Philos Stud (2018) 175:2353–2372
https://doi.org/10.1007/s11098-017-0962-x

Against dispositionalism: belief in cognitive science

Jake Quilty-Dunn[1] · **Eric Mandelbaum**[2]

Published online: 6 September 2017

Abstract Dispositionalism about belief has had a recent resurgence. In this paper we critically evaluate a popular dispositionalist program pursued by Eric Schwitzgebel. Then we present an alternative: a psychofunctional, representational theory of belief. This theory of belief has two main pillars: that beliefs are relations to structured mental representations, and that the relations are determined by the generalizations under which beliefs are acquired, stored, and changed. We end by describing some of the generalizations regarding belief acquisition, storage, and change.

Keywords Belief · Dispositions · Representation · Dissonance · Bias · Inference

Dispositionalism about belief has become fashionable once again. As such it deserves a close review. We aim to provide a comprehensive, critical one.

1 Introduction

Dispositionalism is the view that believing a proposition is nothing more than having a certain set of dispositions. This thesis was popular throughout the twentieth century and, in some form or another, has been endorsed by Ryle (1949), Sellars

✉ Jake Quilty-Dunn
quiltydunn@gmail.com

Eric Mandelbaum
eric.mandelbaum@gmail.com

[1] Faculty of Philosophy, University of Oxford, Radcliffe Humanities Building, ROQ, Woodstock Road, Oxford OX2 6GG, UK

[2] The Graduate Center and Baruch College, CUNY, VC 5271, One Bernard Baruch Way, New York, NY 10010, USA

(1956), Quine (1960), Lewis (1972), Davidson (1984), and Stalnaker (1984) among many others. More recently, Schwitzgebel has defended dispositionalism in an influential series of papers (2001, 2002, 2010, 2013). "To believe that P," according to Schwitzgebel, "is nothing more than to match to an appropriate degree and in appropriate respects the dispositional stereotype for believing that P" (2002, 253). He correctly characterizes dispositionalism as a *superficial* thesis, in that it sees belief not as a matter of whether there is some deep fact realized somewhere in the mind, but rather as a matter of surface phenomena such as behavior and phenomenology. Schwitzgebel's superficialism is instructive: seeing how it goes astray will allow us to critique a broader class of views about belief.

Schwitzgebel's variety of dispositionalism is more nuanced than a flat-footed behaviorism that analyzes all mental states in terms of outwardly observable behavior (2013, 87). His dispositionalism isn't intended to be naturalistic or reductive, so he has no need for behaviorist restrictions on qualia and the like.[1] His theory is therefore a *phenomenal* dispositional account. Each belief has a stereotype, which consists of dispositions both to act and to feel.[2] So, for example, one's belief that there is bourbon in the cabinet has a stereotype containing certain behavioral dispositions (e.g., the disposition to go to the cabinet if one wants to make a Manhattan) and also sometimes phenomenal dispositions (e.g., the disposition to feel disappointment when one opens the cabinet to find rye instead).

The foremost alternative to dispositionalism is *representationalism*, which is arguably the orthodoxy in the philosophy of cognitive science (Fodor 1978, 1987; Field 1978; Loar 1982; Dretske 1988; Millikan 1993; Burge 2010). According to representationalism, to have a belief is to stand in a particular relation to a mental representation. The mental representation is poised to perform certain (typically computational) functions within the mind that often bear only remote connections to stimuli, behavior, and phenomenology. On representationalist views, behaviorism failed not primarily because of its inability to account for consciousness, but rather because of its incompatibility with successful psychological explanations in terms of computational operations on representations.

While Schwitzgebel rejects the anti-phenomenological and reductionist tendencies of the behaviorists, he shares their lack of faith in the explanatory value of mental representation (at least as far as propositional attitudes are concerned). This paper is an exercise in keeping the faith. As we will argue, the science of propositional attitudes offers a wealth of data that is explicable in terms of mental representations but cannot be captured in terms of dispositions. Like Schwitzgebel, we will focus almost entirely on belief as our paradigm case—though, again like

[1] That said, his view is similar to Skinnerian behaviorism in other respects. Just as Skinner thought that what one learned (in, e.g., a concept learning paradigm) was merely to produce a set of designated responses (as opposed to acquiring a concept or belief) so too Schwitzgebel thinks that learning is a matter of acquiring dispositions to respond rather than acquiring some mental structure.

[2] He also includes "cognitive" dispositions to be in other mental states (e.g., Schwitzgebel 2013, 83). Since these mental states will themselves be dispositions to act and feel, we will follow Schwitzgebel in emphasizing the phenomenal and behavioral aspects of his view (though we will discuss cognitive dispositions in Sect. 3).

Schwitzgebel, we take it as a working assumption that the contours of a theory of belief will fit with theories of other attitudes.

First, we will consider (and reply to) Schwitzgebel's arguments for dispositionalism over representationalism. We will then survey some psychological phenomena and show how representationalist explanations succeed where dispositionalist ones do not. Our conclusion sketches a way forward for a fuller account of the attitudes which, unlike dispositionalism, rejects the spirit as well as the letter of behaviorism.

2 In-between belief and the belief box

Schwitzgebel argues that representationalism calls for a discrete yes–no answer to the question of whether a person possesses a relevant belief, and presents cases of so-called "in-between believing" that fail to force a yes-or-no intuition about whether a person possesses the belief. Dispositionalism is then taken to be superior since dispositional stereotypes allow for belief to be a graded phenomenon.

In-between beliefs are posited on the basis of a series of intuitive cases. These include:

(a) A person who gradually, over the course of a lifetime, loses the ability to recall and recognize a person's name (Schwitzgebel 2001, 76–77);
(b) A person who asserts that her son does not smoke marijuana, and yet feels suspicious when he comes home red-eyed late at night and is apt to tell her therapist that she's worried about her son's marijuana use (Schwitzgebel 2002, 260–261);
(c) A person who asserts that white people are not superior to black people, and yet displays behavior indicative of implicit bias (Schwitzgebel 2010, 532);
(d) A person who receives an email that a bridge will be out, and yet takes the route to that bridge and only recalls the email upon approaching the bridge (Schwitzgebel 2010, 533).

In these cases, according to Schwitzgebel, we do not have a clear intuition about whether the person *really* believes p or not-p. This failure of intuition is predicted by a view on which belief is not a discrete, yes-or-no phenomenon, but is rather a matter of degree.

Schwitzgebel argues that such cases resist treatment in terms of relations to mental representations. A representationalist takes statements of the form 'S believes that p' to be true if S is appropriately related to a mental representation whose content is <p>. Whether a representation (e.g., a sentence in the language of thought) is tokened in the mind is a discrete, yes-or-no phenomenon.[3] A

[3] Like Schwitzgebel, we take our paradigm version of representationalism to hold that the representations that underlie propositional attitudes are structured in a roughly language-like fashion (e.g., Fodor 1987; Mandelbaum 2016). It is not mandatory that representationalists accept this further thesis, but it will matter for some of the arguments for representationalism in the following section.

representationalist, according to Schwitzgebel, is thus left without the resources to account (or at least to easily account) for cases like (a)–(d).

