Received: 21 July 2024 | Revised: 16 August 2024 | Accepted: 2 September 2024

DOI: 10.1111/mila.12527

WILEY

SYMPOSIUM

Comments on Favela and Machery's *The concept of representation in the brain sciences: The current status and ways forward*

Frances Egan

Department of Philosophy, Rutgers University, New Brunswick, New Jersey, USA

Correspondence
Frances Egan, Department of Philosophy, Rutgers University, 106 Somerset St., New Brunswick, NJ 08901, USA.
Email: fegan@rutgers.edu

Favela and Machery conclude from their studies that neuroscientists' and psychologists' concept of representation is both *unclear* and *confused*. Rather than advocating reform or elimination of the concept, they suggest that it can serve various theoretical purposes precisely *because* it is unclear and confused. I challenge their claim that the concept of representation, as used by neuroscientists and psychologists, is unclear and confused, and I propose an alternative explanation of why it might appear to be so.

KEYWORDS
intentional content, natural signs, neuroscience, psychology, representation

1 | IS THE SCIENTIFIC CONCEPT OF REPRESENTATION UNCLEAR AND CONFUSED?

Favela and Machery (hereafter F&M) argue that neuroscientists' and psychologists' concept of representation is both unclear and confused. A concept is *unclear* just in case what is required for the concept to apply and what follows from the application of the concept are indeterminate. F&M cite as evidence for the unclarity of the concept of representation their studies showing that neuroscientists and psychologists are indifferent to the *scale* at which the concept applies, to how *specific* the eliciting stimulus must be, and to whether the representation is characterized, explicitly, as serving a *function*.

The results suggest that scientists are indifferent to the implications of a representational attribution regarding questions of scale, specificity, and function or use. F&M interpret their indifference as uncertainty about whether the concept of representation applies, concluding in their earlier paper that sets out their experimental results in detail that "neuroscientists and psychologists do not appear to have a precise idea about what kind of brain structure or pattern counts as representation" (Favela & Machery, 2023, p. 10).

The consequences of such imprecision for theorizing about cognition are potentially serious:

> [Such uncertainty] could breed fruitless debates about whether or not some brain part that responds to some stimulus represents it; barring a clearer concept of representation, such debates cannot be resolved ... imprecision of the concept of representation could prevent neuroscientists from interpreting some experimental results univocally. fMRI adaptation, multi-voxel pattern analysis (MVPA), representational similarity analysis, and others are supposed to determine what kind of representations the brain produces and where. If the concept of representation at play is genuinely imprecise, then it is hard to say what such methods reveal about the brain. (Favela & Machery, 2023, p. 11)

Consequently, in their original paper F&M advocate *reform*, that is, precisification of the concept of representation, or outright *elimination*. In their more recent paper they disavow the need for reform or elimination.

But given the widespread use of representation talk among neuroscientists and psychologists, F&M's interpretation of the survey's results is nonetheless pretty damning. Do neuroscientists and psychologists really not know what their use of the representational idiom commits them to? I suggest we look for a more charitable interpretation of these results, and a credible one is not hard to find.

I see no reason to think that neuroscientists and psychologists are uncertain about the scale at which representations are to be found, or about how specific the response to stimuli must be, or about whether a representation has a downstream function.[1] An explanatory theory of a cognitive capacity that appeals to representation must ultimately answer such questions. But there is no reason to think that there are general answers to these questions or that answers will follow from the application of the concept of representation itself. Rather these questions suggest parameters to be specified by a theory that characterizes the processes it explicates in representational terms. Neuroscientists know exactly what they mean when they characterize large scale structures in visual cortex as representing spatial properties in the world, or cells in the hippocampus as representing elements of their receptive fields. Their apparent indifference to such questions in the survey is unlikely to reflect uncertainty about the application of their concept; more likely it shows simply that appeal to representation alone does not determine specific answers to these questions. Answers will depend upon the details of the individual case. The right conclusion to draw is that their concept of representation has wide application.

A concept is *confused*, according to F&M, just in case it refers to two (or more) different phenomena; in other words, it is ambiguous. They appeal to (1) their study showing neuroscientists'

[1]F&M describe this last finding as follows: "When it comes to representation, we found no evidence that having a function matters" (p. 8). This finding seems to undermine claims by teleosemanticists, following Millikan (1984), that teleological function determines the attribution of representational content in the cognitive sciences. See Neander (2017) and Shea (2018) for the most developed recent teleosemantic proposals.

and psychologists' preference for "thin" causal descriptions ("responds", "processes") or information-theoretic descriptions ("carries information about") of the brain's response to stimuli to intentional descriptions ("represents", "is about"), and (2) the study showing neuroscientists' and to a lesser extent psychologists' seeming unwillingness to characterize a brain state as *misrepresenting*. They conclude from the two studies that scientists do not distinguish natural signs from genuine representations, and hence that their concept of representation is confused, applying indifferently to two quite different kinds of phenomena.

