Dispositionalism, Yay! Representationalism, Boo!

Eric Schwitzgebel
Department of Philosophy
University of California, Riverside
Riverside, CA 92521
USA

January 27, 2023

Dispositionalism, Yay! Representationalism, Boo!

Abstract: We should be dispositionalists rather than representationalists about belief. According to dispositionalism, a person believes when they have the relevant pattern of behavioral, phenomenal, and cognitive dispositions. According to representationalism, a person believes when the right kind of representational content plays the right kind of causal role in their cognition. Representationalism overcommits on cognitive architecture, reifying a cartoon sketch of the mind. In particular, representationalism faces three problems: the Problem of Causal Specification (concerning which specific representations play the relevant causal role in governing any particular inference or action), the Problem of Tacit Belief (concerning which specific representations any one person has stored, among the hugely many approximately redundant possible representations we might have for any particular state of affairs), and the Problem of Indiscrete Belief (concerning how to model gradual belief change and in-between cases of belief). Dispositionalism, in contrast, is flexibly minimalist about cognitive architecture, focusing appropriately on what we do and should care about in belief ascription.

Keywords: belief, cognitive architecture, Fodor, propositional attitudes, representation,

Word count: approx. 8600, plus one figure

Dispositionalism, Yay! Representationalism, Boo!

If belief matters, it matters because what you believe governs what you do, how you feel, and the conclusions you tend to draw. Therefore, as I will explain in this essay, we should be *dispositionalists* about belief rather than *representationalists*. Dispositionalism prioritizes what we do care about and ought to care about in ascribing beliefs to ourselves and others. Representationalism plunges us into a quagmire of simplistic, premature, and needless commitments about deep cognitive architecture. Representationalism overcommits on facts about cognitive architecture of derivative importance, while dispositionalism is wisely minimalist.

1. Representationalism Described.

The core idea of representationalism is this: Normally – at least in typical or standard cases – when someone believes some proposition P, they have a representation with the content P stored somewhere in the functional architecture of their mind, a representation that plays, or stands ready to play, a particular functional role. This functional role, the *belief-like functional role*, has both "backward-looking" and "forward-looking" features. Looking backward, representations play the relevant functional role if they arise through causal processes that normally respond to evidence favoring the truth of P. Looking forward, representations play the relevant functional role if they enter into inferential or inference-like relationships with other representations in theoretical and practical reasoning. Different versions of representationalism can add twists or caveats to this simple sketch (e.g., Mandelbaum 2019 on the "psychological immune system").

Suppose Cynthia believes that there is beer in the fridge. According to representationalism, in the standard case, the history of her belief was this: Evidence that there is beer in the fridge – for example, visual evidence if she has recently looked in the fridge, or testimonial evidence from her roommate – caused her to "token" a representation with the content *there is beer in the fridge*. This representational content was then stored somewhere in her mind. A few hours later, maybe, Cynthia thinks to herself *how pleasant it would be to drink a cold beer*. The representation *there is beer in the fridge* is then activated, or retrieved from storage. Cynthia then employs that representation in theoretical reasoning, perhaps combining it with representations like *the fridge is in the kitchen* and *the kitchen is nearby* to conclude that *beer is nearby*. She employs that representation in practical reasoning too, forming the intention to go to the fridge to retrieve the beer.

It's a neat causal picture. Facts about the world cause representations of those facts, and those representations then cause new inferentially supported representations and practical behavior. To believe that P, in the standard or paradigmatic case, is just to have an internal representation with the content P, playing or standing ready to play that belief-like functional role.

Commonly, representationalists employ the metaphor of the "belief box" – a hypothetical functional region (not necessarily a physiologically distinct brain region) where beliefs are stored and from which they are retrieved. See, for example, Figure 1, from Nichols and Stich (2000).

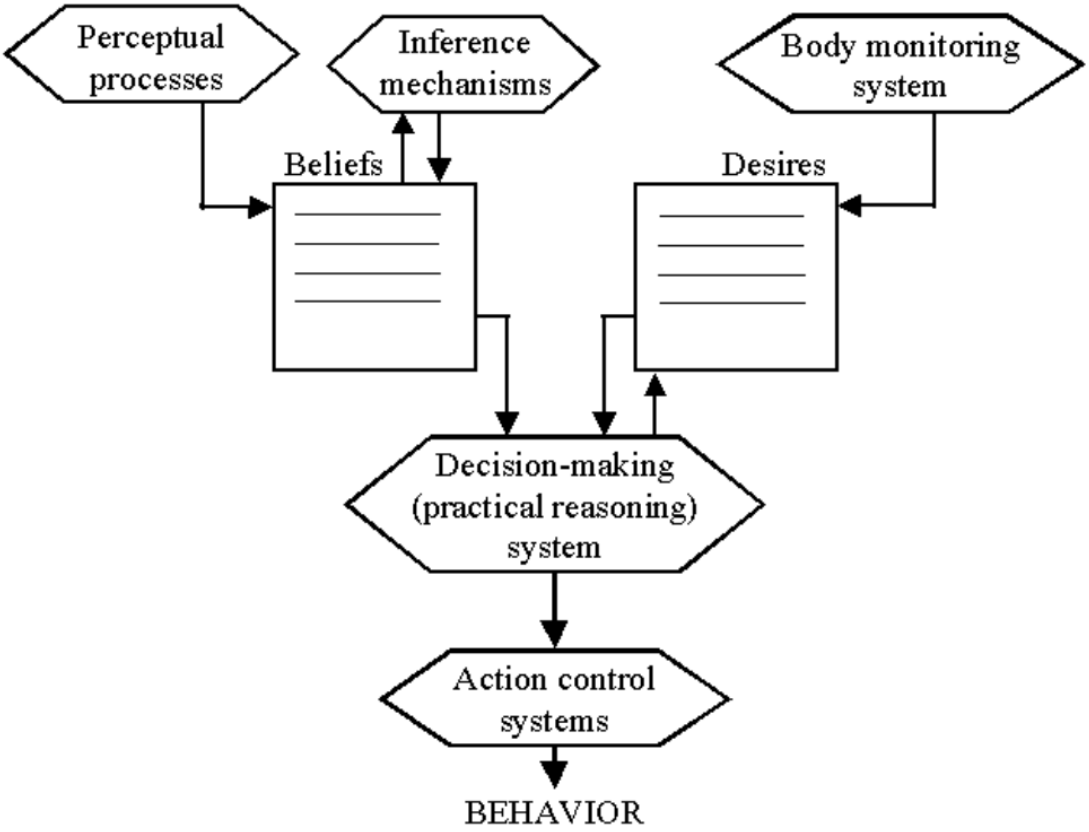

Figure 1: An example of representationalist cognitive architecture, featuring a "belief box" where representational contents are stored and from which representational contents are retrieved for use in inference and decision making, from Nichols and Stich 2000.

