

Dispositionalism, Yay!

Eric Schwitzgebel (2023)

Philosophy of Mind & Cognitive Science
Pascal Ludwig — Sorbonne Université

Recall: Fodor's Language of Thought

Last week we examined Fodor's LOT hypothesis. Its central claims:

1. Thinking occurs in an internal representational system with language-like combinatorial syntax and semantics.
2. Propositional attitudes (beliefs, desires) are computational relations to sentences in this mental language.
3. Beliefs are discrete representational tokens stored in a 'belief box' and causally active in cognitive processes.

Today's question: Is this picture of belief correct? Must believing that P involve storing a discrete representation with the content P?

Two Questions About Belief

The metaphysical question

What constitutes believing that P? What makes it true that someone believes something?

The architectural question

What cognitive structures and mechanisms underlie belief in creatures like us?

Schwitzgebel's key move: these two questions have different answers. Even if the brain stores representations (architecture), belief is constituted by the dispositional profile, not by the representation itself.

Liberal Dispositionalism

To believe that P is to have a suite of dispositions characteristic of believing that P.

Three kinds of dispositions:

Behavioural — acting as if P, choosing accordingly.

Phenomenal — feeling surprise if not-P, having relevant imagery, experiencing confidence.

Cognitive — drawing inferences from P, using P as a premise in reasoning.

‘Liberal’: the view makes no commitment to any particular cognitive architecture. Anything with the right dispositional profile believes — regardless of what underlies that profile.

The Target: Industrial-Strength Representationalism

Schwitzgebel characterises the strongest form of representationalism by four conditions:

Presence. There exists a concrete, persisting representational state constituting the belief.

Discreteness. The representation is a discrete, separable token — not a distributed or holistic pattern.

Kinematics. The representation is causally involved in cognitive processes characteristic of holding the belief.

Specificity. The content of the representation is specific: it encodes the particular proposition believed.

Fodor's LOT satisfies all four. This is the picture Schwitzgebel aims to undermine.

Argument 1: Excessive Representations

You believe there is beer in the fridge. But you also believe:

- There is something to drink in the fridge.
- There is a liquid in the fridge.
- The fridge is not empty.
- There are fewer than a million beers in the fridge.

Does each correspond to a separate stored representation? The representationalist faces a dilemma: either posit an explosive proliferation of discrete tokens, or draw an arbitrary line between 'core' beliefs (explicitly stored) and 'tacit' beliefs (merely derivable).

Argument 2: The Core / Tacit Pseudo-Distinction

The representationalist's escape route: distinguish core beliefs (explicitly stored) from tacit beliefs (derivable but not stored).

Schwitzgebel's four charges:

1. No practical importance: it plays no role in ordinary belief ascription.
2. No ordinary mapping: it does not track how we ordinarily attribute beliefs.
3. No empirical support: we have no evidence and likely cannot test for it.
4. A pseudo-distinction: a fake difference posited when a simple theory meets an unsimple world.

Argument 3: Indiscrete Belief — Implicit Bias

Consider a spectrum of racist belief, from whole-hearted white supremacism to whole-hearted egalitarianism.

On the dispositionalist view, racist belief is constituted by a suite of dispositions — behavioural, phenomenal, cognitive. As we move along the spectrum, dispositions fade, become inconsistent, are replaced by egalitarian dispositions. Belief ascription tracks this gradual shift.

There is no moment at which the belief ‘pops’ from present to absent. Belief is fuzzy-bordered, like a personality trait.

The representationalist must locate a discrete representation that either is or is not present. But implicit bias is not discrete — it is a matter of degree.

Argument 3 (cont.): Gradual Forgetting

The Mengzi example. A student believes that Mengzi debated Gaozi in passage 6A1. Over decades the memory fades across multiple dimensions: the name, the content, the location, the conditions under which it can be recalled.

Representationalist response: the representation persists but retrieval performance declines.

Schwitzgebel's reply:

- Unless no representations are ever lost, there must be a moment when it disappears. When?
- If someone sincerely professes ignorance and typically cannot recall the proposition, it strains credulity to say they still believe it merely because some trace persists.

Even if memory works by discrete storage, belief should be characterised by the dispositional pattern.

The AI Argument: Do LLMs Believe?

Schwitzgebel's arguments gain a new dimension when applied to recent AI — especially large language models (GPT-4, Claude, Llama, etc.).

The observation:

LLMs produce outputs that closely match the dispositional profile of belief. They assert that Paris is the capital of France, draw inferences from it, correct denials, and integrate it consistently across contexts.

The architectural fact:

LLMs store no discrete representation with the content 'Paris is the capital of France'. Their 'knowledge' is distributed across billions of parameters in a high-dimensional weight space. Nothing in their architecture satisfies Presence, Discreteness, or Specificity.