Instead of holding that these cases involve a "deep" fact of the matter as is required by representationalism (Schwitzgebel 2013, 77–78), Schwitzgebel argues that our intuitions lead us to a "superficial" notion of belief. According to this superficial notion, the subjects in these cases sort of do and sort of don't have the relevant beliefs. It is "partly because of its superficiality" that Schwitzgebel's dispositionalism can handle these cases "with a flexible minimalism: Display the dispositional structure and you're done; nothing more to report!" (Schwitzgebel 2013, 86). He argues that representationalism, and its commitment to deeper facts of the matter, is less flexible.[4]

> Deeper approaches, in contrast, invite the worry that something is still left open - for example, that underneath it all, [the implicit racist in (c)] might (or must?) really have "all the races are equally beautiful" in her Belief Box, or "white people are more beautiful" there, or maybe both, and until we have figured this out, we don't know what her attitude really is, even if we know every inch of her superficial dispositional structure.
>
> (Schwitzgebel 2013, 86)

This commitment of deep views like representationalism appears less intuitive than the superficial dispositionalist alternative, so Schwitzgebel concludes that we ought to prefer dispositionalism to representationalism.

In addition to cases of in-between believing, Schwitzgebel argues that the commonly used metaphor of a "belief box" where mental representations are placed and become beliefs leads to additional paradoxes. These paradoxes can be avoided by rejecting the metaphor and endorsing dispositionalism. Schwitzgebel admits that representationalism may allow for a more graded picture of belief than the metaphor suggests (e.g., 2010, 536). Nonetheless, let's grant Schwitzgebel the assumption that for representationalism there is always a determinate fact of the matter about whether one stands in the belief relation to a given mental representation.

Schwitzgebel draws on a perennial criticism of representationalism, viz., that the beliefs we would intuitively ascribe to a subject do not always track the mental representations tokened in their mind. Here he echoes Dennett, who writes, "it should come as no news to any of you that zebras in the wild do not wear overcoats, but I hazard the guess that it hadn't occurred to any of you before just now" (1978, 104). Schwitzgebel's example concerns the belief that there are eight planets:

> It seems that I also believe that there are fewer than nine planets. But do I also believe that there are fewer than ten planets? Fewer than 11? Fewer than 127? That there are $-i^2e^0\sqrt{64}$ planets? More than just the four inner planets? That

[4] Schwitzgebel uses the example of natural kinds to illustrate the distinction between superficial and deep theories of some domain (2013, 77–78). For example, creatures called "cats" on Twin Earth might look and act just like cats here on Earth but have distinct DNA and evolutionary histories. A superficial view of cathood would take these Twin-Earth "cats" to be genuine cats given their superficial similarities, whereas a deep view would identify cats with an underlying structure and/or causal history and thus deny that Twin-Earth "cats" are really cats.

> there are eight planets within the gravitational well of the nearest large hydrogen-fusing body? That there are eight known planet-like entities within half a light year? That Shakespeare probably had too low an estimate of the number of planets? This list is, of course, potentially infinite.
>
> (Schwitzgebel 2013, 88)

Representationalism again seems committed to simple yes-or-no answers to questions that might intuitively appear to lack them, while the graded character of dispositionalism allows it to accommodate such examples. Schwitzgebel's criticisms are all part of the same critique: representationalism is inflexible insofar as it supposes there is a single place where a mental representation is stored and thereby becomes a belief. How can a belief-box picture explain how a belief is only accessible to certain processes (like sorting behavior) and not others (like speech)?

We think these criticisms fail because they rest on an inaccurate depiction of representationalism, and of the belief box metaphor in particular. We share Schwitzgebel's distaste for the metaphor, but not because we deny that fixing beliefs involves storing representations. Instead, we dislike the metaphor because it misleadingly suggests a certain picture of belief storage. The original articulation of the metaphor, originated by Schiffer (1981) but heavily developed by Fodor (1987), was meant to highlight an aspect of the representational theory of mind that has little to do with memory.

One of the central ideas of the representational theory of mind is that different mental states can share contents because they incorporate the same representations. For example, the thought TIGERS HAVE STRIPES and the thought TIGERS ARE ORANGE share a constituent (viz., TIGERS), and it is because they share a constituent that they predicate the property of being striped and the property of being orange of the same object. And since constituents are repeatable in different contexts, the representational theory of mind can explain how we can freely recombine concepts in systematic and productive ways.

Thought is also systematic at the level of attitudes: if you can believe that p, then you can also deny that p, hope that p, desire that p, and so on. The representationalist can explain this datum by positing that differences in propositional attitude types are differences in the relations one bears to mental representations, and therefore allow that distinct attitudes can relate a thinker to the same representation. Thus the idea that we distinguish belief from desire by imagining two "boxes" where we can place tokens of the same type of representation. The insight behind the belief box metaphor is not that there is a single place where beliefs are stored, but rather that propositional-attitude relations are distinct from the type-individuation conditions of mental representations.

It is natural to interpret the phrase 'belief box' as committing to a single, undifferentiated store of beliefs, but in fact the metaphor is best interpreted as making no claims about storage whatsoever. Thus, representationalism, *qua* theory of the metaphysics of belief, has no commitments about the structure of belief storage—it equally allows for the most simple and the most byzantine memory architectures. Schwitzgebel sometimes writes as though the idea of belief storage is

itself metaphorical (Schwitzgebel 2010, 537).[5] However, unlike the case of the belief box, we interpret talk of belief storage concretely: beliefs and other representational states are stored in the mind in just the same sense that representations are stored in (other) computational systems.

Our preferred architecture is one where belief storage is *fragmented* (Mandelbaum 2016). Belief fragmentation is the thesis that, rather than being stored in a single box—or, to borrow another metaphor, in one consistent web—our beliefs are stored in disparate, perhaps mutually inconsistent fragments. The idea that beliefs are fragmented is not an ad hoc stipulation to save representationalism; theorists who reject representationalism nonetheless accept fragmentation (e.g., Lewis 1982; Stalnaker 1984; Egan 2008; Elga and Rayo ms). Anyone who believes that contradictions can persist in a single person's beliefs will have to allow for some degree of fragmentation, no matter what their theory of the metaphysics of belief. However, representationalism is especially well-suited to a fragmented picture of belief. For a representationalist, beliefs are representational states that are literally *stored* in the mind, just as episodic and semantic memories are. The idea that our beliefs are fragmented can therefore be explained by positing architectural divisions between belief stores. It is thus *because* two inconsistent sets of beliefs are stored separately that they persist despite inconsistency, and that they are accessed at different times to produce different behaviors.[6]

Consider Lewis's (1982) classic example of fragmented beliefs: he believed that Nassau St. ran north–south, that it was parallel to a certain railroad, and that the railroad ran east–west. The failure to integrate one's beliefs and resolve inconsistencies seems to call for multiple fragments. Furthermore, cases of implicit bias (Mandelbaum 2016) and the automaticity of belief acquisition (Mandelbaum 2014; Mandelbaum and Quilty-Dunn 2015; Quilty-Dunn 2015) provide independent grounds for dividing beliefs into fragments. But what explains fragmentation itself? A picture that quantifies over representations stored in distinct architectural locations provides a deeper explanation of how it is that our beliefs can be fragmented.

