

Draft July 14, 2022

The Untenable Status Quo: The Concept of Representation in the Neural and Psychological Sciences

Luis H. Favela ^{1,2} and Edouard Machery ^{3,4}


¹ Department of Philosophy, University of Central Florida

² Cognitive Sciences Program, University of Central Florida

³ Department of History and Philosophy of Science, University of Pittsburgh

⁴ Center for Philosophy of Science, University of Pittsburgh

Author Note

Luis H. Favela  <https://orcid.org/0000-0002-6434-959X>

Edouard Machery  <https://orcid.org/0000-0002-9944-6138>

Conflicts of interest: The authors declare no known conflict of interest to disclose.

Ethical approval statement: The study was approved by the Institutional Review Boards at the University of Central Florida (IRB STUDY00002612) and the University of Pittsburgh (IRB STUDY20050065).

Preregistration: Open Science Framework (OSF)
<<https://doi.org/10.17605/OSF.IO/MSKWY>>

Author Contributions: L.H.F. and E.M. contributed equally to the paper.

Correspondence concerning this article should be addressed to Luis H. Favela, Department of Philosophy, University of Central Florida, 4111 Pictor Lane, Building 99, Suite 220, Orlando, FL 32816-1352. E-mail: luis.favela@ucf.edu

Rights and permissions: Creative Commons Attribution 4.0 International License (CC BY 4.0 <<https://creativecommons.org/licenses/by/4.0/>>).

Acknowledgements

We thank our research assistants Maya Best and Dean Allen Walters, Jr. for their support with data collection. For providing helpful feedback on earlier versions of the experiment survey, we thank John Beggs, John Bickle, Tony Chemero, Stephen Fiore, and John Krakauer, the X-Phi lab at the University of Pittsburgh, and Fellows reading group at the Center for Philosophy of Science (Leornaro Bich, Ravit Doran, Heather Douglas, Eugen Fischer, Ruth Kastner, Laura Menatti, Aydin Mohseni, and Serife Tekin). We thank Mary Jean Amon for discussions concerning data analyses. We also thank audiences for feedback on the project from the Philosophy and Neuroscience at the Gulf IV: Fourth Annual Meeting of the Deep South Philosophy and Neuroscience Workgroup and the 58th Annual Meeting of the Alabama Philosophical Society, Neural Mechanisms Online 2022, and the 3rd Joint Conference of the Society for Philosophy and Psychology & European Society for Philosophy and Psychology.

The Untenable Status Quo: The Concept of Representation in the Neural and Psychological Sciences

Abstract

The concept of representation is commonly treated as indispensable to research on brains, behavior, and cognition. We argue that not only is the concept of representation applied with considerable imprecision in such research, but it appears to be used in unclear and confused ways. We present results of a preregistered experiment aimed at making explicit what researchers mean by “representation.” Participants consisted of an international group of psychologists, neuroscientists, and philosophers ($N = 736$). Applying elicitation methodology, participants responded via an online survey platform to four experimental scenarios aimed at invoking applications of various representational concepts. While we find very little disciplinary variation in the application of “representation” and related concepts (e.g., “about,” “carry information,” “processing,” etc.), we show that researchers exhibit confusion about what counts as a representation and are uncertain about what sorts of brain activity involve representations or not. We argue on this basis that the concept of representation—especially in the neural and psychological sciences—should be reformed or eliminated from use. Consequently, the theoretical status quo concerning the concept of representation endorsed by many neuroscientists and psychologists is untenable.

Keywords: representation, conceptual reform, information, scientific concepts

The Untenable Status Quo: The Concept of Representation in the Neural and Psychological Sciences

1. Introduction

The concept of representation is widely applied in research on brains, behavior, and cognition. Psychologists (especially cognitive psychologists) often investigate and explain mental capacities in terms of representations and the computations or operations that process them (e.g., Chomsky, 1980; Fodor, 1981; Anderson, 2015). Examples of this theoretical commitment abound, including research on *attitudes* (“attitudes ... can be conceptualized as mental representations that determine how we evaluate stimuli”; De Houwer, Van Dessel, & Moran, 2021, p. 870), *imagery* (“[d]o learners who understand a picture also construct multiple mental representations in their mind”; Schnotz, Hauck, & Schwartz, 2021, p. 4), and *language* (“language ... is in part shared among us and represented somehow in our minds”; Chomsky, 1980, p. 1). Likewise, the concept of representation is central to the neural sciences, especially the cognitive (e.g., Gazzaniga, Ivry, & Mangun, 2014), computational (e.g., Trappenberg, 2014), and sensory neurosciences (e.g., Reid & Usrey, 2013). A central theoretical commitment in this research is that brains form representations of the organism’s internal state (e.g., proprioceptive experiences) or external environment (e.g., speed and orientation of visual stimuli). Accordingly, neuroscientists commonly aim at identifying and characterizing these representations in order to answer questions such as: what do they represent, what are their vehicles, and how are they used (e.g., Kriegeskorte & Diedrichsen, 2019; Poldrack, 2021)? Following suit, philosophers of psychology and neuroscience have proposed various explications of the concept of representation, sometimes inspired by traditional philosophy of mind (e.g., Ramsey, 2007; Shea, 2018), sometimes by work on signaling (e.g., Planer & Godfrey-Smith, 2021), and sometimes by

the methods used by neuroscientists to identify neural representations, such as representational similarity analysis (e.g., Roskies, 2021). A minority—but increasingly vocal—group of psychologists (e.g., Richardson, Shockley, Fajen, Riley, & Turvey, 2008), neuroscientists (e.g., Buzsáki, 2019), and philosophers (e.g., Chemero, 2009; Hutto & Myin, 2013) disagree with this mainstream representationalism. They argue that the concept of representation need not be central, or even necessary, to investigate and explain brains, behavior, and cognition.

We argue that the theoretical status quo concerning the concept of representation endorsed by many neuroscientists and psychologists is untenable. The “status quo” refers to the apparently general acceptance in the neural and psychological sciences that the widespread—and typically unquestioned—use of the concept of representation means that the term is understood well enough to guide hypothesis development, experimental data interpretation, and explanation. We support our claim by providing evidence that the concept of representation in neuroscience and psychology is both unclear and confused. A concept is *unclear* if the concept user does not know what follows from applying it (e.g., “What follows if some brain pattern is a representation?”) and what must be the case for this concept to apply (e.g., “What properties should a brain pattern have to count as a representation?”). A concept is *confused* if it fails to distinguish two distinct phenomena. As a consequence, we conclude that “representation” should be either substantially reformed or eliminated from use.

Importantly, our main point does not rest on the naive view that a concept can only be appropriately used in scientific research if it is defined by a widely-accepted set of necessary and jointly sufficient conditions. Save for formal systems (e.g., logic and mathematics) or a handful of concepts (e.g., the concept of uncle), few concepts can be defined (Machery, 2009), particularly concepts of entities and processes in the natural world. As such, there is no doubt

that science progresses without defining all of its terms. Moreover, the absence of definitions can be viewed as an indispensable feature of research when scientists are attempting to characterize novel and interdisciplinary targets of investigation, as has been the case in the investigations of genes and viruses (e.g., Rheinberger, 2000). Neurophilosopher Patricia Churchland captures well this idea when she writes that to “force precision by grinding out premature definitions enlightens nobody” (1986, p. 346).

