , 2020, “‘Phenomenal States’ and the Scope of the Phenomenal Concepts Strategy”. Sullivan, A. (ed.). *Sensations, Thoughts, Language: Essays in Honor of Brian Loar*. New York: Routledge, pp. 289–313.
- Levine, J., 1983, “Materialism and Qualia: The Explanatory Gap”, *Pacific Philosophical Quarterly*, 64: 354–361.
- , 1993, “On Leaving Out What it’s Like”, in M. Davies and G. Humphries (eds.), *Consciousness*, Oxford: Blackwell Publishers, 121–136.
- , 2001, *Purple Haze*, New York: Oxford University Press.
- Lewis, D., 1966, “An Argument for the Identity Theory”, *Journal of Philosophy*, 63: 17–25.
- , 1972, “Psychophysical and Theoretical Identifications”, in Block 1980, 207–215.
- , 1980, “Mad Pain and Martian Pain”, in Block 1980, 216–222.
- , 1990, “What Experience Teaches”, in Lycan 1990, 499–519.

- Loar, B., 1981, *Mind and Meaning*, Cambridge: Cambridge University Press.
- , 1987, “Social Content and Psychological Content”, in P. Grimm and D. Merrill (eds.), *Contents of Thought*, Tucson: University of Arizona Press, 99–110.
- , 1997, “Phenomenal States” (revised version), in Block, *et al.*, 1997, 597–616.
- , 1999, “David Chalmers’s *The Conscious Mind*”, in *Philosophy and Phenomenological Research*, 59(2): 439–444.
- Loewer, B., 2002, “Comments on Jaegwon Kim’s Mind and the Physical World”, *Philosophy and Phenomenological Research*, 65(3): 555–662.
- , 2007, “Mental Causation, or Something Near Enough”, in McLaughlin and Cohen 2007, 243–264.
- Ludlow, P. and Y. Nagasawa and D. Stoljar, 2004, *There’s Something About Mary: Essays on Phenomenal Knowledge and Frank Jackson’s Knowledge Argument*, Cambridge, MA: MIT Press.
- Ludwig, K., 1998, “Functionalism, Causation, and Causal Relevance”, *Psyche*, 4(3) (March).
- Lycan, W., 1981, “Form, Function, and Feel”, *Journal of Philosophy*, 78(1): 24–50.
- Lycan, W., 1987, *Consciousness*, Cambridge MA: MIT Press.
- , 1990a, *Mind and Cognition*, Oxford: Blackwell Publishers.
- , 1990b, “The Continuity of Levels of Nature”, in Lycan 1990a, 77–96.
- Lyons, J., 2006, “In Defense of Epiphenomenalism”, *Philosophical Psychology*, 19(6): 767–749.
- Macdonald, C. and G. Macdonald, 1995, “How to be Psychologically Relevant”, in C. Macdonald and G. Macdonald (eds.), *Philosophy of Psychology: Debates on Psychological Explanation* (Volume 1), Oxford: Blackwell, 60–77.
- Malcolm, N., 1968, “The Conceivability of Mechanism”, *Philosophical Review*, 77: 45–72.
- McCauley, R., 2012, “About face: philosophical naturalism, the heuristic identity theory, and recent findings about prosopagnosia”, in Gozzano and Hill 2012, 186–206.
- McCullagh, M., 2000, “Functionalism and Self-Consciousness”, *Mind and Language*, 15(5): 481–499.
- McDowell, J., 1985, “Functionalism and Anomalous Monism”, in E. LePore and B. McLaughlin (eds.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, Oxford: Blackwell Publishers, 387–398.
- McGinn, C., 1989, “Can We Solve the Mind-Body Problem?”, *Mind*, 98: 349–66.
- McLaughlin, B., 2006, “Is Role-Functionalism Committed to Epiphenomenalism?”, *Consciousness Studies*, 13 (1–2): 39–66.
- , 2015, “Does Mental Causation Require Psychophysical Identities?”, in T. Horgan, M. Sabates, and D. Sosa (eds.), *Qualia and Causation in a Physical World: Themes From the Work of Jaegwon Kim*, Cambridge: Cambridge University Press.
- McLaughlin, B. and J. Cohen, 2007, *Contemporary Debates in the Philosophy of Mind*, Malden, MA: Blackwell Publishing.
- Melnyk, A., 2003, *A Physicalist Manifesto: Thoroughly Modern Materialism*, Cambridge: Cambridge University Press.
- Menary, R. (ed.), 2010, *The Extended Mind*. Cambridge, MA: MIT Press.
- Mendelovici, A., 2018, *The Phenomenal Basis of Intentionality*. New York: Oxford University Press.
- Moran, R., 2001, *Authority and Estrangement*. Princeton, NJ: Princeton University Press.
- Mumford, S., 1998, *Dispositions*, Oxford: Oxford University Press.
- Nagel, T., 1974, “What Is It Like To Be a Bat?”, *Philosophical Review*, 83: 435–450.