We are now in a position to see why Schwitzgebel's cases (a)–(d) do not raise a problem for representationalism. In each case, the fact that an easy yes-or-no belief ascription eludes us does not bear on the question of representationalism. Instead, fragmented representationalist architectures can not only allow cases like the implicit racist in case (c), but can *explain* them in causal-mechanistic terms by positing representations that are acquired, stored, and accessed to cause behavior. The implicit racist stores racist and egalitarian beliefs in distinct fragments, and only the egalitarian belief fragment is accessed for conscious planning and speech behavior while the racist belief fragment is accessed in low-level behaviors (such as

[5] We wonder whether he also thinks semantic memory storage is metaphorical. The semantic memory that Trenton is the capital of New Jersey seems like just another belief to be stored, so literalism about semantic memory storage seems to entail literalism about belief storage.

[6] It is instructive to compare representationalism to MIT fragmentationalism. The latter position endorses dispositionalism and so rejects talk of belief storage, in which case it's unclear exactly what is fragmented and how fragmentation works. To our ears the Elga/Egan/Rayo style fragmentation is just a restatement, and not an explanation, of the data (of course this is in part due to different methodological concerns).

crossing the street to avoid someone of a different race).[7] Similarly for the person who lies to herself in case (b) and the person who fails to access a relevant belief in case (d). Other cases are not necessarily explained in terms of fragmentation, but are readily explained in terms of representational architectures, such as the loss of access to a belief in case (a). Schwitzgebel's objections to representationalism fail to recognize the flexibility of representationalist architectures in explaining belief-based behavior in terms of internal mechanisms of acquisition, storage, and access. Since the positing of these mechanisms is not ad hoc, there is no reason to favor non-representationalist views over representationalist ones.

It is true that some of these cases lack obvious answers. Yet this is not because of limitations on representationalist explanations, but rather because the cases are underspecified. For example, in case (a), a person gradually loses the ability to recall or recognize someone else's name. Offhand, it seems hard to pinpoint when the person loses the belief that the name is such-and-such. It doesn't follow that representationalism should merely throw its hands up. A representationalist story distinguishes failures of storage from failures of access—indeed, the distinction between recognition and recall is a distinction about the accessibility of a stored representation. Accessibility comes in degrees. To name a few ways: a representation can be accessible to a greater or lesser number of systems; it can be accessible more or less easily to a single system; it can be accessible under a greater or lesser number of independent conditions; and it can be accessible to a system with greater or lesser impact on behavior (such as central cognition vs. a modular subsystem). Precisely what sort of access failures are involved in case (a), or whether the representation fails to be stored altogether, are simply underdetermined by the description of the case. Testing the conditions under which the subject can recall or recognize a representation of the name would pull different access failures apart.

Schwitzgebel's strategy when discussing cases like (a)–(d) is to pose questions that cannot easily be answered given the limited description of the cases. The aforementioned example of believing there are eight planets is an example of this strategy. He asks whether the person who believes this also believes that there are "fewer than ten planets? Fewer than 11? Fewer than 127? That there are $-i^2e^0\sqrt{64}$ planets?" (Schwitzgebel 2013, 88). While he presents these inquiries as merely rhetorical questions, we take them to be answerable empirical questions. There is a difference between representations that are stored (and thus literally believed) and representations that aren't, but can be inferred from what is stored. All else equal, performing an inference requires more time than merely activating a stored representation. If someone has the belief that there are eight planets stored but not the belief that there are fewer than 127 planets, their judgment of the truth of 'there

[7] Of this case, Schwitzgebel asks "Does it add anything of value—anything besides confusion—to append…the claim that [the implicit racist] believes both P and its negation?" (2010, 544). The answer, on our view, is *yes*: it allows us to limn the architecture of the mind while also saying something true about what data structures a person harbors. Thus we depart from Schwitzgebel by allowing for contradictory beliefs, and we think that, intuition aside, there is all sorts of independent evidence that people have them (see Lewandowsky and Kirsner 2000; Ripley 2009; Hall et al. 2012; Legare et al. 2012).

are eight planets' will be quicker than their judgment of the truth of 'there are fewer than 127 planets'. And in cases where the subject lacks a concept altogether—such as, perhaps, the concept of $-i^2e^0\sqrt{64}$—there will be no amount of time sufficient for the subject to judge the relevant sentence, except the amount of time it would take to acquire the concept and integrate it with stored knowledge to yield a judgment.[8] In sum, there is no reason to doubt that there is a non-ad-hoc answer to the rhetorical questions posed by Schwitzgebel.

There is one aspect of Schwitzgebel's criticism that hits the mark: any empirically respectable version of representationalism makes a hash of our intuitions about belief ascription. Since the acquisition, storage, and access of representations can occur unconsciously and show up in behavior in surprising ways, a thoroughgoing representationalist should expect ordinary yes-or-no belief ascription to fail in a wide range of cases. We thus adopt a methodological modesty about the limits of folk belief ascription and look instead to the science of belief to inform our views about which cases do and don't involve certain beliefs. Folk belief ascription may be, to borrow Schwitzgebel's terminology, superficial, but that doesn't mean that belief itself is. If the empirical facts suggest that, despite superficial messiness in folk belief ascription, there are deep facts of the matter about our beliefs that a superficial account cannot explain, then that is very good evidence in favor of a deep theory. In particular, as we'll argue in the next section, the empirical facts seem to demand a theory of belief that is both deep and representationalist.

3 Evidence for the deep view of belief

3.1 Causation

A desideratum for a theory of beliefs is to explain how beliefs cause behavior (either by interacting with desires or on their own). A dispositionalist theory has in-principle problems in doing so. Dispositions only cause actions when combined with an event: the fragility of the glass alone won't cause the glass to break—to get that you need an event, e.g., the glass being hit. Likewise, a disposition to behave

[8] Although we do think there are empirically ascertainable facts of the matter about whether a belief involves a stored representation or not, we don't mean to imply that absolute reaction times alone can tell us so. Stored beliefs will be more quickly usable than inferred beliefs, but ceteris isn't always paribus: it may turn out that, e.g., retrieving a password for a website you haven't accessed in a while will take longer than verifying whether there are more atoms in the universe or Bush presidents, even though we can assume you have never previously considered the latter question. In other words, our empirical predictions about duration of processing are relative rather than absolute. We expect that stored representations that have previously been accessed will be more easily re-accessed than representations that have not been accessed since being stored.