This is, I think we can all agree, a sometimes useful model. The question is *how realist* to be about such a model. No one, of course, thinks that there is any sort of literal box in the head where beliefs are stored. Nor need beliefs be stored in a single physiologically distinct region, nor even in a single functionally distinct system as opposed to a variety of distributed systems. One can be realist enough – even, so to speak, an "industrial-strength" representational realist (to borrow a term from Dennett 1991) – without committing to any of that. I propose that

we understand industrial-strength representational realism about belief – "representationalism" in the sense employed this essay – in terms of the following four commitments.

Presence. In standard, non-"tacit", cases, belief that P requires that a representation with the content P is present somewhere in the mind. In Section 5, we'll explore the tacitness qualification. This qualification creates, I will argue, a huge headache for the representationalist. For now, note that Presence plausibly constitutes a core commitment of any representationalist realism worth its name.

Discreteness. In standard cases, a representation P will be either discretely present in or discretely absent from a cognitive system or subsystem. It is generally implicit in representationalist models that any particular representation P is either present or absent, either stored or not stored, in any particular cognitive subsystem or set of subsystems. The models typically leave no room for representations being, say, half-present or 23% present or indeterminately hovering between present and absent. Indeterminately present representations cannot readily be made sense of in standard representationalist architectures. Representationalists do not talk normally employ indeterministic language or explain how half-stored representation might work. (In contrast, representationalism can easily accommodate determinately present representations with vague or partly indeterminate *contents*, such as *some people were there;* that's an entirely different matter.) Some marginal cases might violate discreteness – nature has few truly sharp borders, if one zooms in close enough – but these will be brief or rare exceptions.

Kinematics. Rational actions arise from the causal interaction of beliefs that P and desires that Q, in virtue of their specific contents P and Q, or at least in virtue of syntactic or architectural correlates of those specific contents (e.g., Fodor 1987). Similarly, rational

inferences involve the causal interaction of beliefs that P with other beliefs to generate still more beliefs. This is central to the representational realist's causal story.

Specificity. Rational action arises from the activation or retrieval of *specific* sets of beliefs and desires $P_{1...n}$ and $Q_{1...m}$, as opposed to other, related beliefs and desires $P'_{1...j}$ and $Q'_{1...i}$. More accurately, rational action arises from the activation or retrieval of the specific representations whose storage, in the right functional location, constitutes possessing the beliefs and desires $P_{1...n}$ and $Q_{1...m}$. Similarly, rational inference arises from the activation or retrieval of specific sets of representations. This commitment follows from kinematics and discreteness. If the representations are discrete, and we're realist about the kinematics, then there must be a fact of the matter specifically which representations are doing the causal work.

As I interpret them, Jerry Fodor and Eric Mandelbaum are industrial-strength representational realists about belief, accepting these four commitments (Fodor 1981, 1987; Quilty-Dunn and Mandelbaum 2018; Bendaña and Mandelbaum 2021; Porot and Mandelbaum 2023; note that Quilty-Dunn, Porot, and Mandelbaum 2022 defends a thesis that is in some respects substantially weaker). In personal communication, Mandelbaum has confirmed the accuracy of this portrayal.

I hope for two things from this section. First, I hope I have fairly articulated the underlying commitments of an interesting and influential form of representational realism about belief. Second, I hope that readers will be struck by the boldness and implausibility of these commitments.

2. Dispositionalism Described.

The core idea of dispositionalism is this: To believe some proposition P is to match, to an appropriate degree and in appropriate respects, a certain dispositional profile – the dispositional profile characteristic of the belief that P. What exactly belongs in the dispositional profile varies between accounts. Behaviorist dispositionalism (e.g., Marcus 1990) includes only behavioral dispositions in the profile: To believe that P is to be disposed to behave in ways characteristic of a P believer. Phenomenalist dispositionalism (e.g., Smithies this volume) includes only experiential dispositions in the profile: To believe that P is to be disposed to have the phenomenology characteristic of a P believer, for example a feeling of conviction that P is true when the topic arises. Liberal dispositionalism – my own preferred view – includes a broad range of dispositions in the profile: To believe that P is to be disposed very broadly to act and react, to think and feel, inwardly and outwardly, in a manner characteristic of a P believer (e.g., Schwitzgebel 2002, 2021; also Baker 1995; Hunter 2009 [though Hunter 2022 rejects the view]; and in psychology the "functionalism" of De Houwer, Hughes, Barnes-Holmes 2017; De Houwer 2022).

To better understand liberal dispositionalism, consider three broad classes of dispositions: behavioral, phenomenal, and cognitive. As a beer-in-the-fridge believer, Cynthia will have behavioral dispositions like being disposed to go to the fridge if she wants a beer. She'll have phenomenal dispositions like being disposed to say to herself in inner speech with a feeling of assent "there's beer in the fridge". And she'll have cognitive dispositions like the disposition to conclude that Shriya's favorite drink is in the fridge upon learning that beer is Shriya's favorite drink.

Three complications immediately rise. One concerns the nature of cognitive dispositions. Dispositionalism works best if all the relevant dispositions are "person level" rather than

subpersonal (Bantegnie ****; Pober 2022). Relevant cognitive dispositions can include, for example, dispositions to acquire new beliefs and desires, dispositions to make certain assumptions, and dispositions to form certain intentions. If we expanded the set of cognitive dispositions to include, say, the disposition to token a certain type of representation in a certain type of cognitive subsystem, then dispositionalism would risk collapsing into representationalism. (As an aside, it might be attractive to assume that all the relevant cognitive dispositions are eliminable en masse by Ramisification (Lewis 1972): They would then be, so to speak, ontological placeholders for transitional states that can be characterized in terms of behavioral dispositions, phenomenal dispositions, and other placeholder transitional states. They would be Xs, Ys, and Zs definable wholly in terms of their relations to each other and to phenomenal and behavioral events and dispositions.)