While the imprecision of *some* uses of the concept of representation in the neural and psychological sciences can certainly be understood this way, such instances are not what we draw attention to. Consider the following two recent examples from neuroscience. First, in an article on finger movement, the concept of representation is used in a variety of ways including in, but not limited to, contexts such as “different spatial representations,” “low-dimensional representation,” “n members can be represented at time t,” “schematic representation of behavioral mode segmentation,” “the cerebral cortex represents,” and “well-represented in neural state space” (Flint, Tate, Li, Templer, Rosenow, Pandarinath, & Slutzky, 2020). Second, an article on neural network models of symbolic cognitive processes and dynamical systems uses “representation” in an assortment of ways, such as, “agent’s internal representations of the environment,” “distributed representations,” “feature representation in deep learning,” “holographic reduced representations,” “neurobiological representations (i.e., grid cells),” and “structured symbolic representations” (Voelker, Blouw, Choo, Dumont, Stewart, & Eliasmith, 2021). It is not clear to us what “representation” means in all these instances and what would be required for something to be, for example, a “holographic reduced representation” or “represented in the neural state space.” How is a reader to understand if it is reasonable to ask if a structured symbolic representation (Voelker et al., 2021) can be well-represented in neural state

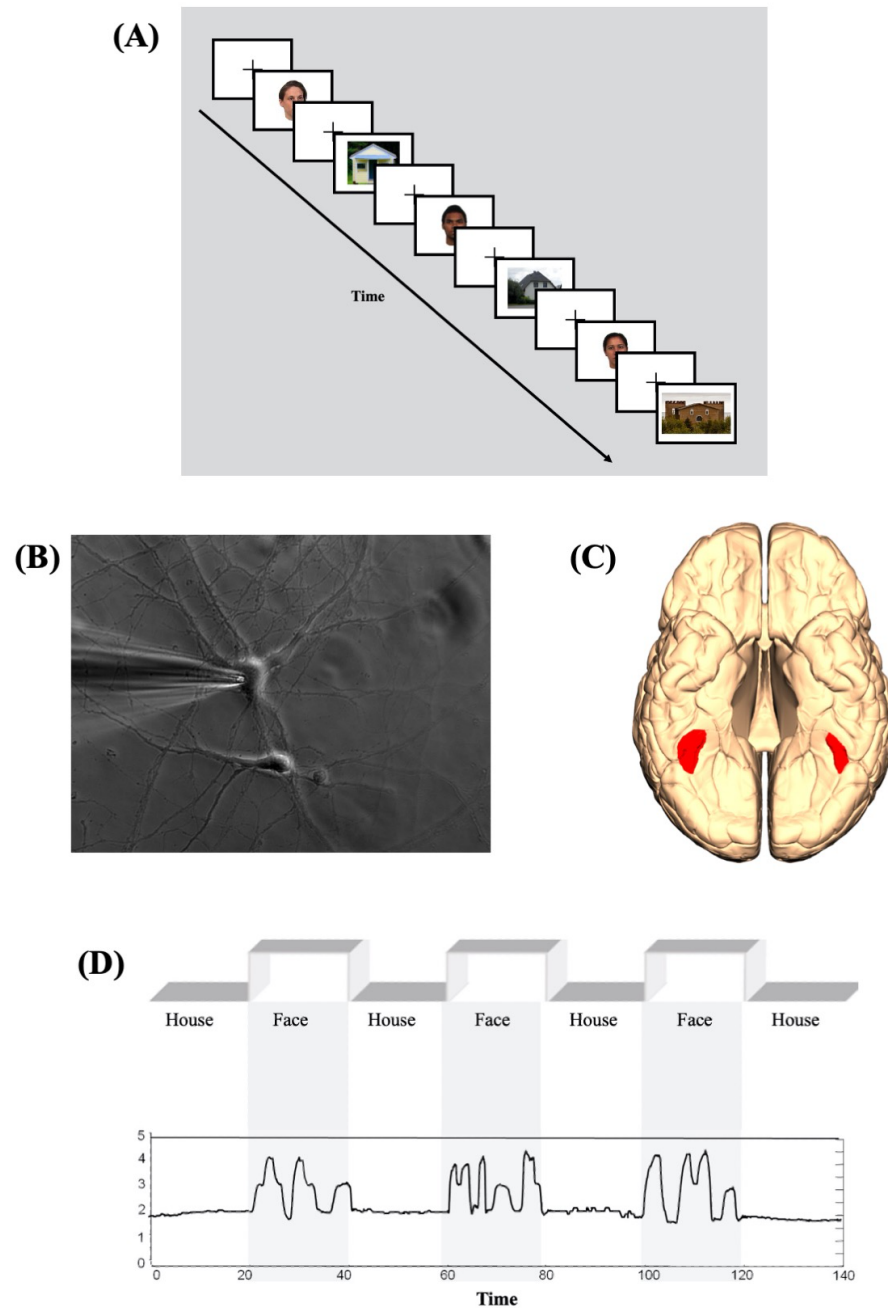
space (Flint et al., 2020)? These examples and the questions they provoke underlie our claim that the status quo is untenable. We do not think there are clearly defined and well-enough understood uses of the concept of representation in the neural and psychological sciences to guide hypothesis development, experimental data interpretation, and explanation. Note that our argument *will not depend* on a disagreement about how to define “representation” or on variation in how scientists understand this definition across disciplines. On the contrary (as our results demonstrate below), there is considerable agreement in the application of “representation” and related concepts, but researchers exhibit confusion about what counts as a representation and ambivalence about what sorts of brain activities are properly characterized as being representations or not.

A couple of handpicked examples is, however, unlikely to be fully convincing, but a literature review of the uses of “representation” in even one discipline would be a Herculean task. So, how can we empirically assess the current state of neuroscientists’ and psychologists’ understanding of the concept of representation? One option is to utilize what linguists call “elicitation studies” (Greenbaum & Quirk, 1970). Instead of asking scientists to reflect and report on their own concepts or examining the natural occurrences of a given concept (e.g., corpus study), scientists are asked to use the target concept and then the experimenter can make inferences about its content on the basis of subjects’ answers (Machery, 2017; Machery, Griffiths, Linquist, & Stotz, 2019). Inspired by the elicitation-study method, we conducted a survey-based experiment with an international group of neuroscientists, psychologists, and philosophers ($N = 736$). The experiment consisted of four studies with the same basic structure. Participants were given a cover story about a neuroscientific study measuring recorded brain reaction to various stimuli, including faces and artifacts (Figure 1). Participants were then asked

to provide a rating on a seven-point scale (“Strongly agree” to “Strongly disagree”) regarding six questions about whether they would agree to describing the brain’s activity as *representing*, *carrying information*, *being about*, *responding*, *processing*, and *identifying* the stimuli.

Figure 1

Sample Experimental Stimuli



Note. The experiment consisted in four studies of similar design, each with a cover story like the one associated with this figure: “In a study published about ten years ago, participants were presented with visual stimuli in a standard block design with alternating images of human faces

and houses (Figure A). Data were obtained via a microelectrode (Figure B) from single neurons in participants' fusiform face area (Figure C). An example of the time series data obtained during the task is presented in Figure D." (Modified and reprinted with permission from Michael J. Tarr <www.tarrlab.org/>. CC BY-NC-SA 3.0, PxHere. CC0 1.0, flickr. CC BY-SA 2.0 and CC BY 2.0 (A); Geissler, Gottschling, Aguado, Rauch, Wetzel, Hatt, & Faissner, 2013. CC BY-NC-SA 3.0 (B); Wikipedia. CC BY-SA 2.1 JP (C); Alkan, Biswal, & Alvarez, 2011. CC BY 4.0 (D).)