- , 2000, “The Psycho-physical Nexus”, in P. Boghossian and C. Peacocke (eds.), *New Essays on the A Priori*, New York: Oxford University Press, 434–471.
- Nelson, J., 1990, “Was Aristotle a Functionalist?”, *Review of Metaphysics*, 43(4): 791–802.
- Nemirow, L., 1990, “Physicalism and the Cognitive Role of Acquaintance”, in Lycan 1990a, 490–498.
- , 2007, “So This is What It’s Like: A Defense of the Ability Hypothesis”, in Alter and Walter 2007, 32–51.
- Ney, A., 2012, “The causal contribution of mental events”, in Gozzano and Hill 2012, 230–250.
- Nida-Rümelin, M., 1996, “Pseudonormal Vision: An Actual Case of Qualia Inversion?”, *Philosophical Studies*, 82: 145–57.
- Nussbaum, C., 2003, “Another Look at Functionalism and the Emotions”, *Brain and Mind*, 4(3): 353–383.
- Pautz, A., 2021, “Consciousness Meets Lewisian Interpretation Theory”, in Uriah Kriegel (ed.), *Oxford Studies in Philosophy of Mind*, Vol 1, New York: Oxford University Press, pp. 263–314.
- Peacocke, C., 1999, *Being Known*, Oxford: Oxford University Press.
- Pereboom, D., 2011, *Consciousness and the Prospects of Physicalism*. New York: Oxford University Press.
- Piccinini, G., 2004, “Functionalism, computationalism, and mental states”, *Studies in the History of Philosophy of Science*, 35: 811–833.
- Pitt, D., 2004, “The phenomenology of cognition, or, what is it like to think that P?”, *Philosophy and Phenomenological Research*, 69(1): 1–36.
- Place, U.T., 1956, “Is Consciousness a Brain Process?” *British Journal of Psychology*, 47: 44–50.
- Polger, T., 2011, “Are Sensations Still Brain Processes?”, *Philosophical Psychology*, 24(1): 1–21.
- Polger, T. and Shapiro, L., 2016, *The Multiple Realization Book*, New York: Oxford University Press.
- Prior, E., and R. Pargetter and F. Jackson, 1982, “Three Theses About Dispositions”, *American Philosophical Quarterly*, 19(3): 251–257.
- Putnam, H., 1960, “Minds and Machines”, reprinted in Putnam 1975b, 362–385.
- , 1963, “Brains and Behavior”, reprinted in Putnam 1975b, 325–341.
- , 1967, “The Nature of Mental States”, reprinted in Putnam 1975b, 429–440.
- , 1973, “Philosophy and our Mental Life”, reprinted in Putnam 1975b, 291–303.
- , 1975a, “The Meaning of ‘Meaning’”, reprinted in Putnam 1975b, 215–271.
- , 1975b, *Mind, Language, and Reality*, Cambridge: Cambridge University Press.
- , 1988, *Representation and Reality*, Cambridge, MA: MIT Press.
- Quine, W.V., 1953, “Two Dogmas of Empiricism”, in W. V. Quine, *From a Logical Point of View*, New York: Harper and Row.
- Rey, G., 1980, “Functionalism and the Emotions”, in A. Rorty (ed.), *Explaining Emotions*, Berkeley, CA: University of California Press.
- , 1997, *Contemporary Philosophy of Mind*, Cambridge, MA: Blackwell.
- , 2007, “Resisting Normativism in Psychology”, in McLaughlin and Cohen 2007, 69–84.
- Rupert, R., 2006, “Functionalism, Mental Causation, and the Problem of Metaphysically Necessitated Effects”, *Noûs*, 40: 256–283.
- , 2009, *Cognitive Systems and the Extended Mind*, New York: Oxford University Press.
- Ryle, G., 1949, *The Concept of Mind*, London: Hucheson.
- Schaffer, J., 2003, “Overdetermining Causes”, *Philosophical Studies*, 114: 23–45.