A second complication concerns the *ceteris paribus* (all else being equal or normal or right) nature of dispositions. Consider the behavioral disposition to go to the fridge if one wants a beer. Someone might perfectly well believe that there is beer in the fridge and yet avoid the fridge despite wanting a beer – for example, if they're occupied with an important video call. But if we're too liberal about springing the *ceteris paribus* clause we risk vacuity. There must be some empirical commitment and risk.

A third complication, especially for liberal dispositionalists, is how to characterize the relevant dispositional profile without leaning on representationalism. Exactly *which* dispositions are constitutive of believing that P? What commonality is shared among the disposition to go to the fridge if one wants a beer, to answer in the affirmative if asked if there's a beer in the fridge, to feel surprise if one were to open the fridge and find no beer, etc.? It is objectionably circular to respond that these are just the dispositions that a beer-in-the-fridge believer would in fact

have. And we'd also better not say: these are just the dispositions that are typically present when one has the representation P stored in one's mind, ready to play a belief-like role in cognition. Such a view would be crypto-representationalism.

The natural way to address the second and third complications becomes evident upon considering the analogous case of personality traits. Plausibly, to have a personality trait is also to match a dispositional profile. To be an extravert is just to be disposed to act and react, think and feel, as an extravert would – to be ready to say "yes" to party invitations, to enjoy meeting new people, to tend to take the lead in social situations. Of course all of these dispositions are ceteris paribus. Extraverts will decline party invitations when they have to stay home to watch the kids. That fact doesn't undermine the usefulness of saying that extraverts tend to say yes to party invitations. The conflicting commitment, we might say, "masks" the disposition; or the disposition contains implicit manifestation conditions such as "absent conflicting commitments". This doesn't collapse into vacuity in the case of personality traits, so we can at least hope, and maybe assume, that it also doesn't collapse into vacuity in the case of belief. In neither case is it a matter of a single disposition. Rather, it's a matter of the overall profile.

As for populating the dispositional profile: The dispositional profile constitutive of extraversion is also complex. We construct it partly through a priori, armchair consideration of what belongs to the concept of extraversion and partly by noticing what patterns of dispositions tend to cluster diagnostically with the dispositions that belong a priori to the concept. In personality psychology, the latter can be formalized through factor analysis, combined with tests of convergent and discriminant validity. The a priori and empirical go hand-in-hand. We can create new dispositional stereotypes for personality traits, as "narcissistic" was invented in the late 19th century (Freud 1914/2012) and "extraverted" in the early 20th century (Jung 1915).

Creation of such stereotypes proceeds partly through conceptual stipulation and partly through empirical observation of what types of thoughts and behavior tend to co-occur. In the case of belief, dispositional profiles must also be constructible for novel beliefs never before entertained, such as the belief that emus despise bowling. Fortunately, we can readily do so using the tools of commonsense folk psychology. Someone who believes that emus despise bowling will be disposed, for example, to say yes if asked "do emus despise bowling?", would be surprised to discover emus who appear to enjoy watching or participating in the sport, and would infer that Adrian despises bowling upon learning that Adrian despises everything that emus despise.

Importantly, dispositionalism – whether behaviorist, phenomenal, or liberal – is silent about deep cognitive architecture. Anyone who matches closely enough the dispositional profile constitutive of believing that P is a P believer, regardless of the structures that underlie those dispositions.

3. The General Architectural Implausibility of Representationalism.

Human heads normally contain a complex organ called the "brain" (Aristotle 4th c. BCE/2019). To a first approximation, this brain consists of eighty billion neurons linked by axons and dendrites, plus supportive glial cells and a healthy supply of blood. The bulk of cognitive action appears to depend on the connectivity and transfer of signals among neurons. There is no obvious functional structure or set of structures where beliefs are stored or where beliefs that P couple with desires that Q to give birth to intentions to R. Of course, beliefs could be stored in some non-obvious way, as the representionalist presumably holds. But from the standpoint of initial biological plausibility, it seems more likely that when Cynthia decides to head to the fridge for a beer, her action arises from complex neural interaction patterns that don't

map neatly onto simple linguistic contents like "there's beer in the fridge". This is of course not decisive. However, it creates an initial explanatory hurdle for the industrial-strength representationalist.

It didn't have to be this way. In fact, it isn't this way for all representational structures. Industrial-strength representational realism is biologically plausible, for example, about some aspects of early visual processing. There really do seem to be cells, for example, that represent the presence, or not, of luminance gradients at certain orientations in certain regions of the visual field, by reacting selectively to that type of input (Wandell and Winawer 2011; Carandini 2012). In principle, we could have found cells or regions or functional systems that represent specific beliefs in a similar way – for example, by having a certain physiological structure if and only if you believe that there's a beer in the fridge. Industrial-strength representational realism would then be impressively confirmed!

The likelier bet is that cognition works via distributed, soft-edged, and complex structures in which representations with folksy contents like *there's beer in the fridge* are not generally either discretely present or absent. Nor do cognitive scientists exploring the architecture of the mind typically appeal to beliefs that P and desires that Q. Although mainstream cognitive scientists are sometimes more representationally realistic than I suspect is fully warranted, the kinds of representations they usually evoke are things like memory traces, object features, and associations among lexical items. There's a world of difference between a low-level amodal representation of an object on a certain trajectory having certain geometrical and color properties and belief that the square is red. Even if a species of representationalism about low-level cognition is correct, the truth of industrial-strength representationalism about belief by no means follows.

Postulating a one-to-one mapping of believed-in contents like *there's beer in the fridge* to stored cognitive representations more closely resembles old-school "Good Old Fashioned AI" than it resembles real, cutting-edge cognitive science. The Cyc Project has devoted decades to attempting to encode every piece of human common knowledge like *Paris is the capital of France* and *Once a physical part has been removed from an object, it generally can't be removed again*. As of 2019, the project had employed over 2000 scientist-years of effort explicitly encoding over 25 million generalizations using about 1.5 million concepts (Cyc.com 2019). The idea is that if an AI stored all of those representations and could access them for reasoning when relevant, it could engage in general-knowledge human-like reasoning. Although there have been some useful applications – such as helping physicians access medical knowledge (Lenat et al. 2010) and a database of information about terrorist organizations (Lenat and Deaton 2008), it is probably fair to say that this project has overall borne little visible fruit relative to the decades of effort committed to it (Davis and Marcus 2015).

