

7-31-2018

Memory and Disjunctivism

Arieh Schwartz

University of California, Davis, agschwartz@ucdavis.edu

Follow this and additional works at: <https://commons.pacificu.edu/eip>

Recommended Citation

Schwartz, Arieh () "Memory and Disjunctivism," *Essays in Philosophy*: Vol. 19: Iss. 2, Article 4. <https://doi.org/10.7710/1526-0569.1611>

Essays in Philosophy

ISSN 1526-0569 | Essays in Philosophy is published by the Pacific University Libraries

Volume 19, Issue 2 (2018)

Memory and Disjunctivism

Arieh Schwartz

University of California, Davis

Abstract

Recent analyses of memory (Robins 2016; Cheng & Werning 2016; Michaelian 2016; Bernecker, 2017) propose necessary and jointly sufficient conditions for a mental state to be a memory, which are meant to set memory apart from related mental states like illusory memory and confabulation. Each of the proposed taxonomies includes accuracy as one of the necessary conditions such that only accurate representations are memories. I argue that inclusion of an accuracy condition implies a sort of disjunctivism about seeming to remember. The paper distinguishes several types of disjunctivism that these taxonomies could be committed to. If these taxonomies are meant to be empirically informed, however, then plausibly they should be seen to endorse the principle of psychological internalism. The causal argument, a standard objection to disjunctivism (Robinson 1985; Burge 2005, 2011), is then used to show that the sort of disjunctivism that endorses psychological internalism is mistaken. The ultimate goal is to underscore a lack of clarity in the status of recent accounts of memory as either epistemic, nonreductively ontological, or reductively ontological in approach.

Essays Philos (2018)19:2 | DOI: 10.7710/1526-0569.1611

Correspondence: agschwartz@ucdavis.edu



© 2018 Schwartz. This open access article is distributed under a Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>)

0. Introduction

A current research project in the philosophy of memory, explicitly initiated by Sarah K. Robins but implicit in earlier work (going back to at least C. B. Martin and Max Deutscher),¹ is to give an account of what memory is that can distinguish instances of remembering from a gamut of related mental states including misremembering, veridical and falsidical confabulation, and veridical and falsidical relearning. Research in this program² admirably takes account of memory science, and attempts to provide an analysis of memory that is not only compatible with this work, but that may be empirically fruitful; that may “improve diagnosis” of memory errors,³ for example, or even get at “how episodic memory should be characterized in order to be validated as a natural kind.”⁴

This line of work is aimed at identifying all of the conditions in virtue of which a mental state is a memory. Moreover, all of the theorists above insist that representations are memories only if they are veridical and factive, in other words, *accurate* representations of the target event. (Of course, other necessary conditions on memory are also invoked.) I will argue that if we are interested in giving an account of what memory is that is compatible with the assumption of psychological internalism (PI)—which is relatively orthodox across the cognitive sciences—that mental states are internal states of the subject, then building an accuracy constraint into our account lets too much epistemology into our ontology. What a given mental state is and what epistemic properties it has are, or ought to be, separate questions that should be given separate answers.

The goal of this paper is to motivate two claims. 1) The accuracy condition on memory implies some form of disjunctivism about seeming to remember. Namely, if only accurate representations are memories, then although an instance of representing a past event accurately and an instance of representing it inaccurately may be subjectively indistinguishable, and although they may be subtended by the same neuropsychological

¹ Sarah K. Robins, “Misremembering,” *Philosophical Psychology* 29, no. 3 (2016) 432–47; C. B. Martin and Max Deutscher, “Remembering,” *Philosophical Review* 75, no. 2 (1966): 161–96.

² See, for example, Sen Cheng and Markus Werning, “What Is Episodic Memory If It Is a Natural Kind?” *Synthese* 193, no. 5 (2016): 1345–85; Robins, “Misremembering”; Kourken Michaelian, “Confabulating, Misremembering, Relearning: The Simulation Theory of Memory and Unsuccessful Remembering,” *Frontiers in Psychology* 7, no. 1857 (2016): 1–13; and Sven Bernecker, “A Causal Theory of Mnemonic Confabulation,” *Frontiers in Psychology* 8, no. 1207 (2017): 1–14.

³ Bernecker, “A Causal Theory,” 11.

⁴ Cheng and Werning, “What Is Episodic Memory,” 1345.

mechanisms, they amount to a disjunction of different kinds of mental states. The accurate representation is a memory (providing other necessary conditions are met), while the inaccurate representation is a mental state that is fundamentally different in kind. 2) If we want to provide an ontology of mind that is compatible with PI, then we should not be disjunctivists about memory, in the sense that would be implied by the accuracy condition. Hence, if we want an ontology of mind that is compatible with PI, we are led to reject the accuracy condition. Disjunctivism about memory is problematic, I shall argue, to the extent that it takes the putative disjunction as one of internal mental state kinds.

The plan for this paper is as follows. The aim of section 1 is to discuss how the accuracy condition functions in recent taxonomies of memory, and to show how it leads to the disjunctive view about seeming to remember. Section 2 distinguishes several different varieties of disjunctivism from the problematic variety that is the target of this paper; namely, reductive ontological disjunctivism. Section 3 aims to clarify why a taxonomy that is compatible with PI should not build epistemic conditions into what the mental state of memory is. This discussion draws on an old line of thought that is sometimes called the causal argument against disjunctivism.⁵ Section 4 concludes.