The goal of Study 1 was to examine whether neuroscientists, psychologists, and philosophers make any assumptions about the vehicle of neural representations, that is, the nature of the brain substrates that represent stimuli. Participants were randomly assigned to one of two conditions. In the neuron condition, they were told that the reaction of a neuron was measured by means of a microelectrode (visually represented) when presented with faces; in the population condition, they were told that the reaction of a whole brain area (i.e., the fusiform face area; FFA) was measured by means of functional magnetic resonance imaging (fMRI).

The goal of Study 2 was to examine what kind of relation, if any, must hold between the brain and stimuli for neuroscientists, psychologists, and philosophers to describe it in various terms. Participants were randomly assigned to one of two conditions. In the high sensitivity condition, a brain area, whose activity was measured by means of fMRI, reacted to faces and only to them; in the low sensitivity condition, it reacted to faces but also to houses.

The goal of Study 3 was to examine whether evidence that the brain's reaction to stimuli is used by a broader neural network, and thus has a function (Cummins, 1975), in addition to a perfect correlation with a stimulus increases neuroscientists, psychologists, and philosophers' willingness to treat the brain's reaction to stimuli in representational terms. In the mere

correlation condition, participants were just given evidence of the brain's reaction to the stimuli; in the function condition, the connection between the relevant brain area and a full network was highlighted verbally and by means of two figures.

Finally, the goal of Study 4 was to examine whether neuroscientists, psychologists, and philosophers are willing to describe the brain's reaction as erroneous, for example, whether it misrepresents stimuli. Philosophers concur that for a state to count as a representation, misrepresentation must be possible. Participants were assigned to a single condition where a brain area that responds to faces happen to also react, once, to a house.

These four studies focus on characteristics that brain states would have to possess if they are to count as representations. Representations must occur at some scale in brain organization (Study 1); the occurrence of representations must causally depend, in some way, on what they represent (Study 2); representations must be used by downstream processes (Study 3); and representations can be misapplied (Study 4). To have a clear concept of representation is to have a sense of the scale at which representations occur, of the nature of representations' causal dependence on what they represent, and on the significance of the use of representations; or at least to have some sense for some of these issues. To have a concept of representation that is not confused is to distinguish concepts for which misapplication matters and those for which it does not (Study 4). In what follows, we report results from each study, which support our claim that the concept of representation in neuroscience and psychology is both unclear and confused.

2. Methods

2.1. Participants

The study was approved by the Institutional Review Boards at the University of Central Florida (IRB STUDY00002612) and the University of Pittsburgh (IRB STUDY20050065).

Hypotheses, data collection methods including the stopping rule, exclusion criteria, and data analytic strategies were preregistered with the Open Science Framework (OSF;

[ANONYMIZED VERSION OF LINK]

https://osf.io/mskwy/?view_only=73cd03040cb74d90acbc93b907233acb).

Two research assistants were tasked to create a database of emails found on the public websites of departments, centers, institutes, and schools at universities around the world. A list of universities in Asia, Australia, Europe, North America, and South America was created and the research assistants were asked to input the names, emails, departmental affiliations of, cognitive scientists, computer scientists, linguists, neuroscientists, philosophers, and psychologists into a data file. Research assistants were ultimately asked to focus on cognitive scientists, neuroscientists, and psychologists in the United States, setting aside computer scientists and linguists as well as academics from abroad. 14,338 recruitment emails were sent, many of which were blocked by university servers. As was indicated in the preregistration, the study was also advertised on blogs, mailing lists, and social media.

736 participants completed the study. We excluded participants who reported being younger than 18, who were not a graduate student, postdoctoral researcher, professor, or researcher with a doctorate, who either did not respond or gave an incorrect answer to the last question of the survey, “Please tell us what this study was about,” and who provided the same numerical answer to questions in all four scenarios (in line with the preregistration). We also limited our analysis to neuroscientists, psychologists, and philosophers (Table 1), setting aside cognitive scientists in light of the small number of participants who self-identified as such and completed the study (52 before exclusion; a departure from the preregistration).

Table 1*Demographic Characteristics of Neuroscientists, Psychologists, and Philosophers*

Discipline	N	Gender	Age			Highest Degree	Location
		W, M, Other	Mean	SD	Range	BA, MA, PhD	USA, UK, Germany
Neuroscientists	177	39, 59, 2	38.3	13.3	22-77	23, 15, 62	88, .5, .6
Psychologists	159	50, 47, 3	36.7	14.6	21-92	15, 33, 52	87, .2, .3
Philosophers	184	14, 84, 2	43.0	14.4	22-87	2, 24, 74	56, 9, 5

2.2 Materials

The recruitment materials included a link to a survey on Qualtrics. Participants were first asked a few demographic questions before being asked to complete successively four studies in random order (described below). They were then asked several philosophical questions related to representation, computation, and their broader commitments related to the foundations of neuroscience and cognitive science (full survey available at the preregistration site: https://osf.io/mskwy/?view_only=73cd03040cb74d90acbc93b907233acb).

Each of the four studies had the same basic structure. Participants were given a cover story about a neuroscientific study measuring brain reaction to various stimuli, including faces and artifacts. A first figure represented the basic structure of the experimental design. Additional figures represented the data observed, including a time series. Participants were then asked six questions about whether they would agree to describing the brain's activity as representing the stimuli, carrying information about the stimuli, being about the stimuli, responding to the stimuli,

processing the stimuli, and identifying the stimuli (each on a 7-point scale anchored as “1” with “strongly agree”).

2.3. Data and analyses

The data are publicly available at the preregistration site (https://osf.io/mskwy/?view_only=73cd03040cb74d90acbc93b907233acb). All analyses were conducted on R (script available at the preregistration site: https://osf.io/mskwy/?view_only=73cd03040cb74d90acbc93b907233acb). As preregistered, the significance level was set at .005 (Benjamin et al., 2018). *P*-values between .05 and .005 are taken to be suggestive and in need for confirmation. All the analyses were redone with participants who had completed a Ph.D. The results did not change.