Reaction times are always a function of which processes—storage, access, inference, etc.—occur, as well as task demands that initiate the processes. Reaction times will thus be graded to some extent, but will not necessarily constitute a perfectly "smooth gradation" (Schwitzgebel 2013, 88), since the number of operations and their respective durations will, together with the duration of sensorimotor processing required to complete the task, determine reaction time.

won't cause behavior without a mental event—such as an activation of a mental representation. Without that, there is no dispositional causation.[9]

The dispositionalist has two ways to respond to this worry. The first is to claim that dispositions aren't causal after all, but their categorical bases are. But if the categorical bases are the causal nexus, then one wonders what these bases are. The natural options are either neural or psychological. As for the former, the program of reducing one's belief that p to any neural area (or kind) is totally dead: there appear to be no generalizations to be had for what areas correspond to beliefs. So one who wanted to reduce attitudes to neural states would be left with beliefs being equivalent to an enormous unprojectible disjunction. If it's this disjunctive categorical base that is doing the causal work, it becomes difficult to see how beliefs could ever generate behavior in reliable, predictable ways.[10]

But if it's not a neural categorical base, then it must be a psychological one. This option fares no better, for the ambient options are mental representations. But if it's mental representations that are doing the causal work that beliefs are supposed to do, then it's unclear in what sense the theory is dispositionalist anymore. At the very least, the debate between the representationalist and the dispositionalist would start to look verbal.

Schwitzgebel could appeal to his superficialism and deny that beliefs are causal, in which case there's no problem for regarding them as dispositions. But this superficialist dispositionalism runs into explanatory dead-ends. For instance, Schwitzgebel accepts that beliefs serve as premises in inferences (see the discussion of "cognitive" dispositions in, e.g., Schwitzgebel 2002, 252). But inference is a causal process whereby (e.g.) believing that P and If P then Q causes one to believe that Q. In other words, inferential promiscuity demands that beliefs are causal. But if beliefs are causal, then the superficialist response won't do, in which case the dispositionalist cannot explain the causal powers of belief without co-opting

[9] A referee raises the worry that this point might prove too much, thus undermining the explanatory value of personality traits, such as the "Big Five". However, personality traits, *qua* standing dispositions, are not elements of causal-mechanical explanations of behavior (in this way trait explanation seems 'superficial' as opposed to 'deep'). For example, Barack Obama's conscientiousness is not a concrete particular that causally interacts with his sensorimotor representations to produce his behavior. That is not to deny that dispositional traits can sometimes provide predictive-explanatory value. Knowing that a person is conscientious can provide a richer understanding of why they don't have credit card debt, and can help us predict that they won't tend to keep their sink full of dirty dishes. Similarly, knowing that glass is fragile can aid understanding of why we don't build cars out of glass, and can help us predict that a baseball flying toward a window will cause it to break. Our claim is therefore not that dispositions play no role in explanation, but rather that they are not elements of causal-mechanical explanations. A purely dispositionalist psychology would not be completely vacuous, but it would lack substantive causal-mechanical explanations of behavior. A fuller theory of personality traits may invoke causally efficacious mental structures, but such a theory would thereby reduce traits in just the sort of way we aim to reduce beliefs to relations to mental representations.

[10] Schwitzgebel (2002, fn18) is admirably upfront about not being sure what to say about dispositional causation. He flirts with the idea of identifying dispositions with their categorical bases and writes "I am willing to allow the identification of believing with being in a certain categorical state as long as that state co-occurs, in all nomologically possible worlds, with the appropriate dispositional profile" (273). Identifying the categorical bases with neural states ensures that the qualifier in the above quote won't be met.

representationalism. Unlike representationalism, dispositionalism does not posit concrete mental particulars that causally interact according to psychological laws to produce new beliefs. To the extent that explanatory work is done by such entities, the theory is a representationalist one.

This last point is important for understanding precisely why the debate between the representationalist and the dispositionalist is not merely verbal. The latter can countenance the existence of representations and posit them to explain low-level subpersonal mental phenomena. But an anti-representationalist metaphysics of belief cannot allow for representations to provide lawlike causal explanation of core features of belief. The metaphysics to which we should commit (at least for entities that fall within the scope of natural science) is the metaphysics determined by the posits that figure in successful explanations. If the best explanation for why various generalizations about belief are true is fundamentally a representationalist one, then we're obligated to endorse a representationalist metaphysics of belief. Ceding the explanatory ground to representations is not compatible with holding that representations are extrinsic to the nature of belief itself.

The remainder of this section will outline various generalizations about belief, both from cognitive science and common sense, and argue that invoking representations consistently provides the best explanation of such generalizations.

3.2 Sorting

A classic counterexample to behaviorist theorizing stemmed from a basic case with a rich pedigree: sorting. Subjects who are asked to sort equivalent classes behave differently based on how they conceptualize the task. So, for instance, people who are given a deck of cards and asked to make a pile consisting of spades and clubs are much faster and less error prone than those who are asked to make a pile of non-diamonds and non-hearts (Bruner et al. 1956). These extensionally equivalent sorts lead to extremely different behaviors based on how one represents what the task is. The well-known effect of negation increasing processing load leads to slower times and worse performance. All we aim to add to this is that this basic behavioral fact is, in the sorting case, also a fact about *belief*. What one takes one's task to be dictates how one performs on the task: if you believe your task to be sorting the non-diamonds and non-hearts you will behave differently than if you believe you are sorting the clubs and spades.

The representational account of belief can explain this fact, while the dispositionalist cannot. Why should these instructions have any effect on performance? For a representationalist, the representations SORT THE SPADES AND CLUBS and SORT THE NON-DIAMONDS AND NON-HEARTS are extensionally equivalent but conceptually (and syntactically) distinct. The latter includes negation operators attached to the predicates DIAMONDS and HEARTS while the former includes no negation. Call SPADES AND CLUBS sorters "positive" sorters, and NON-DIAMONDS AND NON-HEARTS sorters "negative" sorters. When presented with a spade, a positive sorter categorizes it as a spade, matches that category to her belief that she must sort spades and clubs, and sorts it accordingly. A negative sorter will still likely

automatically categorize a spade as a spade, and must then translate SPADE into NON-DIAMOND AND NON-HEART in order to perform the task. It is because these representations have different representational structures that they result in different behavior in the positive and negative sorting cases.[11]

A dispositionalist, on the other hand, cannot advert to representational structure here. While a dispositionalist can posit subpersonal representations, they deny that our person-level beliefs themselves are representational. The representationalist explanation of sorting just sketched does not merely posit subpersonal representations; rather, it posits logically structured, linguistically expressible conceptual representations that explain why *believing* that the task is to sort non-diamonds and non-hearts slows down behavior. As argued in the dilemma posed in the previous subsection, dispositionalists can admit of representations only if they are peripheral to explanations of generalizations about belief. The representationalist explanation of sorting on offer is not peripheral in this way. Here representations are posited specifically to account for how our beliefs about the task affect our performance on the task.