- , 2021, “Ground Functionalism”. In Uriah Kriegel (ed). *Oxford Studies in Philosophy of Mind, Vol. 1*. New York: Oxford University Press, pp. 171–207.
- Schroer, R., 2010, “Where’s the Beef?: Phenomenal Concepts as both Demonstrative and Substantial”, *Australasian Journal of Philosophy*, 88(3): 505–522.
- Schwitzgebel, E., 2010, “Acting Contrary to our Professed Beliefs, or the Gulf Between Occurrent Judgment and Dispositional Belief”, *Pacific Philosophical Quarterly*, 91: 531–53.
- , 2012, “Mad Belief?”, *Neuroethics*, 5: 13–17.
- , 2015, “If Materialism is True, the United States is Probably Conscious”, *Philosophical Studies*, 172: 1697–1721.
- Searle, J., 1980, “Minds, Brains and Programs”, *Behavioral and Brain Sciences*, 3: 417–457
- , 1992, *The Rediscovery of Mind*, Cambridge, MA: MIT Press.
- Sellars, W., 1956, “Empiricism and the Philosophy of Mind”, in M. Scriven, P. Feyerabend and G. Maxwell (eds.), *Minnesota Studies in the Philosophy of Science*, Volume 1. Minneapolis: U.of Minnesota Press.
- Shagrir, O., 2005, “The Rise and Fall of Computational Functionalism”, in Y. Ben-Menahem (ed.), *Hilary Putnam*, Cambridge: Cambridge University Press, 220–250
- Shapiro, L. and T. Polger, 2012, “Identity, variability, and multiple realization in the special sciences”, in Gozzano and Hill 2002, 264–287.
- Shields, C., 1990, “The first functionalist”, in J.-C. Smith (ed.), *Historical Foundations of Cognitive Science*, Dordrecht: Kluwer: 19–33.
- Shoemaker, S., 1984, *Identity, Cause, and Mind*, Cambridge: Cambridge University Press.
- , 1984a, “Phenomenal similarity”, in Shoemaker 1984, 159–183.
- , 1984b, “Functionalism and qualia”, in Shoemaker 1984, 184–205.

- , 1984c, “Some varieties of functionalism”, in Shoemaker 1984, 261–286.
- , 1996, *The first-person perspective and other essays*, Cambridge: Cambridge University Press.
- , 1996a, “Introspection and the self”, in Shoemaker 1996, 3–24.
- , 1996b, “On knowing one’s own mind”, in Shoemaker 1996, 25–49.
- , 1996c, “First-person access”, in Shoemaker 1996, 50–73.
- , 1996d, “Self-knowledge and ‘inner sense’: Lecture I”, in Shoemaker 1996, 201–223.
- , 2001, “Realization and Mental Causation”, in C. Gillett and B. Loewer, *Physicalism and Its Discontents*, Cambridge: Cambridge University Press, 74–98.
- Sider, T., 2003, “What’s so Bad About Overdetermination?”, *Philosophy and Phenomenological Research*, 67: 716–726.
- Smart, J.J.C., 1959, “Sensations and Brain Processes”, *Philosophical Review*, 68: 141–156.
- Smithies, D., 2013a, “The Significance of Cognitive Phenomenology”, *Philosophy Compass*, 8: 731–743.
- , 2013b, “The Nature of Cognitive Phenomenology”, *Philosophy Compass*, 8(8): 744–754.
- Soames, S., 2004, *Reference and Description*, Princeton, NJ: Princeton University Press.
- Sprevak, M., 2009, “Extended cognition and functionalism”, *Journal of Philosophy*, 106(9): 503–527.
- Staffel, J., 2013, “Can there be reasoning with degrees of belief?”, *Synthese*, 190(16): 3535–3551.
- , 2019, *Unsettled Thoughts: A Theory of Degrees of Rationality*. New York: Oxford University Press.
- Stalnaker, R., 2002, “What is it Like to Be a Zombie?”, in Gendler and Hawthorne 2002, 385–400.