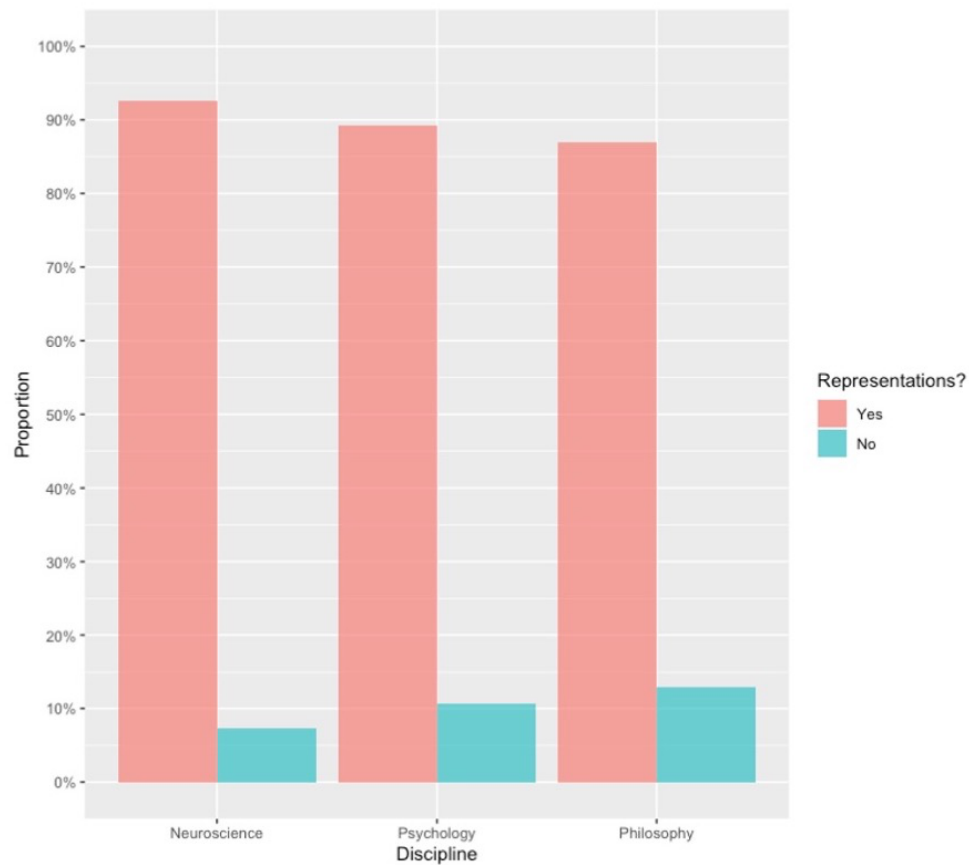
3. Results

3.1. Representationalism

Toward the end of the survey, participants were asked five questions aimed at elucidating positions on foundational issues concerning the nature of cognition. We begin by reporting results from a question probing their commitment to mainstream representationalism: “Does cognition involve representations? Yes or no.” We claimed at the start that representationalism—i.e., mental states involve computations acting on representations and that brains represent those states—is widely-accepted as being necessary to investigate and explain brains, behavior, and cognition. As expected, a very large majority of participants answered this question positively for the three disciplines of interest (Figure 2). It thus appears that mainstream representationalism is embraced by a large majority of psychologists, neuroscientists, and philosophers.

Figure 2

Proportion of “Yes” and “No” Answers to the Representation Question



Note. The overwhelming majority of neuroscientists, psychologists, and philosophers answered “yes” to the question, “Does cognition involve representations?”

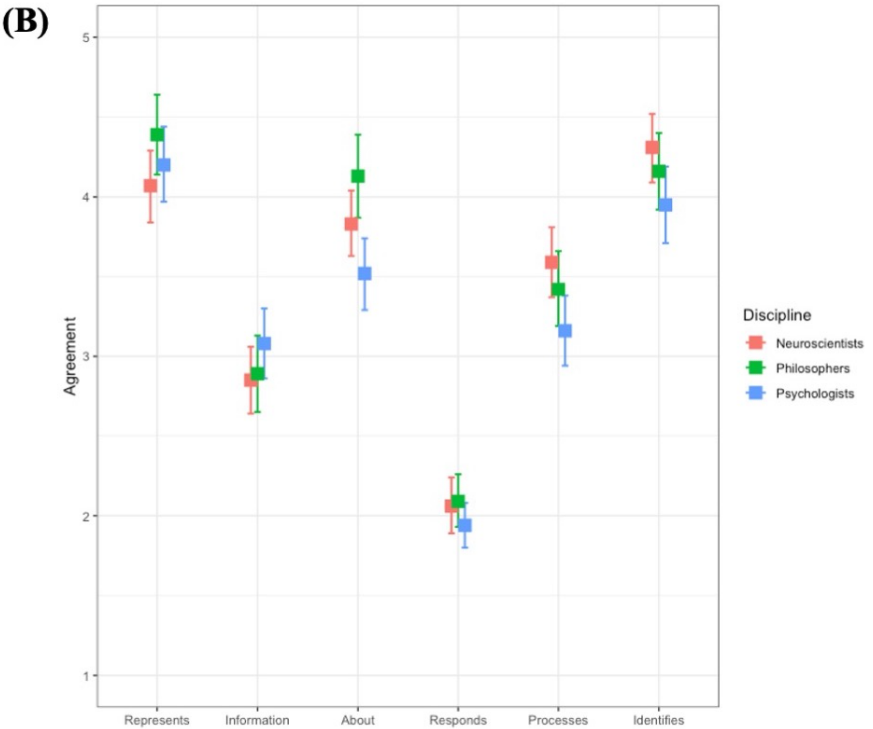
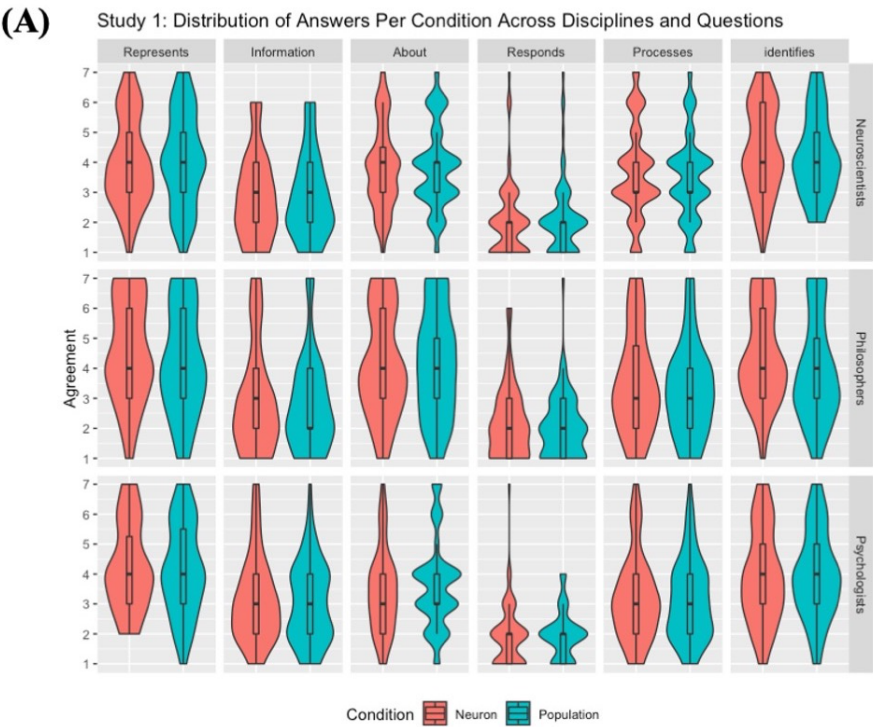
3.2. Study 1: Vehicles of representations

Distribution of responses are presented in Figure 3A. A mixed-design analysis of variance (ANOVA) with Questions as a within-participants factor (6 levels), Discipline as a between-participants factor (3 levels), and Condition as a between-participants factor (2 levels) revealed a main effect (Benjamin et al., 2018) of Question ($F(5, 3083) = 167.8, p < .001$), a

suggestive effect of Discipline ($F(2, 3083) = 5.0, p = .007$), and no effect of Condition ($F(1, 3083) = 3.5, p = .06$). Post hoc analysis revealed that the suggestive effect observed for Discipline is due to a suggestive difference between philosophers and psychologists ($t(3083) = 3.0, p = .007$); no other comparison reaches the .05 level. All post-hoc comparisons between the six questions used were significant except for the non-significant comparison between Represents and Identifies ($t(3083) = .9, p = .9$) and for the suggestive comparison between Is about and Identifies ($t(3083) = -3.3, p = .01$). The main effects of Discipline and Question were qualified by a suggestive two-way interaction ($F(10, 3083) = 2.4, p = .007$; Figure 3B).