The current shape of AI is much more connectionist. Deep learning is hot, as are long vector representations that don't map in any straightforward way onto ordinary human concepts. Philosophers shouldn't bet that the architecture of the human mind is similar to the architecture that was trendy in computer science in the 1970s. It was fine that Jerry Fodor made that bet in 1975, but now we know better.

4. The Problem of Causal Specification: One Billion Beer Beliefs.

Cynthia rises from the couch to go get that beer. If we accept industrial-strength representationalism, in particular the Kinematics and Specificity theses, then there must be a fact

of the matter *exactly which representations* caused this behavior. Consider the following possible candidates:

- There's beer in the fridge.
- There's beer in the refrigerator door.
- There's beer on the bottom shelf of the refrigerator door.
- There's beer either on the bottom shelf of the refrigerator door or on the right hand side of the lower main shelf.
- There's beer in the usual spot in the kitchen.
- Probably there's beer in the place where my roommate usually puts it.
- There's Lucky Lager in the fridge.
- There are at least three Lucky Lagers in the fridge.
- There are at least three and no more than six cheap bottled beers in the fridge.
- In the fridge are several bottles of that brand of beer with the rebuses in the cap that I used to illicitly enjoy with my high school buddies in the good old days.
- Somewhere in the fridge, but probably not on the top shelf, are a few bottles, or less likely cans, of either Lucky Lager or Pabst Blue Ribbon, or maybe some other cheap beer, unless my roommate drank the last ones this afternoon, which would be uncharacteristic of her.

This list could of course be continued indefinitely. Estimating conservatively, there are at least a billion such candidate representational contents. For simplicity, imagine nine independent parameters, each with ten possible values.

If Kinematics and Specificity are correct, there must be a fact of the matter *exactly which subset* of these billion possible representational contents were activated as Cynthia rose from the

couch. Presumably, also, various background beliefs might or might not have been activated, such as Cynthia's belief that the fridge is in the kitchen, her belief that the kitchen entrance is thataway, her belief that it is possible to open the refrigerator door, her belief that the kitchen floor constitutes a walkable surface, and so on – each of which is itself similarly specifiable in a massive variety of ways.

Plausibly, Cynthia believes all billion of the beer-in-the-fridge propositions. She might readily affirm any of them without, seemingly, needing to infer anything new. Sitting on the couch two minutes before the beery desire that suddenly animates her, Cynthia already believed, it seems – in the same inactive, stored-in-the-back-of-the-mind way that you believed, five minutes ago, that Obama was U.S. President in 2010 – that Lucky Lager is in the fridge, that there are probably at least three beers in the refrigerator door, that there's some cheap bottled beer in the usual place, and so on. If so, and if we set aside for now (see Section 5) the question of tacit belief, then Cynthia must have a billion beer-in-the-fridge representations stored in her mind. Specificity requires that it be the case that exactly one of those representations was retrieved the moment before she stood up, or exactly two, or exactly 37, or exactly 814,406. Either exactly one of those representations, or exactly two, or exactly 37, or exactly 814,406, then interacted with exactly one of her desires, or exactly two of her desires, or exactly 37, or exactly 814,406. But which one or ones did the causal work?

Let's call this the Problem of Causal Specification. If your reaction to the Problem of Causal Specification is to think, yes, what an interesting problem, if only we had the right kind of brain-o-scope, we could discover that it was exactly the representation *there are 3 or 4 Lucky Lagers somewhere in the refrigerator door*, then you're just the kind of mad dog representational realist I'm arguing against.

I think most of us will recognize the problem as a pseudo-problem. This is not a plausible architecture of the mind. There are many reasonable characterizations of Cynthia's beer-in-the-fridge belief, varying in specificity, some more apt than others. Her decision is no more caused by a single, precisely correct subset of those billion possible representations than World War I had a single, possibly conjunctive cause expressible by a single determinately true sentence. If someone attempts to explain Cynthia's behavior by saying that she believes there is beer in the fridge, it would be absurd to fire up your brain-o-scope, then correct them by saying, "Wrong! She's going to the fridge because she believes there is Lucky Lager in the refrigerator door." It would be equally absurd to say that it would require wild, one-in-a-billion luck to properly explain Cynthia's behavior absent the existence of such a brain-o-scope. The industrial-grade representationalist presumably would think that as a matter of ordinary folk practice, "there is beer in the fridge" is close enough; but still, according to their causal story the explanation would be, strictly speaking, erroneous.

A certain variety of representationalist might seek to escape the Problem of Causal Specification by positing a single extremely complex representation that encompasses all of Cynthia's beer-in-the-fridge beliefs. A first step might be to posit a map-like representation of the fridge, including the location of the beer within it and the location of the fridge in the kitchen. This map-like representation might then be made fuzzy or probabilistic to incorporate uncertainty about, say, the exact location of the beer and the exact number of bottles. Labels will then need to be added: "Lucky Lager" would be an obvious choice, but that is at best the merest start, given that Cynthia might not remember the brand and will represent the type of beer in many different ways, including some that are disjunctive, approximate, and uncertain. If maps can conflict and if maps and object representations can be combined in multiple ways, further

complications ensue. Boldly anticipating the resolution of all these complexities, the representationalist might then hypothesize that this single, complicated representation is the representation that was activated. All the sentences on our list would then be imperfect simplifications – though workable enough for practical purposes. One could perhaps similarly imagine the full, complex causal explanation of World War I, detailed beyond any single historian's possible imagining.

This move threatens to explode Presence, the idea that when someone believes P there is a representation with the content P present somewhere in the mind. There would be a complex representation stored, yes, from which P might be *derivable*. But many things might be derivable from a complex representation, not all of which we normally will want to say are believed in virtue of possessing that representation. If a map-like representation contains a triangle, then it's derivable from the representation that the sum of the interior angles is 180 degrees; but someone ignorant of geometry would presumably not have that belief that simply in virtue of having that representation (cf. "the problem of logical omniscience": Stalnaker 1991; Égré 2021). Worse, if the representation is complex enough to contain a hidden contradiction, then presumably (by standard laws of logic) literally every proposition that anyone could ever believe is derivable from it.