1.1 Remembering, Misremembering, and Confabulating

Robins posed the challenge that recent work has taken up to distinguish remembering from “misremembering” from confabulation.⁶ “Misremembering,” for Robins, is a technical term that refers to the kind of memory error that is characteristic of subjects in Deese-Roediger-McDermott (DRM) experiments, for example. The DRM paradigm was resuscitated by Henry L. Roediger and Kathleen B. McDermott (1995) from a study by Deese (1959).⁷ Subjects are presented with a series of thematically related items during the study phase, traditionally: a list of semantically related words, that are focused around a “critical lure.” Here is an example study list that is focused around “sweet.” Sour,

⁵ See Howard Robinson, “The General Form of the Argument for Berkeleian Idealism,” in *Philosophy of Mind: Contemporary Readings*, ed. Timothy O'Connor and David Robb (New York: Routledge, 2003), 81–102; and *Perception* (New York: Routledge, 1994). See also Tyler Burge, “Disjunctivism and Perceptual Psychology,” *Philosophical Topics* 33, no. 1(2005): 1–78; and “Disjunctivism Again,” *Philosophical Explorations* 14, no. 1 (2011): 43–80.

⁶ Robins, “Misremembering.”

⁷ Henry L. Roediger and Kathleen B. McDermott, “Creating False Memories: Remembering Words Not Presented in Lists,” *Journal of Experimental Psychology* 21, no. 4 (1995): 803–14; Deese; James Deese, “On the prediction of occurrence of particular verbal intrusions in immediate recall,” *Journal of Experimental Psychology*, 58, no. 1 (1959): 17–22.

candy, sugar, bitter, good, taste, tooth, nice, honey, soda, chocolate, heart, cake, tart, pie (Roediger and McDermott, 1995: 814).⁸ During the test phase, which may ask the participant either to “recall,” i.e., freely produce studied items, or simply to “recognize” which ones had been presented, participants typically purport to remember the critical lures at rates comparable to those subjects exhibit for items that had actually been on the study lists. This is a distinctive sort of error, according to Robins, because making this error—answering that a critical lure had been on a study list when it hadn’t—requires that the subject has some memory of the items that actually *had* been presented. Otherwise, the semantic relatedness of the lure to the targets should have no effect. Yet the false-alarm rate for thematically unrelated lures is usually close to zero, and this is good evidence that the thematic relation between critical lures and targets is very salient.

Although the DRM effect is a particularly clear example of “misremembering”—namely, accurately representing a past event in some aspects, while inaccurately representing it in others—this kind of memory error is very much an everyday phenomenon. That misremembering involves some retention of information from the encoding condition indicates that it is a different kind of memory error from confabulation, which seems not to involve retained information from the target episode at all. Robins argues that it is not possible for either archivalists or simple causal trace theorists to explain why it is that subjects “misremember” critical lures in the first place. For the archivalist, memory functions to store past events and experiences. Similarly, for the causal trace theorist, experience is operative in generating a causal trace of the encoding experience that is eventually operative in generating the retrieved representation through a continuous causal chain of processes. Robins’s charge is that it is hard to square, on this sort of view, the fact that subjects appear to succeed in maintaining a trace of the encoding condition, which is necessary for the thematic relatedness of the critical lure to be operative, and yet fail to accurately remember the word lists at retrieval.⁹ The trace, allegedly an archive of the past experience, is what was supposed to be retrieved.

On the constructivist view, by contrast, there are no particular traces corresponding to particular experiences. On this sort of view, the memory system constructs plausible simulations of past events, in the same way that it constructs plausible models of future

⁸ Roediger and McDermott, “Creating False Memories.”

⁹ Allegedly, the causal theorist cannot explain this error as (1) **guessing** (since participants’ performance is good instead of near chance); (2) **procedural error** or a problem in the search process (since, “if the search did not locate the list, then how could the list influence participant responses?” [437]); (3) **trace decay**, (since misremembering is an error of commission and not an error of omission); or (4) **noise** (since the error is highly systematic).

events (episodic future thinking, EFT), and plausible models of how the past could have been (episodic counterfactual thinking).¹⁰ The problem for constructivism, according to Robins, is that it cannot adequately distinguish what makes it the case that remembering is distinct from misremembering is distinct from confabulation.

In the Constructivist's hands, all representations of past events—whether completely accurate, wholly contrived, or somewhere in between—are, in some sense, confabulations. The view offers no means for differentiation between constructions that are constrained by information retained from a particular past event and those that are not.¹¹

Moreover, the view especially cannot distinguish misremembering from confabulation, for the fact that both errors are simply seen to be inaccurate simulations.

In response, Robins proposes a hybrid theory that takes an archival view of storage and a constructive view of retrieval. Remembering is a matter of “retention of information from a particular past event and construction of an accurate representation of that event at the time of retrieval.”¹² On this proposal, subjects *remember* if and only if both conditions are met. They *misremember* if and only if the retention condition is met but the accuracy condition is not. And they *confabulate* if and only if neither condition is met.

1.2 The Accuracy Conditions

As in the account proposed by Robins, accuracy is established as a necessary condition on memory in the analyses proposed by Cheng and Werning, Michaelian, and Bernecker.¹³ These theories are distinguished from that of Robins in maintaining that it is unnecessary to opt for a hybrid view that is part causalist and part constructivist, because one of the pure views can do all the relevant explanatory work. What is important for our purposes is not the precise nature of the additional conditions that theorists place on the memory process, but rather the accuracy condition they place on its products. Nevertheless, it is good to provide a brief characterization of their views.