Figure 3

Study 1: Vehicles of Representations



Note. Distribution of answers for Study 1 (1: “Strongly agree;” 7: “Strongly disagree”) (A).

Interaction of Question and Discipline in Study 1 (B). Error bars correspond to 95% confidence intervals.

In line of this interaction, an exploratory, not-preregistered mixed-design ANOVA with Question as a within-participants factor (6 levels) and Condition as a between-participants factor (2 levels) was conducted for neuroscientists and psychologists separately. For neuroscientists, we observed a main effect of Question ($F(5, 1049) = 63.5, p < .001$), no effect of Condition, and no interaction (both $ps > .7$). All post-hoc comparisons were significant except for the non-significant comparisons between Represents and Is about, Represents and Identifies, and Is about and Processes ($ps > .5$) and for the suggestive comparisons between Represents and Processes and Is about and Identifies ($ps > .01$). For psychologists, we observed a main effect of Question ($F(5, 941) = 53.9, p < .001$), no effect of Condition, and no interaction (both $ps > .8$). All post-hoc comparisons were significant except for the non-significant comparisons between Represents and Identifies, Is about and Processes, Carries information and Processes (all $ps > .15$), Is about and Carries information, and Is About and Identifies (respectively, $p = .055$ and $.057$).

Three main findings emerge from this first study. First, contrary to our first preregistered hypothesis, neuroscientists and psychologists do not treat all of the descriptions of the brain’s reaction to stimuli identically. Rather than being indifferent, they prefer a lean, causal characterization in terms of responding as well as a characterization in terms of processing. Perhaps more surprisingly, they tend to find an information-theoretic description of the brain’s reaction (carrying information about) to stimuli acceptable. By contrast, they seem to be much more ambivalent and uncertain about intentional characterizations. On average, they choose

“neither agree nor disagree” for “representing,” “identifying,” and “being about.” We will come back to this point in the general discussion below. Importantly, neuroscientists’ and psychologists’ overall ambivalence is not the result of a bimodal distribution, with half of the participants willing to strongly agree to using the concept of representation to describe the brain’s reaction to stimuli, and half of them strongly disagreeing. Rather, the distribution is centered around its mean. (The same is true of the three other studies.)

Second, it made very little difference to neuroscientists and psychologists whether the vehicle of representation was verbally and pictorially represented as a single neuron or as a population. This negative result suggests that neuroscientists and psychologists do not have any expectation about the scale at which representations are to be found in the brain: They may subscribe to the mainstream representationalism, but they have no clear idea about what kind of brain structure or pattern at what level of aggregation (neuron, population, distributed network of populations, etc.) would be a representation.

Third, while philosophers were somewhat less likely to agree with our prompts than psychologists, the variation across disciplines was small. This finding suggests that the concept of representation hasn’t specialized in the disciplines we are considering (see Machery et al., 2019 for discussion of similar results for the concept of innateness in psychology, biology, and linguistics).

3.3. Study 2: Sensitivity and representation

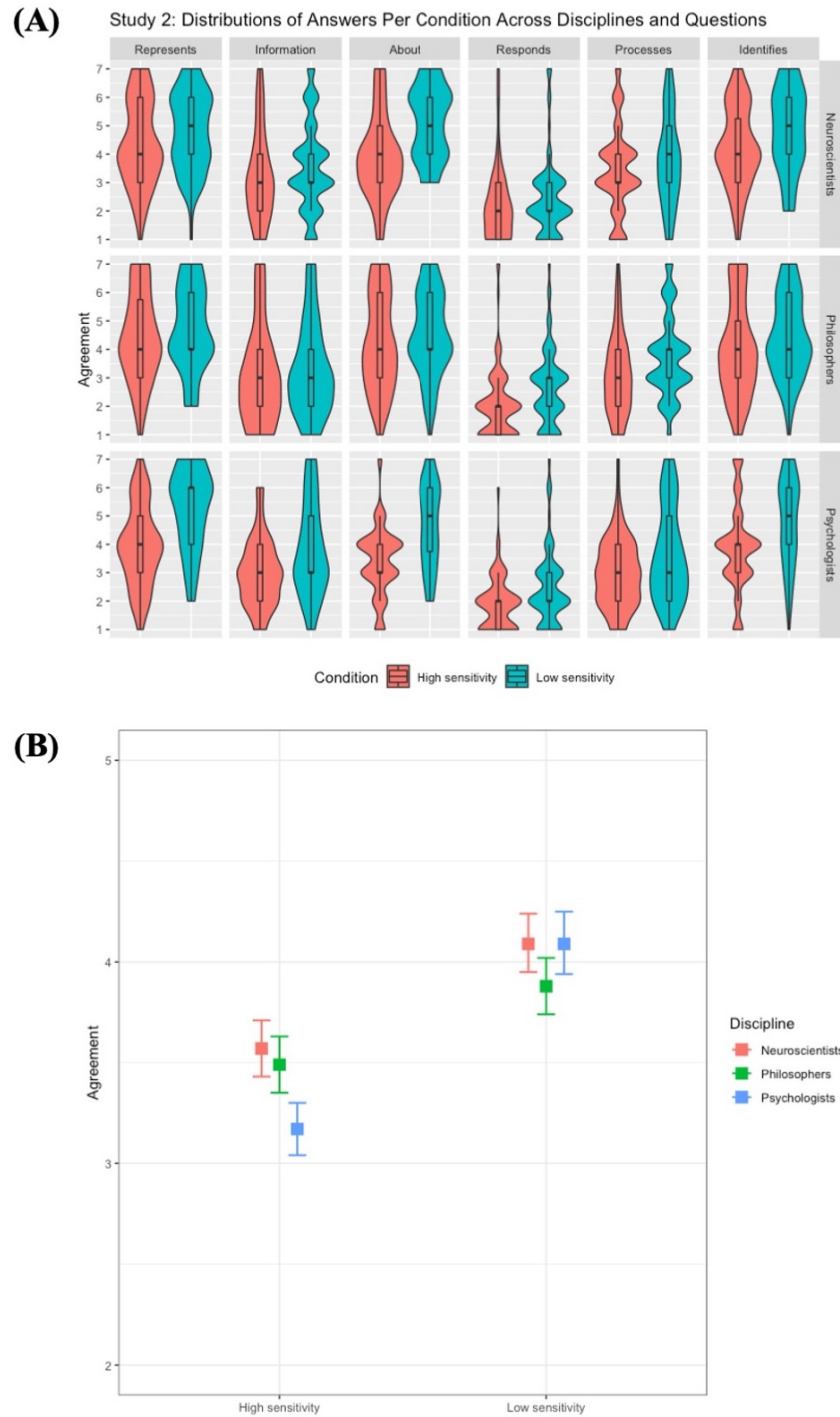
Distribution of responses are presented in Figure 4A. A mixed-design ANOVA with Question as a within-participants factor (6 levels), Discipline as a between-participants factor (3 levels), and Condition as a between-participants factor (2 levels) revealed a main effect of

Question ($F(5, 3083) = 191.4, p < .001$), a suggestive effect of Discipline ($F(2, 3083) = 4.9, p = .007$), and an effect of Condition ($F(1, 3083) = 131.3, p < .001$). Post hoc analysis revealed that the suggestive effect observed for Discipline is due to a difference between neuroscientists and philosophers ($t(3083) = 2.6, p = .03$) and neuroscientists and psychologists ($t(3083) = 3.0, p = .008$). All post-hoc comparisons between the six questions used were significant except for the non-significant comparisons between Represents and Identifies ($t(3083) = 2.2, p = .2$), between Carries information and Processes ($t(3083) = -1.0, p = .9$) and between Is about and Identifies ($t(3083) = -.7, p = .98$) and for the suggestive comparison between Represents and Is about ($t(3083) = 3.1, p = .03$). The main effects of Discipline and Condition were qualified by a two-way interaction ($F(10, 3083) = 9.0, p < .001$): Psychologists are more sensitive to the manipulation of sensitivity than philosophers and neuroscientists.