The move to a single, massively complex representation also creates an architectural challenge. It's easy to imagine a kinematics in which a simple proposition such as *there is beer in the fridge* is activated in working memory or a central workspace. But it's not clear how a massively complex representation could be similarly activated. If the representation has many complex parameters, it's hard to see how it could fit within the narrow constraints of working memory as traditionally conceived. No human could attend to or process every aspect of a

massively complex representation in drawing inferences or making practical decisions. More plausibly, some *aspects* of it must be the target of attention or processing. But now we've lost all of the advantages we hoped to gain by moving to a single, complex representation. Assessing which aspects are targeted throws us back upon the Problem of Causal Specification.

Cynthia believes not only that there's beer in the fridge but also that there's ketchup in the fridge and that the fridge is near the kitchen table and that her roommate loves ketchup and that the kitchen table was purchased at Ikea and that the nearest Ikea is thirty miles west. This generates a trilemma. Either (a.) Cynthia has *entirely distinct* representations for her beer-in-the-fridge belief, her ketchup-in-the-fridge belief, her fridge-near-the-table belief, and so on, in which case even if we can pack everything about beer in the fridge into a single complex representation we still face the problem of billions of representations with closely related contents and an implausible commitment to the activation of some precise subset of them when Cynthia gets up to go to the kitchen. Or (b.) Cynthia has *overlapping* beer-in-the-fridge, ketchup-in-the-fridge, etc. representations, which raises the same set of problems, further complicated by commitment to a speculative architecture of representational overlap. Or (c.) all of these representations are somehow all aspects of one *mega-representation*, presumably of the entire world, which does all the work – a representation which of course would always be active during any reasoning of any sort, demolishing any talk about retrieving different stored representations and combining them together in theoretical inference.

Dispositionalism elegantly avoids all these problems! Of course there is some low-level mechanism or set of mechanisms, perhaps representational or partly representational, that explains Cynthia's behavior. But the dispositionalist need not commit to Presence, Discreteness, Kinematics, or Specificity. There need be no determinate, specific answer exactly what

representational content, if any, is activated, and the structures at work need have no clean or simple relation to the beliefs we ascribe to Cynthia. Dispositionalism is silent about structure. What matters is only the pattern of dispositions enabled by the underlying structure, whatever that underlying structure is.

Instead of the storage and retrieval metaphor that representationalists tend to favor, the dispositionalist can appeal to figural or shaping metaphors. Cynthia's dispositional profile has a certain shape: the shape characteristic of that of a beer-in-the-fridge believer – but also, at the same time, the shape characteristic of a Lucky-Lager-in-the-refrigerator-door believer. There need be no single determinately correct way to specify the shape of a complex figure. A complex shape can be characterized in any of a variety of ways, at different levels of precision, highlighting different features, in ways that are more or less apt given the describer's purposes and interests. It is this attitude we should take to characterizing Cynthia's complex dispositional profile. Attributing a belief is more like sketching the outline of a complex figure – perhaps a figure only imperfectly seen or known – than it is like enumerating the contents of a box.

5. The Problem of Tacit Belief.

Back in the late 1970s to early 1990s, that is, in the heyday of philosophical representational realism about belief, several representationalists noticed a problem closely related to the Problem of Causal Specification, the Problem of Tacit Belief (Field 1978; Lycan 1986; Crimmins 1992; Manfredi 1993). Not all of them regarded it as a problem, exactly. Some regarded it as a discovery. But as a discovery, it proved useless: The literature on tacit belief petered out, rather than proving fruitful.

We can enter the Problem of Tacit Belief by noticing that it's not wholly implausible that people have infinitely many beliefs. A billion beer beliefs is just the start of it! Cynthia believes, presumably, that there are fewer than 100 bottles of beer in her fridge. She therefore also seemingly believes that there are fewer than 101 bottles, and fewer than 102, and fewer than a thousand, and fewer than million, and fewer than 16,423,300.6, and so on. If we accept that Cynthia does in fact believe all that (presumably, she would readily assent to those propositions if asked), then she has infinitely many beliefs about the number of bottles in her fridge. However, it is implausible that each of these beliefs is grounded in a separately stored representational content.

Thus was born the distinction between *core beliefs*, those that are explicitly stored and represented, and *tacit beliefs*, those whose contents are swiftly derivable from the core beliefs. Suppose Cynthia has a stored representation with the content *there are four bottles of Lucky Lager in the refrigerator door*. This is her core belief. From this core belief, an infinite number of tacit beliefs are now swiftly derivable: that there are fewer than five bottles of Lucky Lager in the refrigerator door, that there are fewer than six bottles, etc., and also (given that she knows that Lucky Lager is a type of beer) that there are four bottles of beer in the refrigerator door, and also (given that she knows that whatever is in the refrigerator door is also in the fridge) that there are four bottles of Lucky Lager in the fridge, and also (given that she knows that Lucky Lager is cheap) that there are a few bottles of cheap beer in the fridge. Nearly all of Cynthia's billion beer-in-fridge beliefs might be tacit, grounded in just a few core beliefs.

Although postulating a core/tacit distinction helps the representationalist avoid populating the mind with infinitely many mostly redundant stored representations, a band of merry troubles follows.

First, it's worth noting that this maneuver constitutes a substantial retreat from Presence. As formulated, in the normal or standard case, when someone believes that P they have a stored representation with the content P. I don't think it is uncharitable to characterize representationalists as tending to say this; it's very much how they ordinarily talk. But now it looks like the vast majority of our beliefs might be abnormal or nonstandard. It would be shocking if even 1% of Cynthia's billion beer-in-the-fridge beliefs were explicitly represented: That would be 10 million distinct stored representations for this one minor set of facts about the world. Many other beliefs surely range into the tacit millions or billions: My belief that my wife and I started dating in grad school, your belief that racism was prevalent in Louisiana in the 1920s, Ankur's belief that there's a gas station on the corner of University and Iowa. Each of these beliefs has many, many close neighbors, in combinatorial profusion – many more neighbors, largely redundant, than it's plausible to suppose exist as distinct, robustly real, stored representations. At best, the "normal" case of having a stored representation with exactly the content P when you believe that P is a rarity. Furthermore, we don't distinguish core beliefs from very nearby tacit ones in our ordinary belief attribution, and there is no practical reason to do so.