Cheng and Werning, and Bernecker take a pure causalist position, where what is nec-

¹⁰ Felipe De Brigard, “Is Memory for Remembering? Recollection as a Form of Episodic Hypothetical Thinking,” *Synthese* 191 (2014): 1–31.

¹¹ Robins, “Misremembering,” 444.

¹² *Ibid.*, 445.

¹³ Cheng and Werning, “What Is Episodic Memory”; Michaelian, “Confabulating”; Bernecker, “A Causal Theory.”

essary for a state to count as a memory is that it is appropriately causally based on an accurate experience, and therefore reflects the experience accurately. Bernecker, for example, writes that “memories must be veridical in the sense of being factually correct . . . a mental state qualifies as a memory only if it accurately represents the objective reality and accords with the subject’s initial perception of reality.”¹⁴ This is supposed to occur because the accurate experience leaves a causal trace, an impression upon the mind, that is eventually operative in generating the retrieved memory by a continuous causal process. Here, the causal process must be such that the content of the retrieved representation is counterfactually dependent on the content of the original experience in that had the latter differed, the former would also.

The account proposed by Cheng and Werning also ensures the sort of *accuracy* (to the facts) and *authenticity* (to the past experience) that Bernecker insists upon by stipulating that the encoding experience must be accurate, and must “ground” the content of the retrieved representation. Namely, they require not only that the subject “has a reliable experience of the episode E* . . . called the experiential base,” but also that the content of the retrieved memory representation is “ontologically grounded” by the content of the experiential base, in that the content of the former is either *identical* to that of the latter, a *constituent* of that of the latter, or an *abstraction* of that of the latter.¹⁵ It is plausible that the ontological grounding condition implies the sort of counterfactual dependence that Bernecker has in mind. Moreover, this is a causalist view, like Bernecker’s, because it requires that the memory representation is caused by the target experience, if indirectly, by a “reliable trace” left by the target experience.

The views of Bernecker and Michaelian are distinguished from that of Robins in the further sense that they explicitly allow for the possibility of accurate or veridical confabulation. For Bernecker, a confabulation is either accurate or inaccurate, but necessarily, it is not counterfactually dependent in its content on that of the encoding experience because it is not related to the encoding experience by the relevant sort of causal trace and attendant processes. On Michaelian’s proposal, there is no retention or causal trace condition on memory, but rather a reliability condition on the simulation process by which memories must be constructed. Michaelian’s necessary and jointly sufficient conditions on memory are accuracy, internality, and reliability. Memories must be accurate, and they must be produced by an internal process that is reliable in the sense that it tends to produce accurate representations in close possible worlds. The internality condition is meant to distinguish remembering, misremembering, and confabulating from veridical

¹⁴ Bernecker, “A Causal Theory,” 4.

¹⁵ Cheng and Werning, “What Is Episodic Memory,” 1354–55.

and falsidical relearning, which depend on a process that occurs partially external to the subject.¹⁶ Thus, internality is meant to provide the means for distinguishing a memory of a childhood trip to the seaside, for example, from a representation of the experience that is wholly based on external and mediate sources (family stories, photographs, etc.)

Robins and Michaelian state explicitly that they understand the trio of remembering, misremembering, and confabulating to parallel the categories of perception, illusory perception, and hallucination, respectively.¹⁷ Robins, for example, writes that

the distinction between misremembering and confabulation runs analogously to that between illusion and hallucination in the study of perception. In illusory perception, one perceives an object as having properties that it does not. Hallucination is a more extensive error; it occurs when the entire perceptual experience—both what is perceived and how it is perceived is illusive.¹⁸

Perhaps tellingly, there is no standard nondeverbal noun phrase that refers to the state of misremembering. For convenience, let's call it "illusory memory." In the theories we have discussed, memory is distinguished from illusory memory in that only accurate representations are memories, and illusory memories are inaccurate. *Retention*, for Robins, *causal trace conditions* for Bernecker, and Cheng and Werning, and *reliability*, for Michaelian, are supposed to distinguish veridical confabulations from memories, on one hand, and falsidical confabulations from illusory memories, on the other.¹⁹ Together, the conditions entail that seeming to have a memory of an episode is a disjunctive state of affairs: some cases amount to memory, some amount to illusory memory, some amount to confabulation. The same kind of phenomenology is not a signal of the same kind of

¹⁶ Relearning is generally thought to consist in (1) an encoding experience, (2) imparting information about this experience to an external source, (3) forgetting this information, (4) reacquiring it, and then (5) representing the target event based on the reacquired information. Relearning is said to be "veridical," if the information represented is accurate, and "falsidical" otherwise.

¹⁷ See Robins, "Misremembering," 434; and Michaelian, "Confabulating," 2.

¹⁸ Robins, "Misremembering," 434.