In addition, we explored the impact of sensitivity on representation alone (Figure 4B). For neuroscientists the impact of sensitivity on the description of the brain's reaction in terms of representation was too small to result in a significant or suggestive effect ($t(173.07) = -1.90; p = .059$); by contrast, we found a significant effect for psychologists ($t(155.12) = -5.7; p < .001$).

Figure 4

Study 2: Sensitivity and Representation



Note. Distribution of answers for Study 2 (1: “Strongly agree,” 7: “Strongly disagree”) (A).

Interaction of Question and Discipline in Study 2 (B). Error bars correspond to 95% confidence intervals.

Two main findings emerge from Study 2. First, as we observed in Study 1, neuroscientists and psychologists prefer thin, causal descriptions of the brain's reaction to stimuli (responds to and processes) and information-theoretic descriptions to intentional descriptions, and they are ambivalent about the latter. Second, sensitivity matters for describing how the brain reacts to stimuli (in line with the preregistered second hypothesis). When one aggregates across ways of describing the brain's reaction, neuroscientists, psychologists, and philosophers agree more (although to a different degree) when the brain's reaction is maximally sensitive. Turning to the concept of representation, we only found evidence for the significance of sensitivity for psychologists. It would thus seem that psychologists take sensitivity to be relevant to whether some brain state can count as a representation; neuroscientists might agree to a smaller degree, although we were unable to provide evidence for it. In any case, even perfect sensitivity does not lead neuroscientists and psychologists to abandon their ambivalence when it comes to describing the brain's reaction to stimuli in representational or, more generally, intentional terms.

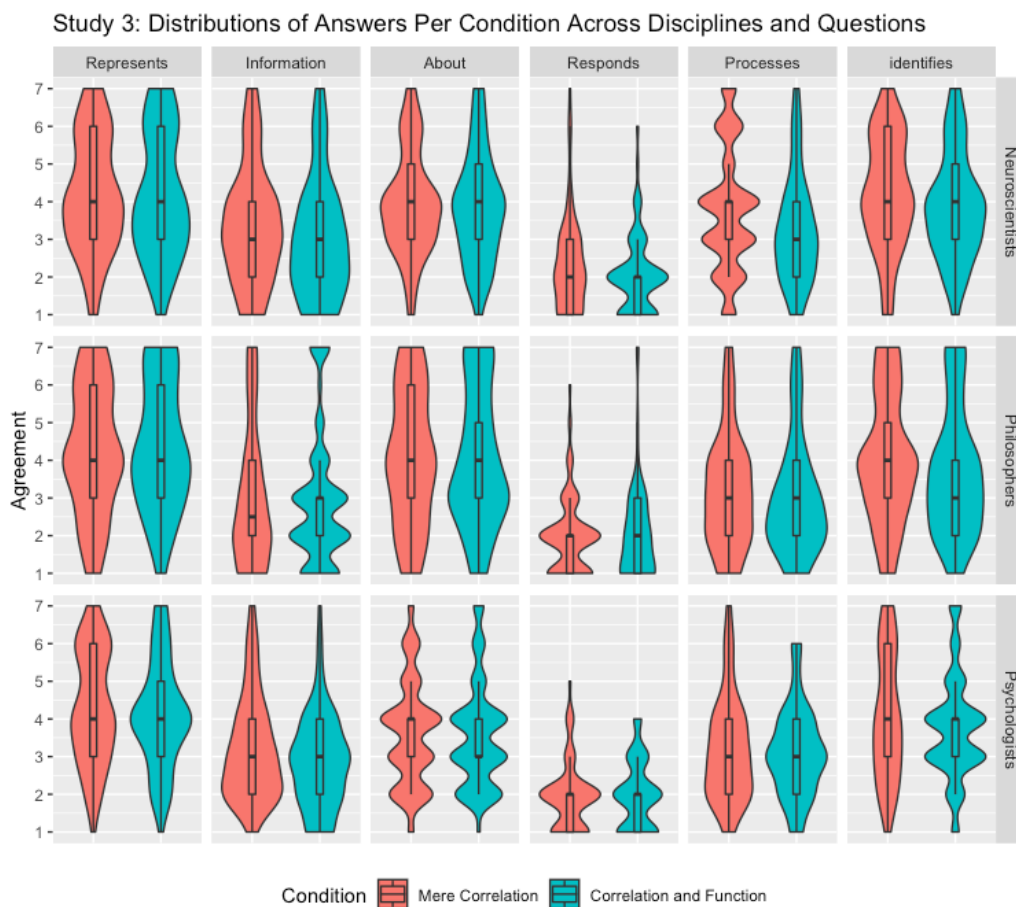
3.4. Study 3: Function and representation

Distribution of responses are presented in Figure 5. A mixed-design ANOVA with Question as a within-participants factor (6 levels), Discipline as a between-participants factor (3 levels), and Condition as a between-participants factor (2 levels) revealed a main effect of Question ($F(5, 3083) = 150.2, p < .001$), and suggestive effects of Discipline ($F(2, 3083) = 4.7, p = .009$) and Condition ($F(1, 3083) = 6.9, p = .009$), but no interaction. Post hoc analysis revealed that the suggestive effect observed for Discipline is due to a suggestive difference between neuroscientists and psychologists ($t(3083) = 2.9, p = .009$); no other comparison was significant

at the .05 level. All post-hoc comparisons between the six questions used were significant except for the non-significant comparisons between Carries information and Processes ($t(3083) = 2.0, p = .4$), between Is about and Identifies ($t(3083) = -1.0, p = .9$) and for the suggestive comparison between Represents and Identifies ($t(3083) = 3.2, p = .02$). To explore the role of function in the assignment of representation, we conducted an ANOVA with Question as a within-participants factor (6 levels) and Condition as a between-participants factor (2 levels). No significant or suggestive effect was observed.

Figure 5

Study 3: Function and Representation



Note. Distribution of answers for Study 3 (1: “Strongly agree;” 7: “Strongly disagree”).

Two main findings emerge from Study 3. First, as was found in Studies 1 and 2, neuroscientists and psychologist prefer thin, causal vocabulary to describe the brain’s reaction to stimuli, and are ambivalent about intentional vocabulary. Second, whether or not the brain area reacting to a stimulus is embedded in a larger network, and thus whether it has a function, influenced how the brain’s reaction was described, but did not influence whether it was described in representational terms. When it comes to representation, having a function does not seem to matter (in line with the preregistered second hypothesis).