Here also I think an alternative interpretation of the results is warranted. Scientists' concept of representation is not ambiguous, or imprecise and confused. The results suggest rather that they have the thin, causal notion characteristic of natural signs in mind when they use representational talk. To cite a famous example,[2] to explain the toad's capacity to recognize and capture prey, the neuro-ethologist must (1) isolate the neural structures that play the appropriate causal role in mediating the prey-recognition process that eventuates (when things go well) in the toad's tongue lashing out and catching a bug, and (2) specify the conditions of these structures' activation—roughly, a *moving worm-like stimulus nearby*. These two tasks can be characterized in representational terms: (a), corresponding to (1), identifying the neural structures that serve as the representational *vehicle*, and (b), corresponding to (2), ascertaining their *content*, but, as I argue in Egan (2022), characterizing the neural structures that play the appropriate causal role in mediating the prey-recognition process—T5-2 cells in the optic tectum, as it happens—as *representations* with the content *moving worm-like stimuli at location x* is just a convenient way of describing the response profiles of the structures whose activation is a crucial part of the causal process.

Importantly, once the theorist has fully specified the structure's role in the process the theoretical heavy lifting is done. Talk of the cell's activation *representing* its distal stimulus conditions is best construed as a *gloss* that adds nothing of theoretical significance.[3] Philosophers are likely to see full-blooded (intentional) representation here. So-called "personal-level" thought processes are typically characterized in terms of their contents and so ascribing content to "subpersonal" processes posited by neuroscientists provides a common way of thinking about the two. Philosophers interested in cognition more generally will look for continuity with higher-level intentional thought. But the neuro-ethologist tasked with providing a causal/mechanical explanation of the toad's visual mechanism has no such motivation. Everything she might want to say about toad prey-recognition, for example, can be expressed more directly in straightforwardly causal terms, deploying a "thin" non-intentional construal of the representational idiom.

It is significant that the neuroscientists (and to a lesser extent the psychologists) in F&M's study do not deny that the intentional notion of representation applies. Rather they chose "neither agree nor disagree", where this result does not reflect a bi-modal distribution. What would explain this? Here is a speculation: Their response may be an expression of disinterest, the verbal equivalent of a shrug: They are not unaware of the "thick" intentional notion, nor are they explicitly denying that there may be grounds for its application; rather they are, in effect, saying "that's not *my* notion", denying its relevance for their own theoretical projects, which are well-served by a "thin" causal notion.

[2]See Neander (2017) for a clear explication of this research. Neander herself concludes that the research supports the attribution of full-blooded (intentional) representation. See Egan (2022) for criticism.

[3]See Egan (2020, 2025) for the view that representational talk is best construed as a gloss serving various pragmatic purposes in computational psychology and neuroscience.

One might insist that the thin, causal notion deployed by scientists is not really a concept of genuine *representation* at all—that it is just the notion of a *natural sign* under another name. One might then go on to argue that in the interests of transparency they should stop using representationalist talk. I agree with F&M's claim in the target article that legitimate purposes are served by talking the talk, even if neuroscientists and psychologists do not walk the walk. More on that below. The important point for now is that scientists are not uncertain about the application of their concept, whether we choose to call it a concept of *representation* or not.

2 | WHY DO SCIENTISTS USE REPRESENTATIONAL TALK? F&M'S PROPOSAL

Since I do not think that scientists' use of representation is unclear and confused I agree with F&M that neither reform nor elimination of their concept is called for. The case for *reform* would require an account of the theoretical purposes to be served by a revised concept, but if scientists' purposes are well-served by their thin, causal notion then reform is not necessary. The case for *elimination* would require an argument that no useful theoretical purpose is served by any concept in the neighborhood.

Still, though, F&M's experimental results do raise a nagging question: Why do neuroscientists and psychologists persist in using representational talk if a purely causal (or information-theoretic) notion suffices for their theoretical purposes? F&M argue that for some theoretical purposes an unclear and confused concept may be useful; indeed, clarity and lack of confusion may sometimes be "detrimental" (p. 10). Citing Rheinberger's (2000) discussion of the concept of *gene* they suggest that confusion and lack of clarity in a concept may allow ideas and techniques to propagate across disciplinary boundaries, sometimes suggesting fruitful new hypotheses.

> Just like the concept of gene, the concept of representation is a cross-disciplinary concept: It is used in different fields, in particular in contemporary artificial intelligence (AI), neuroscience, and psychology. It currently plays prominent roles across these disciplines, particularly AI and neuroscience. (Favela & Machery, 2024, this issue)

F&M point out that research in cognitive neuroscience and on deep neural networks is often characterized in representational terms, allowing the comparison of dynamics across both domains, and facilitating importing hypotheses from AI to the study of visual processing.