Suppose the representationalist acknowledges this, modifying Presence appropriately: To believe that P, in the standard case, is to have a stored representation from which the content of P is swiftly derivable. Now they face the complementary challenge of resisting the conclusion that we believe huge numbers of propositions it's implausible to suppose we believe. To determine if a number is divisible by 3, add its digits. If the sum of its digits is divisible by 3, then the digit itself is. Knowing this, the proposition *112 is not divisible by three* is now, for you, swiftly derivable from propositions that you explicitly represent. But unless you're the type of person

who spends a lot of time thinking about what numbers are divisible by what others, it seems that you don't believe that proposition before actually doing the calculation. Before doing the calculation, you are, so to speak, *disposed to* believe that 112 is not divisible by three. But (dispositionally) believing is one thing and being disposed to believe is quite another. This is a distinction that we do folk-psychologically make, at least in a fuzzy-bordered way (Audi 1994). But the dispositional-belief/disposition-to-believe distinction is decidedly *not* the core/tacit distinction the representationalist wants and needs. Still worse – echoing a trouble I mentioned in Section 4 – if we have any conflicting representations, it will arguably turn out that we tacitly believe literally everything, if everything follows from a contradiction – and presumably swiftly enough given the rules of reductio.

Furthermore, postulating a core/tacit distinction requires abandoning empirical evidence for the sake of an ungrounded and possibly untestable architectural speculation. It requires that there be an important psychological difference between your core beliefs and your tacit ones. Either Cynthia stores *there's beer in the fridge,* leaving tacit *there's Lucky Lager in the fridge,* or she stores *there's Lucky Lager in the fridge,* leaving tacit *there's beer in the fridge,* or she stores both, leaving neither tacit, or she stores neither, both being quickly derivable from some other stored representational content. Cynthia's billion beer beliefs divide sharply into a few core ones and plethora, presumably, of tacit ones. But no evidence from cognitive science speaks in favor of sharply dividing our beliefs into those that are core and those that are tacit. Indeed, it's hard see how such a claim could realistically be tested. Might we, for example, look for different response times to questions about beer versus Lucky Lager? Maybe that would be a start. But it seems unlikely that we could really separate out such patterns from linguistic processing time and other sources of difficulty or facilitation of response (see also Quilty-Dunn and Mandelbaum

2018, note 8). Could we look for higher levels of activity in brain regions associated with explicit inference? Maybe. But again, there are many reasons that such regions might be active when considering whether there is beer in the fridge.

To avoid an impossible proliferation of representations, the industrial-strength representationalist needs a sharp distinction between core and tacit beliefs. But the distinction has no practical importance, doesn't map onto ordinary patterns of belief attribution, has no empirical support, and it's unlikely that we could even realistically test for it with existing methods. It's a useless posit of a fake difference, a pseudo-distinction required when the representationalists' simplistic theory crashes against our unsimple world.

6. Indiscrete Belief.

Imagine a spectrum of racism. On one end stand whole-hearted believers in White supremacy. On the other end stand whole-hearted believers in the equality of the races. Those at the racist end are very aptly describable as believing that White people are superior to Black people. Those at the egalitarian end are very aptly describable as not believing that White people are superior to Black people. Between these extremes runs a multidimensional spectrum of intermediate, half-hearted, inconsistent, or partial racist belief. On a liberal dispositionalist approach, the racist belief that White people are superior is constituted by the suite of racist dispositions – behavioral, phenomenal, and cognitive. There is nothing more to having the racist belief than having enough of those dispositions. As we move along the spectrum, the dispositions fade away, become less consistent, are replaced by egalitarian dispositions, and it becomes ever less appropriate to ascribe the racist belief. There is no moment at which – pop! – the racist belief moves from determinately present to determinately absent. Structurally, having

a racist belief is like having a personality trait: It's fuzzy-bordered matter of your overall dispositional structure. It's a matter of whether your behavioral, phenomenal, and cognitive posture toward the world more closely resembles the posture of the racist or the posture of the egalitarian. (See Schwitzgebel 2010 for a more patient treatment of how a dispositional approach to belief handles implicit bias.)

Or consider a belief that at first blush looks like just the sort of belief that might be underwritten by a discrete representation: the belief that Mengzi debated Gaozi in passage 6A1 of the *Mengzi*. One on end of our spectrum stands a student of Chinese philosophy familiar with the passage, which stands fresh in their memory. On the other end stands that same student at the end of their life, unfortunately lost deep in dementia, remembering nothing about Chinese philosophy, much less about Mengzi in particular. Between the two extremes, imagine a multidimensional spectrum of forgetting. One dimension is the name of the author: Perhaps our student slides from being able to freely recall "Mengzi" to being able to recognize the name on a list of very close alternatives to being able to recognize the name only given more distant alternatives to knowing only that it has two syllables and doesn't start with "Z". Another dimension concerns who Mengzi was, independent of the fact about his name: the second most-prominent ancient Confucian and also an advocate of the view that human nature is good, or just an ancient Confucian, or just a Chinese philosopher, or just a philosopher. Another dimension, still concerning the Mengzi portion of this proposition, concerns the likelihood of memory under various conditions given various prompts. We might imagine increasing uncertainty, or being disposed to generate the correct answer only in rare contexts with a good helping of chance. Of course, we can construct similar spectra for the other features of this proposition: the fact that Mengzi debated *Gaozi*, the fact that he *debated* Gaozi, and where the passage occurred.

On a dispositionalist approach, the appropriateness of attributing someone the belief that Mengzi debated Gaozi in *Mengzi* 6A1 depends on how closely they match the dispositional profile characteristic of believing that proposition. If some newcomer to Chinese philosophy were to ask where the Mengzi-Gaozi debate appears, would they be ready to answer? If they were to open an authoritative edition of the *Mengzi* to the beginning of Book 6 and not see the name Gaozi, would they feel surprised and confused? If they felt an urge to refresh themselves on Gaozi's views, would they flip to that part of the Mengzi (perhaps among checking other resources)? The farther they drift from being characterizable this way, the less apt the belief ascription, until we might not want to say that they believe Mengzi debated *Gaozi* in 6A1 or that Mengzi debated Gaozi *in 6A1*, etc. It's implausible that there are sharp dividing lines or always a single best summary of the propositional content.