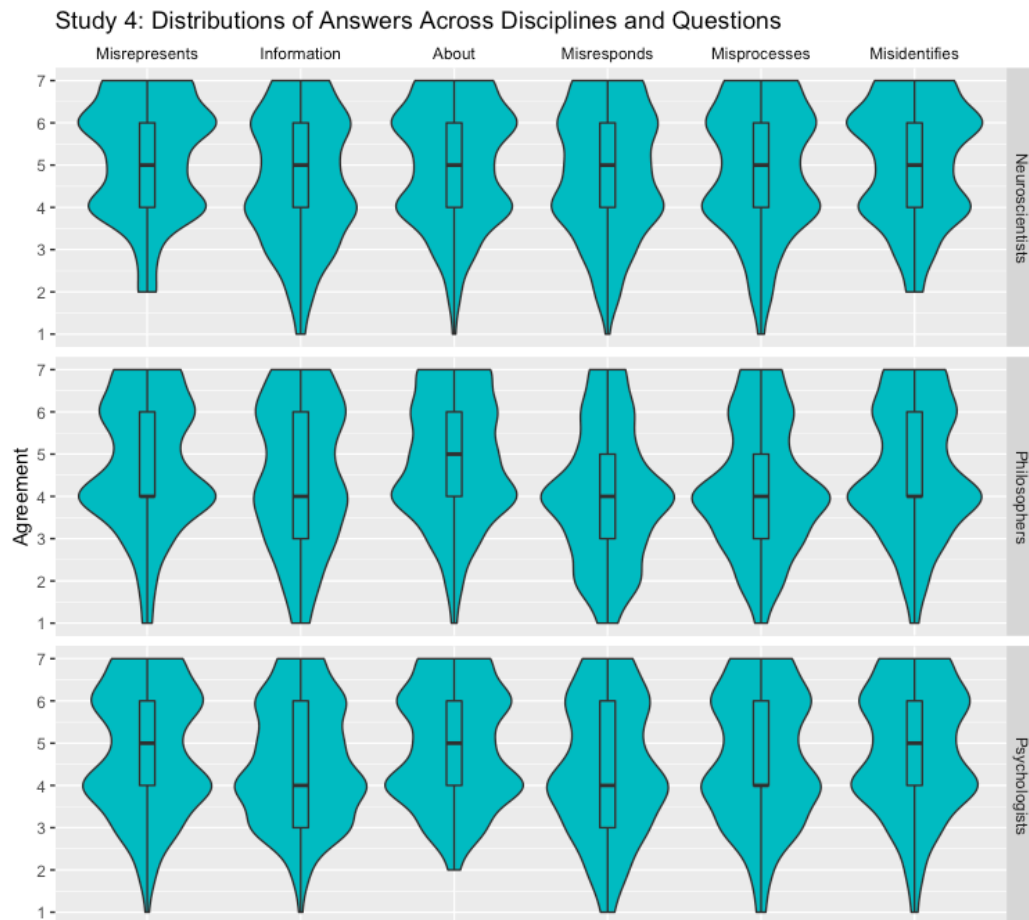
3.5. Study 4: Misrepresentation

Distribution of responses are presented in Figure 6. A mixed-design ANOVA with Question as a within-participants factor (6 levels) and Discipline as a between-participants factor (3 levels) revealed a main effect of Question ($F(5, 3101) = 11.5, p < .001$) and of Discipline ($F(2, 3101) = 26.6, p < .001$), but no interaction. Post hoc analysis revealed that the effect observed for Discipline is due to significant differences between all disciplines, philosophers being less unwilling to view the brain’s reaction as erroneous (in line with the fourth preregistered hypothesis). All pairwise comparisons between questions were significant at the .005 level, except for Represents and Is About, Represents and Identifies, Carries information and Represents, Carries information, and Processes, Is about and Identifies, and Responds and Processes, which were not significant at the .05 level, and Carries Information and Identifies and Processes and Identifies, which were only suggestive ($.01 < ps < .05$). We also found that

neuroscientists' and psychologists are unwilling to assign misrepresentation (both mean answers significantly higher than the neutral point, "neither agree nor disagree"; $ps < .001$).

Figure 6

Study 4: Misrepresentation



Note. Distribution of answers for Study 4 (1: "Strongly agree;" 7: "Strongly disagree").

The main finding to emerge from Study 4 is neuroscientists', and to a smaller extent psychologists', unwillingness to describe the brain's reaction to stimuli as erroneous, that is, as

failing to do what the brain is meant to do (contrary to the preregistered third hypothesis). In particular, neuroscientists and psychologists are unwilling to say that it misrepresents something as something else.

4. Discussion

Neuroscientists, psychologists, and philosophers believe that the concept of representation is important to understand brains, behavior, and cognition. While neuroscientists, psychologists, and philosophers occasionally differ in their responses to the stimuli, those differences appear to be very small. The concept of representation does not appear to have specialized in different directions as scientific concepts tend to do when they fulfill genuine explanatory or experimental roles (Hull, 1988; Machery et al., 2019). While our sample size was not large enough to investigate whether the concept of representation varies within disciplines (e.g., between molecular and system neuroscientists), we found no evidence for this possibility since the data were not bimodally distributed.

Furthermore, having a clear concept of representation requires having some sense of what follows from something being a representation (or of what is required for something to count as a representation), including the scale at which it occurs, the way it depends on stimuli, or how it features in downstream processes. Having a non-obscure concept of representation requires distinguishing representations from other kinds of signs. Despite the purported centrality of representations (Figure 2), Studies 1 to 4 show that the concept of representation is unclear and confused.

First, in none of the four studies were neuroscientists and psychologists willing to describe the brain's reaction as representing its stimulus. They are unwilling not because they think this reaction is not an instance of representation, but because they are ambivalent: They

neither agree nor disagree, probably because it is not clear to them what counts as a representation. This pattern is found for other intentional descriptions such as the idea that the brain's reaction is about its stimulus or what it identifies as its stimulus. This ambivalence stands in contrast with neuroscientists and psychologists' willingness to apply thinner, causal descriptions, such as responding and processing, to the brain's reaction to stimuli. Surprisingly, neuroscientists and psychologists were also willing to describe the brain's reaction in information-theoretic terms, suggesting perhaps that they understand information in more a causal sense than an intentional one.

Second, neuroscientists and psychologists do not appear to have a precise idea about what kind of brain structure or pattern counts as representation. Whether the brain's reaction was described at the neuronal (single neuron) or at the population level (area) made little difference to their answers.

Third, neuroscientists appear not to require the brain's reaction to be used in a broader neural network and thus to have a function (Cummins, 1975) to count as a representation. They could be indifferent to the role of function for representations either because they endorse a non-functional, correlation-based account of representation or because they have no clear idea about what is required for something to count as a representation. Their ambivalence in applying the concept of representation noted above suggests that the latter is more likely the case. For psychologists, on the other hand, representation requires sensitivity; that is, brain states cannot be representations if they occur in response to different types of stimuli. Thus, psychologists' concept of representation is clearer than neuroscientists: They appear to endorse a necessary condition for the application of this concept.

These first three points suggest that psychologists' and, to an even greater extent, neuroscientists' concept of representation is unclear: Psychologists and neuroscientists are unsure what properties a brain pattern must have to count as a representation, and unsure about what follows from calling a brain pattern a representation. This ambivalence extends to other intentional notions, and contrasts with thinner, causal notions.