> Whether or not such hypotheses turn out to be correct is an open empirical question, but for present purposes, what matters is that the concept of representation facilitates the comparison of two different systems—brains and artificial networks—at least in part due to the imprecision (and lack of clarity) of the concept's definition. (Favela & Machery, 2024, this issue)

The examples cited do support the claim that the notion of representation is employed in multiple disciplines—it is *common coin*—but it is the thin, causal notion that is employed. I have argued that this notion is not confused and unclear. In any event, F&M provide no argument that it is the (alleged) imprecision of the concept that enables the cross-disciplinary

connections. I suggest it is more likely that the thin causal notion is common coin precisely because it is relatively clear.

I do agree with F&M though when they say:

> [I]t is unjustified to assume that these research products provide grounds for assigning genuine representational status to the relevant states, with all the properties (e.g., functional ones) that representations have. (Favela & Machery, 2024, this issue)

In other words, there is no justification for thinking that theorists in the various disciplines deploying representational language are attributing full-blooded (intentional) representation.

3 | AN ALTERNATIVE EXPLANATION FOR SCIENTISTS' USE OF REPRESENTATIONAL TALK

There is a more plausible explanation for why neuroscientists and psychologists persist in using representational talk when they really mean *causes* or *carries information about*. The full-blooded (intentional) notion of representation is in widespread use. It is a central component of the commonsense picture of the world and our relation to it that we *represent* the world, and in doing so we typically get things right but are occasionally subject to error. Our representational capacities (our cognitive capacities more generally) are the explanatory targets for theorizing in neuroscience and psychology. These explananda are typically characterized in intentional terms, as, for example, an organism's *knowing the 3-D structure of the scene*, or *knowing the location of an object in view*. Representational talk is the "connective tissue" linking the subpersonal causal processes characterized in neuroscientific and psychological theory and the manifest personal-level, intentionally characterized, capacities that are the theory's explanatory target. To cite just one implication, though an important one, neuroscientists and psychologists, when writing their grant proposals, need to make clear that their research addresses these explananda. And so they use representational talk in articulating their theories.

It does not follow, however, that the full-blooded intentional notion of representation is doing the explanatory work in the theories themselves. Deploying their thin, causal notion neuroscientists and psychologists can *reconstruct* what is wanted—the distinction between a successful application and a mistake—without making use of intentional notions. Consider once again the neuro-ethologist's explanation of the toad's prey-capture capacity. T5-2 cells in the optic tectum are activated by a moving wormlike stimulus in the toad's visual field. Their activation in turn causes the toad's sticky tongue to lash out. Often enough in the toad's normal environment the moving wormlike stimulus is a bug (the successful cases). Occasionally it is not (an error). There is no appeal to *representation* in the explanation itself—the process is described in purely causal terms. But *glossing* the causal process in representational terms (as I like to put it) serves the purpose described above, that is, making clear that the theory addresses its intentionally characterized explanatory target.

DATA AVAILABILITY STATEMENT

There is no data available.

ORCID

Frances Egan https://orcid.org/0000-0002-5073-8340

REFERENCES

Egan, F. (2020). A deflationary account of mental representation. In J. Smortchkova, K. Dolega, & T. Schlicht (Eds.), *What are mental representations?* (pp. 26–54). Oxford University Press.

Egan, F. (2022). The elusive role of normal-proper function in cognitive science. *Philosophy and Phenomenological Research*, *105*, 468–475.

Egan, F. (2025). *Deflating mental representation: The Jean Nicod lectures*. MIT Press.

Favela, L. H., & Machery, E. (2023). Investigating the concept of representation in the neural and psychological sciences. *Frontiers in Psychology: Cognition*, *14*, 1165622. https://doi.org/10.3389/fpsyg.2023.1165622

Millikan, R. (1984). *Language, thought, and other biological categories*. MIT Press.

Neander, K. (2017). *A mark of the mental: In defense of informational teleosemantics*. MIT Press.

Rheinberger, H. J. (2000). Gene concepts: Fragments from the perspective of molecular biology. In P. Beurton, R. Falk, & H. J. Rheinberger (Eds.), *The concept of the gene in development and evolution: Historical and epistemological perspectives* (pp. 219–239). Cambridge University Press.

Shea, N. (2018). *Representation in cognitive science*. Oxford University Press.

How to cite this article: Egan, F. (2025). Comments on Favela and Machery's *The concept of representation in the brain sciences: The current status and ways forward*. *Mind & Language*, 1–6. https://doi.org/10.1111/mila.12527