¹⁹ Cheng and Werning do not explicitly discuss confabulation. Although they say that "a mnemonic representation of that episode will also fail to be a case of episodic memory if it is based on the imagination or confabulation of that encounter," it appears that they intend "confabulation" more in the sense of hallucination than in the sense of memorial confabulation ("What Is Episodic Memory," 1360). Nevertheless, because the "sequence analysis" is intended to provide necessary and jointly sufficient conditions for episodic memory, it must be capable of determining that confabulations are not episodic memories, and we can assume that this job would fall to (S6) "S's representation with content *E* at *t*₁ is causally grounded in S's experience of *E** through a reliable memory trace" (1354).

mental state. As Cheng and Werning explain, “a mnemonic representation of an episode, say an encounter with a black panther, may fail to be a case of episodic memory because the representation is based on a misperception, illusion or even a hallucination.”²⁰ This view—which is common among the theories in question—that memories are different in kind from illusory memories and confabulations, is somewhat analogous to how disjunctive theories of perception have been articulated since J. M. Hinton.²¹ Here, seeming to remember is really a disjunction of different kinds of mental states, as the disjunctivist takes seeming to perceive as a disjunction of different kinds of mental states.

Disjunctive views of perception have been defined in different ways. Versions of disjunctivism say there is no “common mental core,” or “highest common factor,” or “most specific common kind” of mental state that is shared by the subject in a “good case” (who perceives something veridically) and the subject in a “bad case” (whose experience is introspectively indistinguishable from a good case, but nonveridical in some respects).²² Saying the mental state of memory requires accuracy establishes a disjunction between memory, on the one hand, and illusory memory and (falsidical) confabulation, on the other. The mental state of the subject in the good memory case is different in kind than the mental state of the subject in the bad memory case. The good case and the bad case are understood to lack a common, most specific answer to the question of what kind of mental state is in play. There are, however, various sorts of memory disjunctivism to which the theorists in question could be committed, only some of which are virulent.

2. Varieties of Memory Disjunctivism

Disjunctivism, I shall argue, is problematic to the extent that it denies that—with respect to everything that is going on *inside* the subject—memory, illusory memory, and confabulation can sometimes comprise all and only the same mental kinds. Some types of disjunctivism are not bound to deny this possibility. For example, epistemological

²⁰ Bernecker (2017) appears to make the same point: “On this view, one can fail to remember something not only because there is something wrong with one’s memory but also because the representation fed into the memory process is false. Memory neither allows for a mistake of inheritance nor for the inheritance of a mistake” (4). My own view is that memory, considered as an ontological kind, should allow for the inheritance of a mistake. The inherited mistake need have nothing to do with memory.

²¹ J. M. Hinton, “Visual Experiences,” *Mind* 76, no. 302 (1967): 217–27.

²² Timothy Williamson is the source of the cases terminology (*Knowledge and Its Limits* [Oxford: Oxford University Press, 2000]). A case is a centered possible world, i.e., a world with a marked subject and time. For an overview of the different kinds of disjunctivism, see Alex Byrne and Heather Logue, “Either/Or,” in *Disjunctivism: Perception, Action, Knowledge*, ed. Adrian Haddock and Fiona Macpherson (Oxford: Oxford University Press, 2011), 57–94.

disjunctivism (ED) claims that seeming to have a memory is a disjunction of epistemic states that are different in kind. This type of disjunctivism is basically anodyne. Memory, illusory memory, and confabulation are in fact epistemically different kinds of states. They are evidentially different in kind, for example. This view leaves it ambiguous, however, whether instances of these different epistemic states correspond to different internal states of the subject. Therefore, ED is not at odds with the argument of this paper.

A different form of disjunctivism, which I shall call nonreductive ontological disjunctivism (NOD), may also be consistent with the argument of this paper. One way to be a nonreductive ontological disjunctivist is to hold that mental states are irreducible *propositional attitudes*, in the sense of two-place relations that subjects stand in to propositions that are the objects of their thoughts.²³ A proposition might be understood as a structured (i.e., ordered) complex of the referents of subsentential expressions to which the subject is related under a certain guise (Russellian). Alternately, a proposition might be understood as a structured complex of the meanings of subsentential expressions, which has a certain cognitive significance (Fregean). The metaphysics of propositions need not detain us as long as we stipulate that the relation the subject bears to the proposition in question, the content of the subject's thought, is not wholly determined by the nature of the mental representations involved. Adopting this sort of view, a theorist can maintain that "subjects are in mental states, not vice versa"²⁴ and thus that memory, e.g., is not some internal state of the subject, or, obviously, of the subject's brain. Hence, that memory, illusory memory, and confabulation are considered specifically different mental states is consistent with the possibility that all three mental states may be accompanied by all and only the same kinds of internal states. However, it is important to note that NOD is in tension with the common assumption of psychological internalism (PI), according to which mental states are (or at least supervene exclusively on) the internal states of the subject.²⁵ (A fuller discussion of PI is given later in this section.) Thus, insofar as the cog-

²³ For example, my believing 'It is 5pm,' and your believing 'It is 5pm,' amount to our occupying the same state of mind in assuming the same mental attitude, i.e. belief, toward the same proposition, 'It is 5pm.' But when I merely desire that 'It is 5pm,' I occupy a different mental state, because I take a very different attitude toward this propositional content.

²⁴ Williamson, "Replies to Critics," In *Williamson on Knowledge*, ed. Patrick Greenough and Duncan Pritchard (Oxford: Oxford University Press, 2009), 331.