Barring representational explanations, one would expect that, when being an X and being a Y are mutually exclusive and exhaustive, a disposition to sort Xs and a disposition to sort non-Ys would be identical. What disposition could be called upon to explain the data? Perhaps the dispositionalist would say: to believe that one is to sort the spades and clubs is just to be disposed to sort the spades and clubs in a quick and error-free manner, while to believe that one is to sort the non-diamonds from the non-hearts is to be disposed to sort them slowly and poorly. But where do these dispositions come from?[12]

This example highlights a slipperiness in Schwitzgebel's view. For instance, he writes, "Once the dispositions are fully characterized the question of what the subject believes is closed" (2002, 273). His idea is that the categorical bases of belief don't matter once we note what dispositions the subject has. But the vehicles of representation themselves (in part) dictate how one is disposed to behave, so representations cannot drop out of the picture in favor of dispositions. On the contrary, the effect—an effect of what the subject *believes* on how she behaves—must be explained in terms of differences in the structural features of representational vehicles.

3.3 Opacity and truth evaluability

Two core aspects of belief are naturally explained by taking belief to be representational: opacity and truth evaluability. It's well known that one can believe

[11] Even if one thinks that negation is special in some way, and thus should be ignored, there are generalizations to be had about the mere form of representations outside of negation—for example, conjunctive concepts are applied faster and with fewer errors than disjunctive ones (Wason and Johnson-Laird 1972).

[12] The question of why any belief has the stereotype it does is never answered on Schwitzgebel's view. Why assertion is central to the stereotype of some beliefs and not others is just stated, but never explained (e.g., 2013, 81–82).

X to be F while not believing Y to be F even though X and Y are identical. Frege cases are legion, and a theory of belief should at least provide a sketch of how to solve the issue. The representationalist offers a response: one can have two representations with the same content without knowing that they corefer (Fodor 1998; Edwards 2014). The dispositionalist just says that you are disposed to behave differently with respect to Xs and Ys even though they are coextensive, and offers no explanation of Frege cases. Schwitzgebel (2002, 265) argues that Frege cases are another instance of in-between belief wherein we sort of do and sort of don't believe that X is F. His view seems to take for granted that belief is opaque, however, rather than explaining it; differences in representational form generate opacity while differences in behavioral dispositions are merely a consequence of opacity.[13]

Similarly, an account of belief should account for the fact that beliefs are truth evaluable—some beliefs are true, whereas others aren't. A representationalist view can easily explain this fact. Representations can represent or misrepresent, and sentence-like structures are themselves truth evaluable. It is unclear in what sense dispositions are truth evaluable, in which case it's difficult to see how dispositionalists can explain how beliefs are truth evaluable. Dispositionalists can say that it is simply constitutive of belief that it is truth evaluable, so the dispositions that realize belief are *ipso facto* truth evaluable—and they can seek to offer some supplementary theory of truth evaluability. The representationalist, however, can both grant the constitutivity claim and also provide a deeper explanation of opacity and truth evaluability in one fell swoop, since representations have truth-evaluable contents and type-distinct representations can corefer.

3.4 Beliefs' similarity to other propositional attitudes

Although beliefs differ from other propositional attitudes, there are truisms that arise between the attitudes and serve as desiderata for a theory of beliefs. For example, beliefs can be focused on the same content as any other propositional attitude. One can believe that P, or doubt that P, or hope that P, and so on. As mentioned in the previous section, the representationalist can easily explain how this is so: each propositional attitude is just a relation to a given representation. The differences between the attitudes are differences in the sort of relations that are instantiated. The representationalist theory thus provides a substantive explanation of the datum that our various attitudes may concern the same propositions. The dispositionalist theory isn't inconsistent with this datum, but it cannot explain it either.

Moreover, there is a curious parallel between what it is possible to say and what it is possible to believe (see the discussion of "Vendler's Condition" in Fodor 1978).

[13] A referee suggests that perhaps we are implicitly restricting the dispositionalist too much by assuming they don't have access to a de dicto reading of beliefs. However, these same problems arise even for de dicto beliefs in Mates cases (see, e.g., Fodor 1998 for discussion).

In short, it appears that anything you can say you can also believe, and vice versa.[14] Again, if beliefs are relations to (language-like) mental representations, this is easily explained, since both thought and language are composed of representations and linguistic representations express corresponding conceptual representations. It is unclear how the dispositionalist could explain this datum, since dispositionalism does not allow for beliefs to exhibit any language-like structure.

3.5 Dissonance and belief change

Throughout, we have been quantifying over relations to mental representations. It's time to say a bit about what these relations are, for our preferred view isn't just representationalist, it's *psychofunctionalist*. Psychofunctionalism is an answer to the question of what type of relations differentiate one propositional-attitude type from another. It says that the relations are to be given from law-like generalizations uncovered in cognitive science (as opposed to, say, analytic functionalism, which says that the relations are to be given from commonsense platitudes).

As we see it, the basic problem that classic versions of psychofunctionalism (e.g., Block and Fodor 1972; Block 1980) ran into for the attitudes was that the generalizations weren't forthcoming; the only one ever posited seemed to be the practical syllogism. But that wasn't because of any deep fact—it was just a function of philosophers not investigating the relevant data from social psychology.[15] Nevertheless, there is an enormous, important science of belief that provides data that any theory of belief must explain.

Consider dissonance theory (e.g., Festinger 1957; Cooper 2007). In the abstract, dissonance theory tells you that if someone believes that P, and receives information that not P, that disconfirming information will *hurt* (Elliot and Devine 1994). In particular, it will put the agent into a negatively valenced state (the dissonance), one that the agent will be motivated to escape because of its painful nature. To do that, one needs to assuage the dissonance by dealing with the inconsistency. The theory garners its predictions from the multiple ways one can reduce dissonance.[16] One paradigm is *induced compliance*, whereby subjects are manipulated into behavior that goes against their standing beliefs. In a classic case (Festinger and Carlsmith 1959) subjects spend time doing a boring, pointless task (turning a bunch of knobs 90 degrees until they are fully rotated to their original position). After completing the task, subjects are then asked (but, importantly, not forced) to tell other subjects

[14] There may of course be cases where one is unable to express a belief (e.g., implicit bias). Vendler's condition holds that any *type* of belief could be expressed by some type of sentence, not that any particular token belief is poised for linguistic expression.

[15] This is due to an interesting historical accident. The philosophers most associated with psychofunctionalism—Ned Block and Jerry Fodor—were also most closely aligned with cognitive psychology. Social psychology went mostly undetected. As such the role of social psychologists (such as Leon Festinger and Stanley Milgram) in the downfall of behaviorism has been greatly underappreciated as compared to the role of cognitive psychologists (George Miller) and linguists (Chomsky).