The AI Argument: The Representationalist's Dilemma

The representationalist faces a concrete dilemma — not science fiction, but engineering fact:

Option 1. Deny that LLMs have anything like beliefs. But this requires explaining away a system that passes every behavioural and cognitive test we associate with believing that P. The denial looks increasingly stipulative.

Option 2. Accept that LLMs have beliefs (or belief-like states). But then discrete stored representations are not necessary for belief, and the core of industrial-strength representationalism collapses.

Dispositionalism handles this naturally: if the dispositional profile is right, the system believes. The underlying architecture is secondary.

The AI Argument: Connecting the Threads

LLMs make Schwitzgebel's arguments vivid in a way that pure thought experiments cannot:

Excessive representations. An LLM answers questions about indefinitely many consequences of 'there is beer in the fridge' without storing any of them discretely. It generates appropriate responses on the fly — as dispositionalism predicts.

Indiscrete belief. LLMs exhibit graded confidence and partial 'knowledge'. A model may produce correct answers in some contexts but not others, with no sharp boundary — mirroring the spectrum cases.

Core vs. tacit. The distinction has no purchase in an architecture where all outputs are generated from the same undifferentiated parameter space. There is nothing it is for one 'belief' to be stored and another merely derivable.

The Alien Thought Experiment

Space aliens arrive. They learn our languages, tell stories, make discoveries, express surprise, feel doubt. They match the full dispositional profile of believers.

Schwitzgebel: if we know all this, we know they have beliefs. We need not know anything about their inner architecture.

The representationalist's dilemma:

Horn 1. Agree the aliens believe. But then the dispositional profile suffices regardless of architecture, and representationalism loses its empirical content.

Horn 2. Say belief depends on further architectural facts. But this redefines belief in terms of something we do not and should not care about.

Note the parallel with the LLM case: a concrete, non-fictional system that satisfies the dispositional profile without the representationalist's architecture.

Superficialist Minimalism

Schwitzgebel's positive conclusion combines two ideas:

Superficialism. Belief is present if the 'surface' is right — the pattern of behavioural, phenomenal, and cognitive dispositions — regardless of what transpires behind the surface.

Minimalism. The account attributes to believers exactly the properties we care about in belief ascription, and nothing more. It avoids the puzzles generated by over-commitment to a specific architecture.

Analogy: personality psychology studies extraversion without positing a causally efficacious inner switch set to 'E'. We can study belief without positing stored sentences in a mental language.

Reconciliation with Representationalism?

Schwitzgebel ends on a conciliatory note:

Since dispositionalism is silent about underlying structure, an underlying representationalist architecture is entirely consistent with it.

The reconciliation:

Representationalism might succeed as an account of the architecture of cognition (in some creatures), while dispositionalism captures what makes that architecture count as the architecture of belief.

LOT may describe the computational structure of the human mind. But what makes a state a belief is not its being a sentence in mentalese — it is the dispositional role that state plays.

Discussion

1. Does dispositionalism adequately distinguish belief from other attitudes (desire, hope, fear)? Is the dispositional profile specific enough?
2. Is the LLM case genuinely parallel to the alien case? What role does phenomenology play? Can a system without phenomenal dispositions count as a believer?
3. Quilty-Dunn & Mandelbaum accuse dispositionalism of being unscientific. Is this charge fair?
4. If representationalism describes architecture and dispositionalism characterises belief, what explanatory work does 'belief' do that the architecture does not?

Against Dispositionalism: Belief in Cognitive Science

Jake Quilty-Dunn & Eric Mandelbaum (2018)

Philosophy of Mind & Cognitive Science
A critical response to Schwitzgebel's dispositionalism

The Paper's Aim

Quilty-Dunn & Mandelbaum (henceforth QD&M) argue that dispositionalism about belief is explanatorily inadequate. They defend a psychofunctionalist, representationalist alternative.

Two-part strategy:

1. Schwitzgebel's arguments against representationalism rest on a caricature of the belief box. A fragmented representationalist architecture can handle all the 'in-between' cases.
2. Cognitive science provides a wealth of generalisations about belief — concerning causation, sorting, opacity, dissonance, inference — that are explicable in terms of mental representations but not in terms of dispositions.

Central claim: the metaphysics of belief should be determined by the posits that figure in successful explanations. If those posits are representations, we are obligated to endorse a representationalist metaphysics.

Reinterpreting the Belief Box: Fragmentation

Schwitzgebel's 'in-between' cases (implicit bias, gradual forgetting, self-deception, contextual access failures) are meant to show that representationalism cannot handle graded or conflicting beliefs.

QD&M's reply:

The belief box metaphor is misleading. It was never meant to denote a single undifferentiated store. Representationalism is compatible with the most simple and the most byzantine memory architectures.