²⁵ Functional decomposition, which is plausibly the principal methodology of cognitive science (see, e.g. Robert Cummins, "'How Does It Work?' versus 'What Are the Laws?'" In *Explanation and Cognition*, ed. Frank C. Keil and Robert A. Wilson (Cambridge, MA: MIT Press, 2000), seems to evidence PI, in that personal-level capacities are analyzed into component functions of internal subpersonal capacities (see, e.g., Zoe Drayson, "The Personal/Subpersonal Distinction," *Philosophy Compass* 9, no. 5 [2014]: 338–46).

nitive sciences endorse PI, it is unclear whether and how a proponent of NOD could provide a taxonomy of memory that is empirically informed, let alone empirically fruitful.

There is a third kind of disjunctive view of memory, however, which we may call reductive ontological disjunctivism (ROD), and which, I shall argue, we should repudiate. On this sort of view, psychological states are thought to reduce to internal states of the subject.²⁶ Nevertheless, it is thought that the *internal* mental states of subjects must be different when they remember / misremember / confabulate. This is the sort of view that is suggested when Cheng and Werning adduce necessary conditions that are meant to validate *episodic memory* as a natural kind. These authors are quite aware (in fact, Cheng is a neuroscientist of memory) that the term *episodic memory* refers to a neuropsychological kind.²⁷ Hence, claiming that episodic memory is a necessarily accurate *natural kind* (which is a result of the conditions discussed above,) is to say that memory is a different kind of neuropsychological state than those of illusory memory and confabulation. This is an empirical claim, however, and it requires empirical evidence that has not been provided. The problem with this claim is actually worse than that it is just empiri-

²⁶ The reduction referenced here need not be a reduction of psychological states to neuropsychological states to biochemical states and so on; namely, it need not be an intertheoretic scientific reduction of psychology to neuroscience and so forth. Rather, it is merely the assumption that the terms of folk psychology will turn out to be either reducible to items in *some* science of the mind or eliminable.

²⁷ *Episodic memory* was coined by Endel Tulving to name the sort of memory for experienced events that is rich in contextual details (e.g., regarding the what, the where, and the when of the relevant episode's occurrence) (Tulving, "Episodic and Semantic Memory," in *Organization of Memory*, ed. Endel Tulving and Wayne Donaldson [New York: Academic Press, 1972], 38–403). Episodic memory is contrasted with semantic memory for facts that lacks substantial contextual detail of when the information was acquired. Episodic and semantic memory are the two forms of what is called declarative memory, memory whose content can be consciously accessed and manipulated across a range of retrieval conditions (see Larry R. Squire and Stuart Zola-Morgan, "Memory and Brain Systems: 1969–2009." *Trends in Neuroscience* 11, no. 4 [1988]: 170–75). It remains controversial how to precisely characterize episodic memory. Tulving's more recent account (see *Elements of Episodic Memory* [Oxford: Oxford University Press, 1983], and "Memory and Consciousness," *Canadian Psychology* 26, no. 1 [1985]: 1–12), which defines episodic memory as the sort of memory that comes with the conscious phenomenology of reexperiencing a past event, is a further approach. It is widely known, (and Cheng and Werning particularly stress this point ["What Is Episodic Memory," 1364–76]) that episodic memories are primarily dependent (at least before consolidation), upon the hippocampus, while semantic memories are known to depend primarily on neocortical regions. Although these memory systems are to some extent dissociable, the extent to which they work interdependently is an interesting area of ongoing research (see, e.g., Muireann Irish and Olivier Piguet, "The Pivotal Role of Semantic Memory in Remembering the Past and Imagining the Future," *Frontiers in Behavioral Neuroscience* 7, no. 27 [2013]: 1–11; and Daniel Greenberg and Mieke Verfaellie, "Interdependence of Episodic and Semantic Memory: Evidence from Neuropsychology," *Journal of the International Neuropsychology Society* 16, no. 5 [2010]: 748–53).

cally unsubstantiated, however (the next section explains how).

One reason to suspect that ROD is the sort of disjunctivism to which Robins, Michaelian, and Bernecker are also committed is that these theorists engage quite closely with the empirical science of memory; they intend their work to be useful in clinical settings, or at the very least, compatible with a scientific view of what mental states are. Bernecker, for example, says his “paper is to be understood as a friendly offer of help to the medical sciences.”²⁸ Michaelian describes his project as “drawing on the resources of philosophy but aiming for coherence with psychology.”²⁹ And Robins based the taxonomic program upon the need to explain the “DRM effect, one of the most well-established effects in the psychological study of memory.”³⁰ Yet psychological internalism appears to be a fairly common commitment of cognitive science. Therefore, if Robins, Michaelian, and Bernecker are on board with the basic assumption that mental states are internal states of subjects, and they are committed to the claim that memory, illusory memory, and confabulation are different kinds of mental states—not just different epistemic profiles that the same kind of mental state can assume—then they appear committed to the claim that these are different internal states of subjects.

Stephen P. Stich’s maxim, “what knowledge adds to belief is psychologically irrelevant,” gets at the core of psychological internalism.³¹ Psychology, with the cognitive sciences generally, aims to account for mental states in whatever way is most explanatory of behavior. Yet whether a belief is accurate does not make a difference to its causal profile; it will cause the same behaviors regardless of its accuracy. Hence, what knowledge adds to belief, what perception adds to illusion, what remembering adds to misremembering, is not relevant to the psychological explanation of behavior. As Jaegwon Kim explains,

It is only the element of belief in knowing that is causally productive of the action. Similar comments apply to believing truly. My truly believing that something is so is not more efficacious in producing actions than my merely believing that something is so.