[16] In our preferred reading of the theory, dissonance is created by a logical inconsistency between two beliefs, and this inconsistency generally occurs by friction between one's three core beliefs (that one is a good person, a smart person, and a reliable, consistent person; see Aronson 1992).

who are waiting to take part in the study that the study was really fun; that is, the subjects are asked to lie to their peers. In exchange for lying subjects were given either a large reward ($20) or a small reward ($1). Finally, the subjects are asked to rate how much they liked the original task.

Contra reinforcement theory's predictions, subjects who are paid less money report liking the task *more*. Reinforcement theory wrongly predicts that the greater the reward, the more the subject will like the task, but the dissonance theorist correctly realizes that self-justification is key to understanding how beliefs change. Before lying, all subjects detested the task equally. Those who were paid $20 continue to believe the task was terrible. They know they lied to the others, but they did so for a good reason: $20 is a decent amount of money now (and was a ton of money in 1959). On the other hand, subjects who were paid $1 also lied, but $1 isn't enough money to justify why they lied (compare: $1000 would justify doing the electric slide in the middle of a subway car, but one penny wouldn't). If these subjects believe that they lied and believe that they did so for no real reason, then they must have lied because they were bad people. But dissonance theory posits that most everyone believes that they are morally good, so the $1 group thus feels dissonance: they know they aren't bad people but have evidence that they have done something bad. They reduce this dissonance by changing their attitude about the task: they no longer believe the task was boring but now instead they believe the task was fun. Thus the dissonance is assuaged because the low-reward subjects no longer took themselves to be lying. The same mechanism of unconscious self-justificatory reasoning is at play in other well-known effects, such as effort justification (Aronson and Mills 1959) and counterattitudinal essay writing.

Dissonance reduction is perfectly general and ubiquitous. It appears nearly every time we make a free choice. Moreover, the reasoning is *unconscious*—those who cannot explicitly remember their original attitudes show more movement than those who do. Take the "Spreading of Alternatives" (or "free choice") paradigm, where subjects are asked to rank items. Say E. J. is given 12 Billy Joel albums (or household appliances, or motivational posters, etc.) and asked to rank them in order of their desirability, such as it is. E. J. ranks *The Stranger* and *The Nylon Curtain* 6 and 7 so that he has very little difference of opinion between them. He is then told he can choose one to take home, and does. If later asked to re-rank his choices again, the album he chose will move up his rankings, and the one he didn't will move down (Brehm 1956). But this holds only for people who don't remember their original rankings. In fact, the effect is considerably larger if you use anterograde amnesiacs, who cannot form any explicit new memories, as your subjects (Lieberman et al. 2001).

Let's recap by returning to one of Schwitzgebel's examples mentioned above: case (b). In that case a mother of a teenage son implicitly suspects he is smoking pot, and shows this by acting suspicious and feeling dread when her son comes home late with bloodshot eyes. Yet in most moods she cannot bring herself to consciously consider the possibility. But it's worth focusing on a nice, subtle detail that Schwitzgebel adds to the case: that the mother openly deplores her friend's parenting because her friend's child smokes pot. This detail feels so familiar and reasonable it's easy for us to move right past it, but it's not just a random addendum.

Rather, it is predicted by dissonance theory. When the mother is reminded of teenagers smoking pot, she has two relevant, dissonant beliefs activated: the belief that her son smokes pot and the belief that he doesn't smoke pot.[17] This induces dissonance. She then expels the dissonance not by resolving the inconsistency but rather through a mode of projection: by being extra-specially judgmental of her friend. Schwitzgebel recognizes that this is a normal, predictable mode of behavior, but his theory doesn't explain why it comes about. The psychofunctionalist has an answer: it's because beliefs are concrete particulars with causal powers, governed by, *inter alia*, the laws of dissonance.

The laws of dissonance state, roughly, that beliefs that will generate a negative, motivational, phenomenologically salient discomfort whenever one encounters counterattitudinal evidence. We will then be moved to assuage this feeling not via the most rational route available, but generally by any easily available route. We have also seen that the mechanism for doing so works best when unconscious. But of course all of this evidence takes belief to be a state, and a deep one–one that has laws about what things it causes and what causes it to change. And there is much more evidence of this sort.

Beliefs lead to attentional effects: if you believe in X, you will selectively avoid information that reflects poorly on X, while seeking out information that is consistent with your belief (Brock and Balloun 1967). Beliefs also lead to polarization: if you believe extremely strongly in X and receive information against X, then you will, counterintuitively, increase your belief in X (Festinger et al. 1956). Say you've joined a cult and you've told all your friends and family that it's the greatest, most beneficent group ever. If you are presented with unequivocal and undeniable evidence as to the cult's failings you will not be able to merely accept the evidence and give up your belief—instead you will *increase* your belief in the cult by focusing more on the cult and it's good qualities (Mandelbaum ms). In fact, thinking about a proposition will not only increase its accessibility (Krosnick and Petty 1995, 10), but also merely thinking about a proposition ("my cult is good!") will increase your strength of belief in the proposition (Tesser 1978).[18] Even beliefs that you've supposedly rejected will often stretch their inferential tentacles over time. Say you read that Obama is a Muslim, but since the source is the New York Post, you reject the information immediately. Nevertheless, the "sleeper effect" dictates that over time this belief will come untethered from its source and increase in strength even though it was originally rejected (Kumkale and Albarracin 2004). It is unclear how, or why, a disposition would increase with strength over time if that disposition is never exercised. Moreover, there is some reason to think beliefs are acquired automatically even when subjects fail to consciously endorse them (Mandelbaum 2014), which, together with the sleeper effect, can help explain why

[17] In which case the answer to Schwitzgebel's question ("But what does [the mother] believe *now*, while she's working intensely on a client's account and not giving the matter any thought") is: both that her child smokes pot and that her child doesn't smoke pot.

[18] This is why self-affirmation works—thinking "I'm a good person!" leads people to believe it more strongly (Mandelbaum 2014).

large percentages of otherwise rational populations have odd beliefs (see Mandelbaum and Quilty-Dunn 2015 for examples).

These are all generalizations about belief, and *pace* superficialists (like Dennett and Schwitzgebel), they are counterintuitive and deep. Such generalizations are the bread and butter of a mature psychofunctionalism. By the same token, they undermine the explanatory usefulness of a superficialist dispositionalism.

3.6 Concepts, inference, and the representational theory of mind

The examples discussed so far have been specific to belief. This acquiesces to the dialectic as Schwitzgebel sets it up: representationalism and dispositionalism are two approaches to propositional attitudes, and the dispute should be adjudicated by appeal to evidence that is specific to propositional attitudes. While we agree that data about belief are most salient to the dispute, other data are relevant as well. Representationalism about the attitudes is part of a more general representational theory of mind. The key claim of the representational theory of mind is that mental states are built up out of representations. We turn now to evidence for the representational theory of mind as it applies to thoughts (and therefore to propositional attitudes like belief).