Representational realists about gradual forgetting will likely jump on the competence-performance distinction: Our hero remains *competent* to retrieve the representation as long as there are some cases in which the representation can be successfully retrieved, but their *performance* nonetheless becomes ever less dependable. However, that move only obscures and postpones the issue. Suppose that they really store the representation *Mengzi debated Gaozi in* Mengzi *6A1* somewhere in their memory, though the representation becomes more difficult to retrieve over time. Unless we stipulate that no representations are ever lost, that representation will be gone after dementia has taken its course, and it's unsupported and untestable architectural speculation (akin to that of imagining a discrete list of core beliefs about beer in the fridge) to suppose that there's a specific moment when that representation disappears.

Even if there is a specific moment at which belief ceases, it strains credulity and scrambles our belief ascription practices to suppose that someone believes a proposition as long

as it is retrievable in *some* circumstances, including when they are so deep into forgetting that only by the rarest chance would they retrieve it and most of the time they would either guess incorrectly or answer that they don't know. If our erstwhile student wracks their brain and simply cannot come up with the answer, sincerely professing ignorance of material unstudied since student days decades past, and if this is furthermore their *typical* response, it is not the case that they believe, in the face of their sincere denials of that belief, that Mengzi debated Gaozi in *Mengzi* 6A1. This is so even if some little trace remains such that a rare alignment of factors could in principle make the answer pop to mind. We neither do or should employ the concept of belief in this manner. Thus, even if memory works by the storage of discrete representations with contents like *Mengzi debated Gaozi in passage 6A1*, we still ought to think of belief in terms of the dispositional pattern.

Industrial-strength representationalism thus entangles us in a nest of difficulties concerning in-between and indeterminate cases of implicit bias and gradual forgetting – not to mention self-deception, ambivalence, referential or conceptual confusion, momentary forgetting, procedural knowledge, inconsistency, gradual learning, and mood- or cue-dependent states (see also Schwitzgebel 1999, 2001, 2002). Call this the Problem of Indiscrete Belief.

7. The Sweets of Superficialist Minimalism.

Quilty-Dunn and Mandelbaum (2018) accuse dispositionalism of being unscientific because it does not attempt to describe the lower-level mechanisms of belief. I reply that industrial-strength representationalism about belief reifies a mere cartoon-sketch of the mind. Scientifically speaking, it's better not to commit to a mechanism than to postulate one with no real empirical support and a wagonload of implausible implications.

Science can stay at the surface. Consider, again, personality psychology. Few personality psychologists attempt to describe the underlying mechanisms of extraversion. Instead, they explore relationships among the dispositional patterns constitutive of the trait and the relationships between those patterns and other social and psychological facts. To hold that we are being unscientific unless we posit a causally efficacious inner representations with contents like *there's beer in the fridge* is like holding that a scientific approach to extraversion requires positing an ontologically real, causally efficacious inner switch set to "E" rather than "I".

Space aliens arrive in glittering ships. They learn Chinese, English, and Esperanto, and they teach us their own language. They tell hilarious and tragic tales from their homeworld, complete with slideshows and artifacts. They reveal the secrets of fusion drives, gravity control, and better cotton candy. Not only do they outwardly behave in just the sorts of ways we would expect organisms with beliefs to behave, but they also have belief-like phenomenology: They feel surprise, they have appropriate imagery and inner speech, they experience confidence and doubt. If we know all this about them, we know they have beliefs. We needn't know anything about their inner architecture except that somehow it supports these patterns of behavior, cognition, and phenomenology. Dispositionalism, of course, makes neat, lovely sense of this.

The representationalist, in contrast, faces a dilemma: Either agree that alien visitors of this sort would have beliefs, or contend that whether or not they have beliefs would depend on further facts about their cognitive architecture.

The first option strips representationalism of its empirical content. Claims about the necessity of a specific underlying cognitive structure lose their force. Having a belief-like dispositional profile proves sufficient for believing, whatever the architectural facts. No specific

architectural commitments remain. At best, the representationalist will need to make an *a priori* case – an armchair philosophical case – that steering spaceships, feeling surprise, and speaking passable English requires robustly real discrete internal representations that satisfy Presence, Discreteness, Kinematics, and Specificity.

The second option redefines belief in terms of something we don't and shouldn't care about nearly as much. It is part of the ordinary conception of belief that aliens of this sort would have beliefs, as manifested by every treatment of comprehensible aliens in science fiction. (Incomprehensible aliens, of the sort imagined by Lem (1961/1970) and Watts (2006/2020), might not believe, but they don't meet the conditions described above.) A representationalist might acknowledge this, proposing that we revise our ordinary concept of belief to exclude such aliens unless they have the right cognitive architecture beneath the surface; but this sacrifices everything that is important in belief. Faced with such a representationalist linguistic proposal, we would have to invent a new term for what aliens and humans have in common when they are inclined to act and react as if P is the case; that new term would be the better and more useful term; and it would just be a synonym for what we meant all along by "belief".

Set aside the science fiction. In a mundane sense, what do we care about in belief ascription? Relatedly, but not identically, what should we say *makes it true* that someone believes or fails to believe? Not that people have one architecture rather than another. Rather, that they behave in certain ways, are prone to make certain inferences rather than others, and have certain images, feelings, and reactions rather than others. The underlying architecture that implements all this is only derivatively important – important exactly because it enables those patterns of behavior, phenomenology, and cognition, and only to the extent that it does so. Ontologically and practically it's the patterns that count, not the stuff from which the patterns are

built. Beneath it all could be vortices of peanut butter, eddies of magnetism in compressed gas, or noumenal properties of ectoplasm; regardless, if we match the dispositional profile we believe. Some aliens might even have belief and desire boxes that store propositions in the language of thought. Even if that were so, dispositionalism would still be the correct way to characterize their attitudes, because it would capture what makes their beliefs worth calling beliefs.

This last observation suggests a point of possible reconciliation between dispositionalism and a form of representationalism. Since dispositionalism is silent about underlying structure, an underlying representationalist architecture is entirely consistent with dispositionalism; and perhaps a case can be made for some form of underlying representationalist architecture, even if the specific form described in this essay fails. If so, representationalism could potentially succeed as an account of the architecture of cognition while dispositionalism still captures why that architecture counts as the architecture of belief. (Compare De Houwer, Hughes, and Barnes-Holmes 2017 on reconciling cognitive and functional approaches to psychology; and Pober's 2022 hybrid functionalism.)