- Sterelny, K., 1993, “Categories, categorisation and development: Introspective knowledge is no threat to functionalism”, in *Behavioral and Brain Sciences*, 16(1): 81–83.
- Stich, S., 1981, “Dennett on Intentional Systems”, *Philosophical Topics*, 12: 39–62.
- , 1983, *From Folk Psychology to Cognitive Science*, Cambridge, MA: MIT Press.
- Strawson, G., 1994, *Mental Reality*, Cambridge, MA: MIT Press.
- Stoljar, D., 2001, “Two Conceptions of the Physical”, *Philosophy and Phenomenological Research*, 62: 253–81.
- Strevens, M., 2012, “The Explanatory Role of Irreducible Properties”, *Noûs*, 46(4): 754–780.
- Thompson, E., 2015, *Waking, Dreaming, Being*, New York: Columbia University Press.
- Tooley, M., 2001, “Functional concepts, referentially opaque contexts, causal relations, and the definition of theoretical terms”, *Philosophical Studies*, 105(3): 251–279.
- Tye, M., 2000, *Consciousness, Color, and Content*, Cambridge, MA: MIT Press.
- Van Gulick, R., 1989, “What Difference Does Consciousness Make?” *Philosophical Topics*, 17: 211–230.
- Velmans, M. and Schneider, S. (eds.), 2007, *The Blackwell Companion to Consciousness*, Maldon, MA: Blackwell Publishing.
- Walter, S., 2010, “Cognitive Extension: The Parity Argument, Functionalism, and the Mark of the Cognitive”, *Synthese*, 177(2): 285–300.
- Wedgwood, R., 2007, “Normativism Defended”, in McLaughlin and Cohen 2007, 85–101.
- White, S., 1986, “Curse of the Qualia”, *Synthese*, 68: 333–368.
- White, S., 2007, “Property Dualism, Phenomenal Concepts, and the Semantic Premise”, in Alter and Walter 2007, 210–248.

- Witmer, G., 2003, “Functionalism and Causal Exclusion”, *Pacific Philosophical Quarterly*, 84(2): 198–214.
- Wittgenstein, L., 1953, *Philosophical Investigations*, New York: Macmillan.
- Yablo, S., 1992, “Mental Causation”, *Philosophical Review*, 101: 245–280.
- , 2000, “Textbook Kripkeanism and the Open Texture of Concepts”, *Pacific Philosophical Quarterly*, 81(1): 98–122.
- , 2002, “Coulda, Shoulds, Woulda”, in Gendler and Hawthorne 2002, 441–492.
- Yates, D., 2012, “Functionalism and the Metaphysics of Causal Exclusion”, *Philosophers Imprint*, 12: 1–25.
- Yi-Vakkuri, J. and Hawthorne, J., 2018, *Narrow Content*, Oxford: Oxford University Press.

Academic Tools

-  How to cite this entry.
-  Preview the PDF version of this entry at the Friends of the SEP Society.
-  Look up topics and thinkers related to this entry at the Internet Philosophy Ontology Project (InPhO).
-  Enhanced bibliography for this entry at PhilPapers, with links to its database.

Other Internet Resources

[Please contact the author with suggestions.]

Related Entries

behaviorism | belief | dualism | mental causation | mental representation | mind/brain identity theory | mind: computational theory of | multiple realizability | physicalism | qualia | qualia: knowledge argument | Turing machines

Copyright © 2023 by the author

Janet Levin