Two of the classic arguments for structured representations are the arguments from systematicity and productivity (e.g., Fodor 1987). The capacity to form the thought (whatever its attitude) that the tiger sees the mouse tends to co-occur with the capacity to form the thought that the mouse sees the tiger. The fact that being able to think thoughts of the form aRb goes along with being able to think thoughts of the form bRa is explained by a view on which thoughts literally have *forms*, and are constructed out of atoms that can compose into various other thoughts. The same holds for productivity, i.e., the capacity to think new thoughts. This capacity can be explained by the capacity to compose atomic representations in new ways.

The debates over systematicity and productivity are well-worn, and we do not wish to get embroiled in them here. Instead, we'd like to focus on an underappreciated argument for the idea that thoughts are composed out of concepts: the existence of semantic priming. Semantic priming is one of the most robust and well-validated effects in cognitive science. When a subject reads the word 'doctor', and then has to discriminate words from non-words (e.g., hit the YES key in response to 'bread' and the NO key in response to 'drabe'), her reaction time will be faster in identifying semantically related words like 'nurse' than in identifying unrelated words like 'tree' (e.g., Meyer and Schvaneveldt 1971).

One plausible explanation of priming is that mental representations are stored in associative networks such that activating one representation (by, e.g., reading the word that expresses it) activates representations connected to it in the network. The basic apparatus used to explain semantic priming involves representations that are (literally) stored in semantic networks and that compose into larger representations that express propositional contents (which is why, for example, you're more likely to think doctor-related thoughts if you've just read the word 'doctor'). This apparatus seems to implicate the very same architecture implicated to explain systematicity and productivity. And while systematicity and productivity have been

controversial (see Gendler Szabo 2012 for details), the existence of semantic priming is a datum for everybody (though behavioral priming is of course another story altogether; see Doyen et al. 2012).

We are not sure what Schwitzgebel would have to say about this datum, or about systematicity and productivity for that matter. Dispositionalism does not have the resources to say why thought is compositional—dispositionalism does not even allow for the idea that thoughts have constituents! What would the constituents of a disposition be? And why think they could compose at all, let alone in the specific ways mental representations do?

Schwitzgebel might respond that his theory is not meant to explain these features of thought. That is fair enough, but something must explain them, and the most plausible candidate is structured mental representations. Given that we already have independent reason to invoke mental representations to explain well-known features of thought, and given that propositional attitudes are, after all, kinds of thoughts, an account that construes propositional attitudes in terms of structured mental representations has some independent verification.

The marriage between representationalism about belief and the representational theory of mind more generally becomes especially clear when we revisit the nature of inference. There is arguably no better candidate for a constitutive feature of belief than the fact that beliefs are inferentially promiscuous. If you believe that tigers are striped, and you believe that X is a tiger, then you have all the materials you need to infer that X is striped. While conscious reasoning often feels like a whirlwind of "mental chaos" (Siegel 2017, 99), unconscious inference seems to be automatic and syntactic. When subjects are presented with a semantically sparse but syntactically well-formed sentence like 'If there is a 3 then there is an 8', and are then subliminally presented with '3', they show facilitation for '8' (Reverberi et al. 2012). Subliminally presenting '8', however, does not facilitate '3'.

Results like these suggest that inference operates on representations in respect of their syntactic, formally specifiable properties: any thoughts of the form P and IF P THEN Q will trigger an inference to Q because a rule of mental logic specifies types of constituent structure that conform to something like modus ponens (Quilty-Dunn and Mandelbaum 2017). Once again, the dispositionalist seems to be without an explanation. In this instance, mental representations cannot be dismissed as independent of propositional attitudes. It's structural features of those mental representations that explain a core feature of belief, viz., inferential promiscuity. The inferential promiscuity of beliefs is explained by the same syntactic apparatus implicated by semantic priming and other data suggestive of conceptual compositionality. It is thus not open to the dispositionalist to accept representational explanations in central cognition but deny their role in explaining core features of propositional attitudes. The representational structures that explain these various generalizations are not mere subpersonal attendants to genuine belief. It is because *beliefs themselves* have representational structures that they exhibit these generalizations. We need structured mental representations in the mind generally, and we need them in the metaphysics of belief as well.

4 Conclusion

Let's take stock. Schwitzgebel critiques the idea of psychofunctional theory by arguing that none has been presented "except as an optimistic promise or simplistic cartoon sketch of the mind" (Schwitzgebel 2013, 94). In contrast, we have presented a psychofunctional theory that identifies beliefs as relations to mental representations, with the relations characterized by the psychological generalizations that hold over belief. These generalizations include that beliefs are acquired ballistically and automatically, put subjects into a negatively valenced motivational state when encountering disconfirming evidence, are changed in ways that will assuage that state, will increase in strength over time if left alone, will increase in strength even more if repeatedly tokened, and will increase in accessibility the more they are activated.[19] The mental representations themselves allow for beliefs to be causal and can explain the opacity of beliefs. On this view, beliefs look deep.

Compare to the dispositionalist view. The dispositionalist cannot explain mental causation or belief opacity, and, in virtue of its superficialism, cannot explain any of the generalizations about belief. The superficialist must deny that these effects exist, or otherwise ignore them. And it's not just dispositionalists that are up the creek. Interpretationists—those who think that what we believe is just a matter of interpretation, perhaps from those of us taking the "intentional stance" or using a principle of charity—are also a species of superficialist. They see nothing deep about beliefs. But if we look a bit deeper, we find that there is much more to belief than appears on the surface.

Acknowledgements Thanks to Joseph Bendaña, Michael Brownstein, Zoe Jenkin, Rob Long, Jesse Rappaport, Eric Schwitzgebel, and Jennifer Ware.

References

Aronson, E. (1992). The return of the repressed: Dissonance theory makes a comeback. *Psychological Inquiry, 3,* 303–311.

Aronson, E., & Mills, J. (1959). The effect of severity of initiation on liking for a group. *Journal of Abnormal and Social Psychology, 59,* 177–181.

Block, N. (1980). Troubles with functionalism. In N. Block (Ed.), *Readings in the philosophy of psychology* (Vol. 1 and 2, pp. 268–305). Cambridge, MA: Harvard University Press.

Block, N., & Fodor, J. (1972). What psychological states are not. *The Philosophical Review, 81*(2), 159–181.

[19] There's much more we could add to this list to constrain the relation, such as the wisdom of crowds effect, anchoring and adjusting, the efficacy of self-affirmation, etc. See Mandelbaum 2010 for details.