The dispositionalist view is superficialist in the sense that – if we think of our behavioral, phenomenal, and cognitive dispositions as the "surface" – belief is present if the surface is right, regardless of what transpires behind the surface. The view is minimalist in the sense that it attributes to believers the properties that we do care about and should care about in belief ascription, and nothing more, avoiding the unnecessary puzzles and misalignments that result from overcommitment. What we need from an account of belief is an ontology that follows good attributional practice, without screwing things up. Liberal dispositional delivers exactly that.

Acknowledgements

For useful comments and discussion, thanks to audiences at University of Antwerp and Princeton University; Brice Bantegnie, Eric Mandelbaum, Bence Nanay, and Jeremy Pober; and commenters on relevant posts on my blog and social media accounts.

References

Aristotle (4th c. BCE/2019). Generation of animals and history of animals I, parts of animals I. Trans. C. D. C. Reeve. Hackett.

Audi, Robert (1994). Dispositional beliefs and dispositions to believe. *Noûs, 28,* 419-434.

Baker, Lynne R. (1995). *Explaining attitudes*. Cambridge University Press.

Bendaña, Joseph, and Eric Mandelbaum (2021). The fragmentation of belief. In C. Borgoni, D. Kindermann, and A. Onofri, eds., *The Fragmented Mind*. Oxford University Press.

Carandini, Matteo (2012). Area V1. *Scholarpedia, 7* (7):12105.

Crimmins, Mark (1992). Tacitness and virtual beliefs. *Mind & Language, 7,* 240-263.

Cyc.com (2019). *Technology overview*. https://www.cyc.com/wp-content/uploads/2019/09/Cyc-Technology-Overview.pdf [accessed Jan. 16, 2023].

Davis, Ernest, and Gary Marcus (2015). Commonsense reasoning and commonsense knowledge in Artificial Intelligence. *Communications of the ACM, 58* (9), 92-103.

De Houwer, Jan (2022). On the merits and challenges of treating conscious and unconscious thoughts and feelings as behavior. *PsychArchives.* https://doi.org/10.23668/psycharchives.5332 [accessed Jan 16, 2023].

De Houwer, Jan, Sean Hughes, and Dermot Barnes-Holmes (2017). Psychological engineering: A functional-cognitive perspective on applied psychology. *Journal of Applied Research in Memory and Cognition, 6,* 1-13.

Dennett, Daniel C. (1991). Real patterns. *Journal of Philosophy, 88,* 27-51.

Égré, Paul (2021). Logical omniscience. In D. Gutzmann, L. Matthewson, C. Meier, H. Rullmann, and T. Zimmermann, eds., *The Wiley Blackwell companion to semantics.* Wiley-Blackwell.

Field, Hartry (1978). Mental representation. *Erkenntnis, 13,* 9-61.

Fodor, Jerry A. (1981). *Representations.* MIT Press.

Fodor, Jerry A. (1987). *Psychosemantics.* MIT Press.

Freud, Sigmund (1914/2012). *On narcissism.* Ed. J. Sandler, E. S. Person, and P. Fonagy. Karnac Books.

Hunter, David (2009). Guidance and Belief. *Canadian Journal of Philosophy, 35 supp.,* 63-90.

Hunter, David (2022). *On believing.* Oxford University Press.

Jung, C. G. (1915). On psychological understanding. *Journal of Abnormal Psychology, 9,* 385-399.

Lem, Stanislaw (1961/1970). *Solaris.* Trans. J. Kilmartin and S. Cox. Harcourt.

Lenat, Douglas B., and Chris Deaton (2008). *Terrorism Knowledge Base (TKB).* https://www.researchgate.net/publication/235042082 [accessed Jan. 16, 2023].

Lenat, Douglas B., et al. (2010). Harnessing Cyc to Answer Clinical Researchers' Ad Hoc Queries. AI Magazine, 31 (3), 13-32.

Lewis, David K. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy*, *50*, 249–258.

Lycan, William G. (1986). Tacit belief. In R. Bogdan, ed., *Belief: Form, content, and function.* Oxford University Press.

Mandelbaum, Eric (2019). Troubles with Bayesianism: An introduction to the psychological immune system. *Mind & Language, 34,* 141-157.

Manfredi, Pat A. (1993). Tacit beliefs and other doxastic attitudes. *Philosophia, 22,* 95-117.

Marcus, Ruth Barcan (1990). Some revisionary proposals about belief and believing. *Philosophy and Phenomenological Research, 50* (supplement), 133-153.

Nichols, Shaun, and Stephen Stich (2000). A cognitive theory of pretense. *Cognition, 74* (2), 115-147.

Pober, Jeremy (2022). *Reduce them all: A theory of psychophysical reduction for all mental phenomena.* PhD dissertation, University of California, Riverside.

Porot, Nicolas, and Eric Mandelbaum (2023). The science of belief: A progress report. In J. Musolino, J. Sommer, and P. Hemmer, eds., *The cognitive science of belief.* Cambridge University Press.

Quilty-Dunn, Jake, and Eric Mandelbaum (2018). Against dispositionalism: Belief in cognitive science. *Philosophical Studies, 175,* 2353-2372.

Quilty-Dunn, Jake, Nicolas Porot, and Eric Mandelbaum (2022). The best game in town: The re-emergence of the Language of Thought Hypothesis across the cognitive sciences. *Behavioral and Brain Sciences*, 1-55. doi:10.1017/S0140525X22002849

Schwitzgebel, Eric (1999). Gradual belief change in children. *Human Development, 42,* 283-296.

Schwitzgebel, Eric (2001). In-between believing. *Philosophical Quarterly, 51,* 76-82.

Schwitzgebel, Eric (2002). A phenomenal, dispositional account of belief. *Noûs, 36,* 249-275.

Schwitzgebel, Eric (2010). Acting contrary to our professed beliefs, or the gulf between occurrent judgment and dispositional belief. *Pacific Philosophical Quarterly, 91,* 531-553.

Schwitzgebel, Eric (2021). The pragmatic metaphysics of belief. In C. Borgoni, D. Kindermann, and A. Onofri, eds., *The Fragmented Mind*. Oxford University Press.

Stalnaker, Robert (1991). The problem of logical omniscience, I. *Synthese, 89,* 425-440.

Wandell, Brian A., and Jonathan Winawer (2011). Imaging retinotopic maps in the human brain. *Vision Research, 51* (7), 718-737.

Watts, Peter (2006/2020). *Blindsight.* Tom Doherty Associates.