Their preferred architecture: belief fragmentation (Mandelbaum 2016). Beliefs are stored in disparate, mutually inconsistent fragments. The implicit racist stores egalitarian and racist beliefs in distinct fragments, accessed in different contexts.

This is not ad hoc: fragmentation is independently motivated by data on contradictory beliefs, implicit bias, and the automaticity of belief acquisition.

The Causation Problem

A desideratum: beliefs cause behaviour. QD&M argue that dispositionalism faces a principled difficulty here.

The problem: dispositions only cause outcomes when combined with an event (fragility does not break the glass; the impact does). A disposition to behave will not cause behaviour without a mental event — such as the activation of a representation.

Dilemma for the dispositionalist:

Horn 1. The categorical bases (not the dispositions) do the causal work. But if those bases are neural states, we get an unprojectible disjunction. If they are psychological states, those states are representations — and the theory is no longer dispositionalist.

Horn 2. Deny that beliefs are causal (Schwitzgebel's superficialism). But Schwitzgebel himself accepts that beliefs serve as premises in inferences, and inference is a causal process.

Sorting and Representational Structure

A classic case (Bruner et al. 1956): subjects asked to sort ‘spades and clubs’ are faster and more accurate than subjects asked to sort ‘non-diamonds and non-hearts’ — even though the two tasks are extensionally equivalent.

The representationalist explanation:

The representations SORT THE SPADES AND CLUBS and SORT THE NON-DIAMONDS AND NON-HEARTS are extensionally equivalent but syntactically distinct. The latter includes negation operators, which increase processing load. The structural features of representational vehicles explain the behavioural difference.

The dispositionalist’s problem:

If belief is nothing more than the dispositional profile, then two extensionally equivalent beliefs should yield the same dispositions. But they do not. The vehicles of representation dictate how one is disposed to behave, so representations cannot drop out of the picture.

Opacity and Truth Evaluability

Two core features of belief that representationalism explains naturally:

Opacity. You can believe X is F without believing Y is F , even if $X = Y$ (Frege cases). The representationalist explains this: one can have two representations with the same referent without knowing they co-refer. The dispositionalist takes opacity for granted rather than explaining it — differences in representational form generate opacity, while differences in dispositions are merely a consequence of it.

Truth evaluability. Beliefs are true or false. Sentence-like representations are truth-evaluable. It is unclear in what sense dispositions are truth-evaluable. The dispositionalist can stipulate that dispositions constituting beliefs are ipso facto truth-evaluable, but the representationalist provides a deeper explanation.

Dissonance and Psychofunctionalism

QD&M's positive proposal: beliefs are relations to mental representations; the relations are determined by law-like generalisations uncovered in cognitive science (psychofunctionalism).

Central exhibit: cognitive dissonance theory (Festinger 1957).

- Encountering disconfirming evidence for a belief generates a negatively valenced state (dissonance).
- Agents are motivated to escape this state, often by non-rational routes (self-justification, projection, attitude change).
- The mechanism works best when unconscious (anterograde amnesiacs show larger effects).
- Beliefs left unexercised increase in strength over time (the sleeper effect); beliefs that are repeatedly tokened also increase in strength.

These are deep, counterintuitive generalisations about belief as a state — a state with laws about what it causes and what causes it to change. Dispositionalism, committed to the surface, cannot explain them.

Inference and Compositionality

Representationalism about belief is part of a broader representational theory of mind. QD&M argue that evidence for the RTM also supports representationalism about belief specifically.

Semantic priming. Reading 'doctor' facilitates recognition of 'nurse' (Meyer & Schvaneveldt 1971). This is explained by representations stored in associative networks: activating one representation activates connected ones.

Unconscious inference. Subjects presented with 'If there is a 3 then there is an 8' and then subliminally flashed '3' show facilitation for '8', but not the reverse (Reverberi et al. 2012). Inference operates on representations in virtue of their syntactic, formally specifiable properties.

The dispositionalist has no account of why thought is compositional. Dispositions do not have constituents, and it is unclear why they could compose at all, let alone in the specific ways mental representations do.

Taking Stock: Discussion

QD&M's conclusion: the science of belief demands a theory that is both deep and representationalist. Dispositionalism, in virtue of its superficialism, cannot explain the generalisations.

Questions for discussion:

1. Does fragmentation save representationalism, or does it concede too much? If beliefs are stored in disparate, inconsistent fragments, how different is this from a dispositional profile with conflicting elements?
2. Schwitzgebel accepts that representationalism might describe cognitive architecture while dispositionalism characterises belief. Does the QD&M critique successfully block this reconciliation?
3. Do the dissonance and sorting data show that beliefs are representations, or merely that representations figure in the causal mechanisms underlying belief?
4. How does the LLM case bear on QD&M's arguments? LLMs exhibit sorting effects, priming, and inferential promiscuity — but without discrete stored representations.