Brehm, J. (1956). Postdecision changes in the desirability of alternatives. *The Journal of Abnormal and Social Psychology, 52*(3), 384–389.
Brock, T., & Balloun, J. (1967). Behavioral receptivity to dissonant information. *Journal of Personality and Social Psychology, 6*(4.1), 413–428.
Bruner, J., Goodnow, J., & Austin, G. (1956). *A study of thinking*. New York: Wiley.
Burge, T. (2010). *The origins of objectivity*. Oxford: OUP.
Cooper, J. (2007). *Cognitive dissonance: 50 Years of a classic theory*. London: Sage Publications Ltd.
Davidson, D. (1984). *Inquiries into truth and interpretation*. Oxford: Clarendon.
Dennett, D. (1978). A cure for the common code. In D. Dennett (Ed.), *Brainstorms: Philosophical essays on mind and psychology* (pp. 90–108). Cambridge: Bradford Books.
Doyen, S., Klein, O., Pichon, C. L., & Cleeremans, A. (2012). Behavioral priming: It's all in the mind, but whose mind? *PLoS ONE, 7*(1), e29081.
Dretske, F. (1988). *Explaining behavior*. Cambridge, MA: MIT Press.
Edwards, K. (2014). Keeping (direct) reference in mind. *Noûs, 48*(2), 342–367.
Egan, A. (2008). Seeing and believing: Perception, belief formation and the divided mind. *Philosophical Studies, 140*(1), 47–63.
Elga, A., & Rayo, A. (ms). Fragmentation and information access.
Elliot, A. J., & Devine, P. G. (1994). On the motivational nature of cognitive dissonance: Dissonance as psychological discomfort. *Journal of Personality and Social Psychology, 67,* 382–394.
Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.
Festinger, L., & Carlsmith, J. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology, 58,* 203–210.
Festinger, L., Riecken, H., & Schachter, S. (1956). *When prophecy fails*. Minneapolis: University of Minnesota Press.
Field, H. (1978). Mental representation. *Erkenntnis, 13*(1), 9–61.
Fodor, J. (1978). Propositional attitudes. *The Monist, 61,* 501–523.
Fodor, J. (1987). *Psychosemantics*. Cambridge, MA: MIT Press.
Fodor, J. (1998). *Concepts: Where cognitive science went wrong*. Oxford: OUP.
Gendler Szabo, Z. (2012). The case for compositionality. In W. Hinzen, E. Machery, & M. Werning (Eds.), *The oxford handbook of compositionality* (pp. 64–80). Oxford: OUP.
Hall, L., Johansson, P., Strandberg, T., & Martinez, L. M. (2012). Lifting the veil of morality: Choice blindness and attitude reversals on a self-transforming survey. *PLoS ONE, 7*(9), e45457.
Krosnick, J. A., & Petty, R. E. (1995). Attitude strength: An overview. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 1–24). Hillsdale, NJ: Erlbaum.
Kumkale, G., & Albarracin, D. (2004). The sleeper effect in persuasion: A meta-analytic review. *Psychological Bulletin, 130*(1), 143–172.
Legare, C. H., Evans, E. M., Rosengren, K. S., & Harris, P. H. (2012). The coexistence of natural and supernatural explanations across cultures and development. *Child Development, 83*(3), 779–793.
Lewandowsky, S., & Kirsner, K. (2000). Knowledge partitioning: Context-dependent use of expertise. *Memory and Cognition, 28*(2), 295–305.
Lewis, D. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy, 50,* 249–258.
Lewis, D. (1982). Logic for equivocators. *Noûs, 16*(3), 431–441.
Lieberman, M., Ochsner, K., Gilbert, D., & Schacter, D. (2001). Do amnesics exhibit cognitive dissonance reduction? The role of explicit memory and attention in attitude change. *Psychological Science, 12*(2), 135–140.
Loar, B. (1982). *Mind and meaning*. Cambridge: Cambridge University Press.
Mandelbaum, E. (ms). Troubles with Bayesianism: An introduction to the psychological immune system.
Mandelbaum, E. (2010). The architecture of belief: An essay on the unbearable automaticity of believing. Doctoral Dissertation, University of North Carolina, Chapel Hill.
Mandelbaum, E. (2014). Thinking is believing. *Inquiry, 57*(1), 55–96.
Mandelbaum, E. (2016). Attitude, inference, association: On the propositional structure of implicit bias. *Noûs, 50*(3), 629–658. doi:10.1111/nous.12089.
Mandelbaum, E., & Quilty-Dunn, J. (2015). Believing without reason, or: Why liberals shouldn't watch Fox News. *The Harvard Review of Philosophy, 22,* 42–52.
Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90,* 227–234.
Millikan, R. (1993). *White queen psychology and other essays for alice*. Cambridge, MA: MIT Press.

Quilty-Dunn, J. (2015). Believing in perceiving: Known illusions and the classical dual-component theory. *Pacific Philosophical Quarterly, 96*(4), 550–575.
Quilty-Dunn, J., & Mandelbaum, E. (2017). Inferential transitions. *Australasian Journal of Philosophy*. doi:10.1080/00048402.2017.1358754
Quine, W. V. (1960). *Word and object*. Cambridge: MIT Press.
Reverberi, C., Pischedda, D., Burigo, M., & Cherubini, P. (2012). Deduction without awareness. *Acta Psychologica, 139,* 244–253.
Ripley, D. (2009). Contradictions at the borders. In *International workshop on vagueness in communication* (pp. 169–188). Berlin: Springer.
Ryle, G. (1949). *The concept of mind* (2002nd ed.). Chicago: University of Chicago Press.
Schiffer, S. (1981). Truth and the theory of content. In H. Parret & J. Bouveresse (Eds.), *Meaning and understanding* (pp. 204–222). Berlin: de Gruyter.
Schwitzgebel, E. (2001). In-between believing. *Philosophical Quarterly, 51,* 76–82.
Schwitzgebel, E. (2002). A phenomenal, dispositional account of belief. *Noûs, 36,* 249–275.
Schwitzgebel, E. (2010). Acting contrary to our professed beliefs, or the gulf between occurrent judgment and dispositional belief. *Pacific Philosophical Quarterly, 91,* 531–553.
Schwitzgebel, E. (2013). A dispositional approach to attitudes: Thinking outside of the belief box. In N. Nottelman (Ed.), *New essays on belief: Constitution, content, and structure* (pp. 75–99). New York: Palgrave Macmillan.
Sellars, W. (1956). Empiricism and the philosophy of mind (New edition, 1997). Cambridge, MA: Harvard University Press.
Siegel, S. (2017). *The rationality of perception*. New York: Oxford University Press.
Stalnaker, R. (1984). *Inquiry*. Cambridge, MA: MIT Press.
Tesser, A. (1978). Self-generated attitude change. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 11, pp. 229–338). New York: Academic Press.
Wason, P., & Johnson-Laird, P. (1972). *Psychology of reasoning: Structure and content*. Cambridge: Harvard University Press.