Seeming remembrance can do all of the explanatory work done by remembering. Thus, when I act in a certain way in part because of my remembering a certain thing, then under the same circumstances my replica will act in the same way

²⁸ Bernecker, “A Causal Theory,” 2.

²⁹ Michaelian, “Confabulating,” 1.

³⁰ Robins, “Misremembering,” 445.

³¹ Stephen P. Stich, “Autonomous Psychology and the Belief-Desire Thesis,” *Monist* 61, no. 4 (1978): 574.

because of his seeming to remember the same thing. Whether or not his seeming remembrance is a genuine case of remembering will not affect his behavior.³²

The moral is entirely general. Whether your mental state is accurate is *not* a feature of the physical nature of your internal states, while it is only these features that impact what behaviors are caused. Surely whether your mental state is accurate may impact what your behavior achieves in the world—whether saying “I do” starts your marriage, for example, or whether sliding your card pays for the opera—but “the job of psychological explanation is done once it has explained the bodily action[s].”³³ This is the justification many have seen for the common practice in cognitive science to classify mental states according to their internal features. Here are some other philosophers’ articulation of the principle:

No psychological state, properly so called, presupposes the existence of any individual other than the subject to whom that state is ascribed.³⁴

The properties and relations to be invoked in an explanatory psychological theory must be supervenient upon the current, internal physical properties and relations of organisms (i.e., just those properties that an organism shares with all of its replicas).³⁵

Internal psychological states are the only psychological states that psychological theory needs to invoke in explaining human behavior—the only states needed for psychology.³⁶

The idea of psychological internalism is simply that mental states, as understood in cognitive science, are internal states of subjects.³⁷ The difference between accurate and inaccurate mental representations need not imply differences in the internal states of subjects. Therefore, it is improper to establish accuracy as an individuation criterion for

³² Jaegwon Kim, “Psychophysical Supervenience,” *Philosophical Studies* 41 (1982): 63, 65.

³³ *Ibid.*, 64.

³⁴ Hilary Putnam, “The Meaning of Meaning,” in *Mind Language, and Reality* (Cambridge: Cambridge University Press, 1975), 220.

³⁵ Stich, “Autonomous Psychology,” 575.

³⁶ Kim, “Psychophysical Supervenience,” 59.

³⁷ Cf. Drayson (“Personal/Subpersonal Distinction”) for discussion of PI and a line of resistance Williamson, and Clark and Chalmers have offered to it.

mental states, given PI. In light of these points, the choices it appears that Robins, Michaelian, and Bernecker are faced with are either to endorse NOD, a position that takes a fundamentally different view of the nature of mental states than that which is orthodox in cognitive science; endorse ROD, which is implausible for reasons we shall explore shortly; or reject ontological disjunctivism about memory.

3. Rejecting ROD

What is often called the causal argument against disjunctivism was advanced by Robinson,³⁸ (although he traces this way of thinking back to Berkeley). More recently, the argument has been also adapted by Burge.³⁹ The first step of Robinson's version of the argument is the premise that there does not need to be anything internally, i.e., neurally, that distinguishes an instance of veridical seeing from an instance of hallucination. Namely, (1) it is possible, at least in principle, to artificially initiate the same neural mechanisms that are involved in a case of veridical perception, and thereby cause an instance of hallucination that is subjectively indistinguishable from the veridical perception. The second step of the argument is to motivate the claim that "it is necessary to give the same account of both hallucinating and perceptual experiences . . . if they have the same proximate—i.e. neural—cause."⁴⁰ The second step of the argument is supported, according to Robinson, by the general principle that "same proximate cause, same immediate effect."⁴¹ If we hold fixed the salient background conditions across the cases in question, and we are interested in the nature of the effect, i.e., the mental state, seen from a purely causal perspective, the principle appears plausible.

In Burge's version of the argument, you are asked to consider a case of veridically perceiving an object, which, as you blink, is replaced by a duplicate object, which is then removed at the same time that you experience a visual illusion—caused by a strange confluence of light—that is indistinguishable from the experience of the original object.⁴² The first step of Burge's argument is not intended to establish the absence of a difference in the internal proximal causes of a case of veridical perception and a case of hallucination, but rather the absence of an internal causal difference between the causes of a veridical perception and two cases of illusion. What is important here is that the sensory

³⁸ Robinson, "General Form."

³⁹ Burge, "Disjunctivism"; "Disjunctivism Again."

⁴⁰ Robinson, "General Form," 89.

⁴¹ *Ibid.*, 90.

⁴² Burge, "Disjunctivism," 25.

impingement on the retina, and across the entire body, is supposed to be qualitatively the same across these cases. Although Burge's case is no doubt contrived, the fact that different objects in the distal environment can cause qualitatively the same proximal stimulation upon the organism is a familiar symptom of the general underdetermination of our perceptual representations by the perceptual data. Burge does not rest the second stage of the argument on the truth of the general principle, "same proximal cause, same immediate effect," but rather by the commitment of vision science to this principle within the specific domain of perception.

He describes the proximity principle (PP) as follows:

Holding constant the antecedent psychological set of the perceiver, a given type of proximal stimulation (over the whole body), together with associated internal afferent and efferent input to the perceptual system, will produce a given type of perceptual state, assuming that there is no malfunctioning in the system and no interference with the system.⁴³

Thus, given the assumption that an illusory perception can, at least in principle, be caused by the same proximal stimulation (given the same residual state of the perceiver) as a veridical perception, the proximity principle implies that the resultant illusory perception amounts to the same type of perceptual state as the veridical perception. It appears that reductive ontological disjunctivism about perception is mistaken insofar as it denies this.

It is time to apply this way of thinking to the case of memory. One immediate obstacle is that experience does not stand to memory as a proximate cause stands to an immediate effect. Rather, numerous processes intervene between the experience encoded and the memory retrieved. Thus, the principle of Robinson and Burge that (other things equal) the same proximate cause leads to the same immediate effect does not itself imply that, given the same kind of perceptual state as input, we get the same kind of memory state as immediate output; memory is no immediate effect of perception.⁴⁴ Nevertheless, the familiar considerations of "causal contiguity and continuity" apply here as ever.⁴⁵ There is no way for a cognitive system, including memory systems, to be sensitive to differences in the distal causes of perceptions unless these differences affect the sensible proximal stimulation upon the organism. Otherwise, such differences in the distal facts are simply

⁴³ Burge, "Disjunctivism," 22.

⁴⁴ Thanks to an anonymous reviewer for this point.

⁴⁵ Kim, "Psychophysical Supervenience," 65.

not available for processing. As Burge says, “the effects of distal causes are entirely exhausted by their effects on proximal causes.”⁴⁶

If we take our lead from Robinson and Burge that accurate and inaccurate perceptual representations are sometimes the same kind of mental state because they arise from the same kind of proximal stimulation, then we can assume that these mental states would have to provide the same kind of input to memory. Therefore, assuming no further evidence is forthcoming that one of these input experiences had been inaccurate, it would be difficult to explain how the memory system could instantiate a different kind of internal state, because the memory of this inaccurate experience would itself be inaccurate in turn. But the claim that the internal memory system instantiates different kinds of mental states when it represents the past accurately versus inaccurately, and that it does this *of necessity*, is the central thesis of ROD about memory. To illustrate the problem, imagine the following: At some time, three physically identical subjects register qualitatively identical proximal stimuli, which cause each to have an experience *as of* a red tomato. Ever after, each subject takes herself to have seen a red tomato on the occasion in question, never having received any evidence to the contrary. But in fact, while one subject had truly seen a red tomato, the second had experienced an illusion of a green tomato’s being red, and the proximal stimulation on the third subject had miraculously been caused by a strange confluence of light. These experiences are input to memory, and at some later time, each subject retrieves a representation of the experience. The first subject’s memory representation is accurate, but the second and third subjects’ representations are inaccurate. Does it follow that these subjects’ internal mental states are different in kind?

It would be unaccountable for the subjects’ memory systems to somehow access the nature of the distal things in themselves that occasioned the encoding experiences, and to instantiate different kinds of internal states accordingly. If we are to suppose that the mind/brain can take on different internal states in response to differences in the distal environment that made no difference to the proximal stimulation upon the organism, then we are left with a mystery of how this could possibly be accomplished. Are we to suppose that the distal environment acts upon the mind at a distance, or that mind is sensitive to the distal facts through ESP? The difference that made one perception veridical and the other nonveridical was at no point available for processing, so it is unclear how it could cause a difference between the internal memory states.

This argument does not support the idea that inaccurate representations of the past are *never* a different kind of mental state than accurate representations of the past. This will

⁴⁶ Burge, “Disjunctivism,” 22.

likely occur when the representations in question are produced by different kinds of cognitive mechanisms. Nevertheless, if internal memory states differ in kind, it is not by virtue of whether their content is accurate, but rather because of a difference in their causal profiles. It is incorrect to see the internal memory system as capable of infallibly tracking the accuracy of its representations, and of taking on different kinds of internal states accordingly. It is possible for the same kinds of internal memory states to bear content that is accurate or inaccurate.

4. Conclusion

There are numerous ways to resist the causal argument against disjunctivism, but the most obvious of them abandon the reductive view of mental states that is part and parcel of PI, and which seems to be a common assumption in the cognitive sciences. The correct moral to take away is not that the recent line of work in the philosophy of memory that seeks to distinguish memory from illusory memory from confabulation is misdirected. Rather, the point is that the job that recent memory taxonomies are intended to perform is unclear. It is unclear whether they are, perhaps, solely intended as epistemological taxonomies of the circumstances in which mental states count as having certain epistemic properties; whether they are only committed to ED. Or perhaps they are ontological taxonomies of when a mental state counts as taking the memory relation to a mind-independent proposition, as opposed to some different propositional attitude; i.e., perhaps they are only committed to NOD. Last, perhaps they are *indeed* meant as taxonomies of internal mental state kinds, (as clearly seems to be Cheng and Werning's view). If this is how the recent memory taxonomies are intended—if they are committed to reductive ontological disjunctivism—then I claim it is a mistake for them to make accuracy a necessary condition on memory.

Acknowledgements

I am very grateful to Zoe Drayson, Bernard Molyneux, Nico Orlandi, Adam Sennet, and two anonymous reviewers for helpful feedback on an earlier draft of this paper.

References

- Bernecker, Sven. "A Causal Theory of Mnemonic Confabulation." *Frontiers in Psychology* 8, no. 1207 (2017): 1–14. <https://doi.org/10.3389/fpsyg.2017.01207>
- Burge, Tyler. "Disjunctivism and Perceptual Psychology." *Philosophical Topics* 33, no. 1(2005): 1–78. <https://doi.org/10.5840/philtopics20053311>

———. “Disjunctivism Again.” *Philosophical Explorations* 14, no. 1 (2011): 43–80. <https://doi.org/10.1080/13869795.2011.544400>

Byrne, Alex, and Heather Logue. “Either/Or.” In *Disjunctivism: Perception, Action, Knowledge*, edited by Adrian Haddock and Fiona Macpherson, 57–94. Oxford: Oxford University Press, 2011.

Cheng, Sen, and Markus Werning. “What Is Episodic Memory If It Is a Natural Kind?” *Synthese* 193, no. 5 (2016): 1345–85. <https://doi.org/10.1007/s11229-014-0628-6>

Cummins, Robert. “‘How Does It Work?’ versus ‘What Are the Laws?’” In *Explanation and Cognition*, edited by Frank C. Keil and Robert A. Wilson. Cambridge, MA: MIT Press, 2000.

De Brigard, Felipe. “Is Memory for Remembering? Recollection as a Form of Episodic Hypothetical Thinking.” *Synthese* 191 (2014): 1–31. <https://doi.org/10.1007/s11229-013-0247-7>

Deese, James. “On the prediction of occurrence of particular verbal intrusions in immediate recall.” *Journal of Experimental Psychology* 58, no. 1 (1959): 17–22. <https://doi.org/10.1037/h0046671>

Drayson, Zoe. “The Personal/Subpersonal Distinction.” *Philosophy Compass* 9, no. 5 (2014): 338–46. <https://doi.org/10.1111/phc3.12124>

———. (2017) “Extended Minds and Prime Mental Conditions: Probing the Parallels.” In *Extended Epistemology*, edited by J. Adam Carter, Andy Clark, Jesper Kallestrup, S. Orestis Palermos, and Duncan Pritchard, 147–61. Oxford: Oxford University Press.

Greenberg, Daniel, and Mieke Verfaellie. “Interdependence of Episodic and Semantic Memory: Evidence from Neuropsychology.” *Journal of the International Neuropsychology Society* 16, no. 5 (2010): 748–53. <https://doi.org/10.1017/S1355617710000676>

Hinton, J. M. “Visual Experiences.” *Mind* 76, no. 302 (1967): 217–27. <https://doi.org/10.1093/mind/LXXVI.302.217>

Irish, Muireann, and Olivier Piguet. “The Pivotal Role of Semantic Memory in Remembering the Past and Imagining the Future.” *Frontiers in Behavioral Neuroscience* 7, no. 27 (2013): 1–11. <https://doi.org/10.3389/fnbeh.2013.00027>

Kim, Jaegwon. “Psychophysical Supervenience.” *Philosophical Studies* 41 (1982): 51–70. <https://doi.org/10.1007/BF00353523>

Martin, C. B., and Max Deutscher. “Remembering.” *Philosophical Review* 75, no. 2 (1966): 161–96. <https://doi.org/10.2307/2183082>

Michaelian, Kourken. “Confabulating, Misremembering, Relearning: The Simulation Theory of Memory and Unsuccessful Remembering.” *Frontiers in Psychology* 7, no. 1857 (2016): 1–13.

———. “Generative Memory.” *Philosophical Psychology* 24, no. 3 (2011): 323–42. <https://doi.org/10.1080/09515089.2011.559623>

Putnam, Hilary. "The Meaning of Meaning." In *Mind, Language and Reality*, 215–71. Cambridge: Cambridge University Press, 1975. <https://doi.org/10.1017/CBO9780511625251.014>

Robinson, Howard. *Perception*. New York: Routledge, 1994.

———. "The General Form of the Argument for Berkeleian Idealism." In *Philosophy of Mind: Contemporary Readings*, edited by Timothy O'Connor and David Robb, 81–102. New York: Routledge, 2003.

Robins, Sarah K. "Misremembering." *Philosophical Psychology* 29, no. 3 (2016): 432–47. <https://doi.org/10.1080/09515089.2015.1113245>

Roediger, Henry L., and Kathleen B. McDermott. "Creating False Memories: Remembering Words Not Presented in Lists." *Journal of Experimental Psychology* 21, no. 4 (1995): 803–14. <https://doi.org/10.1037/0278-7393.21.4.803>

Squire, Larry R., and Stuart Zola-Morgan. "Memory and Brain Systems: 1969–2009." *Trends in Neuroscience* 11, no. 4 (1988): 170–75. [https://doi.org/10.1016/0166-2236\(88\)90144-0](https://doi.org/10.1016/0166-2236(88)90144-0)

Stich, Stephen P. "Autonomous Psychology and the Belief-Desire Thesis." *Monist* 61, no. 4 (1978): 573–91.

Tulving, Endel. *Elements of Episodic Memory*. Oxford: Oxford University Press, 1983.

———. "Episodic and Semantic Memory." In *Organization of Memory*, edited by Endel Tulving and Wayne Donaldson, 38–403. New York: Academic Press, 1972.

———. "Memory and Consciousness." *Canadian Psychology* 26, no. 1 (1985): 1–12. <https://doi.org/10.1037/h0080017>

Williamson, Timothy. *Knowledge and Its Limits*. Oxford: Oxford University Press, 2000.

———. "Replies to Critics." In *Williamson on Knowledge*, edited by Patrick Greenough and Duncan Pritchard, 331. Oxford: Oxford University Press, 2009.