One of the few things philosophers working on representation agree upon is that representation requires misrepresentation (e.g., Bechtel, 1998; Haugeland, 1998; Ramsey, 2007; Shea, 2018). That is to say, representations can be misapplied; for example, a map can misrepresent the region it is about; we can call a dog a "wolf;" etc. By contrast, a natural sign cannot misrepresent (Dretske, 1988): The smoke produced by the fire carries information about the fire, but it cannot misrepresent it; tree rings carry information about the age of the tree but cannot misrepresent it; and so on. Neuroscientists and psychologists are unwilling to describe the brain's reaction as erroneous, including as being a misrepresentation. Their reluctance suggests that their concept of representation is confused: They fail to distinguish natural signs and representations.

The lack of clarity and confusion of the concept of representation are not innocuous. They can 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. For instance, in the embodied cognition literature, cognitive neuroscientists have provided ample fMRI evidence that at least sometimes (e.g., Kiefer & Pulvermüller, 2012), motor and perceptual areas of the brain are activated when participants retrieve and use concepts, but critics have responded that those activations are incidental: They are not the conceptual representations themselves (e.g., Mahon & Caramazza, 2008). Without greater clarity of what it

means for a brain pattern to be a representation and some operationalization of the concept of representation, this controversy is unlikely to be resolved. Furthermore, lack of clarity and confusion of the concept of representation 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. Since the concept of representation at play is so unclear, it is hard to say what they reveal about the brain: What do we learn from them at all? The situation is even made worse by the lack of convergence across these methods.

What is to be done with an unclear and obscure scientific concept like the concept of representation? A common view is that unclear and obscure concepts must be reformed or, as philosophers now say, “explicated” (Carnap, 1950), “prescriptively analyzed” (Machery, 2017), or “engineered” (Cappelen, 2018). Alternatively, one could propose to eliminate the concept of representation from neuroscience and psychology (on elimination, see, e.g., Churchland, 1979 for folk psychological concepts; Griffiths, 1997 for the concept of emotion; Griffiths, Machery, & Linquist, 2009 for the concept of innateness).

We suspect most neuroscientists and psychologists would strongly prefer the former option, and most philosophers of psychology and neuroscience would agree. Elimination might be costly, or even impracticable. In our view, a reform of the concept of representation would specify to a sufficient degree of precision the characteristics of representation that make something a representation, including its use and its causal dependence on what it represents, and it would distinguish representations from natural signs. Operationalizations would have to be examined as well. To implement the proposed reform, a consensus conference, which would bring together neuroscientists, psychologists, and philosophers would result in a white paper

published in a leading scientific journal, might be needed, and would probably have to be supported by leading scientific organizations.

Alternatively, one might push for the elimination of the concept of representation, an option critics of mainstream psychology and neuroscience would prefer. If the concept of representation is to be eliminated, neuroscience would have to put its results, methods, and theories in non-representational terms. One might wonder what such a neuroscience would look like. While it would be presumptuous for us to dictate the shape of a future neuroscience, we note that neuroscientists are willing to describe the brain's reaction in causal and informational terms, and that tools already exist to describe the dynamics of neural processes in non-representational terms (e.g., Cunningham & Yu, 2014; Dumas, de Guzman, Tognoli, & Kelso, 2014; Honey & Sporns, 2008; Izhikevich, 2007; Shenoy, Sahani, & Churchland, 2013; Sussillo & Barak, 2013; Zhang, Kalies, Kelso, & Tognoli, 2020; for additional review see Favela, 2020, 2021).

We remain neutral here about which of these two options is preferable. But we insist that the status quo is untenable and that the concept of representation must either be reformed or eliminated.

References

- Alkan, Y., Biswal, B. B., & Alvarez, T. L. (2011). Differentiation between vergence and saccadic functional activity within the human frontal eye fields and midbrain revealed through fMRI. *PLoS One*, 6(11). <https://doi.org/10.1371/journal.pone.0025866>
- Anderson, J. R. (2015). *Cognitive psychology and its implications* (8th ed.). New York, NY: Worth Publishers.
- Bechtel, W. (1998). Representations and cognitive explanations: Assessing the dynamicist's challenge in cognitive science. *Cognitive Science*, 22(3), 295-317.
- Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R., ... & Johnson, V. E. (2018). Redefine statistical significance. *Nature Human Behaviour*, 2(1), 6-10. <https://doi.org/10.1038/s41562-017-0189-z>
- Buzsáki, G. (2019). *The brain from inside out*. New York, NY: Oxford University Press.
- Cappelen, H. (2018). *Fixing language: An essay on conceptual engineering*. Oxford, UK: Oxford University Press.
- Carnap, R. (1950). *Logical foundations of probability*. Chicago, IL: University of Chicago Press.
- Chemero, A. (2009). *Radical embodied cognitive science*. Cambridge, MA: MIT Press.
- Chomsky, N. (1980). Rules and representations. *Behavioral and Brain Sciences*, 3, 1-61.
doi:10.1017/S0140525X00001515
- Churchland, P. M. (1979). *Scientific realism and the plasticity of mind*. Cambridge: Cambridge University Press.
- Churchland, P. S. (1986). *Neurophilosophy: Toward a unified science of the mind/brain*. Cambridge, MA: MIT Press.

- Cummins, R. (1975). Functional analysis. *The Journal of Philosophy*, 72(20), 741-765.
<https://doi.org/10.2307/2024640>
- Cunningham, J. P., & Byron, M. Y. (2014). Dimensionality reduction for large-scale neural recordings. *Nature Neuroscience*, 17(11), 1500-1509.
- De Houwer, J., Van Dessel, P., & Moran, T. (2021). Attitudes as propositional representations. *Trends in Cognitive Sciences*, 25(10), 870-882. <https://doi.org/10.1016/j.tics.2021.07.003>
- Dretske, F. (1988). *Explaining behavior: Reasons in a world of causes*. Cambridge, MA: The MIT Press.
- Dumas, G., de Guzman, G. C., Tognoli, E., & Kelso, J. S. (2014). The human dynamic clamp as a paradigm for social interaction. *Proceedings of the National Academy of Sciences*, 111(35), E3726-E3734. <https://doi.org/10.1073/pnas.1407486111>
- Favela, L. H. (2020). Dynamical systems theory in cognitive science and neuroscience. *Philosophy Compass*, 15(8), e12695, 1-16. <https://doi.org/10.1111/phc3.12695>
- Favela, L. H. (2021). The dynamical renaissance in neuroscience. *Synthese*, 199(1-2), 2103-2127. <https://doi.org/10.1007/s11229-020-02874-y>
- Flint, R. D., Tate, M. C., Li, K., Templer, J. W., Rosenow, J. M., Pandarinath, C., & Slutzky, M. W. (2020). The representation of finger movement and force in human motor and premotor cortices. *eNeuro*, 7(4). <https://doi.org/10.1523/ENEURO.0063-20.2020>
- Fodor, J. A. (1981). *Representations: Philosophical essays on the foundations of cognitive science*. Brighton, Sussex: The Harvester Press.
- Gazzaniga, M. S., Ivry, R. B., & Mangun, G. R. (2014). *Cognitive neuroscience: The biology of mind* (4th ed.). New York, NY: W. W. Norton & Company Ltd.

Geissler, M., Gottschling, C., Aguado, A., Rauch, U., Wetzel, C. H., Hatt, H., & Faissner, A.

(2013). Primary hippocampal neurons, which lack four crucial extracellular matrix molecules, display abnormalities of synaptic structure and function and severe deficits in perineuronal net formation. *Journal of Neuroscience*, 33(18), 7742-7755.

<https://doi.org/10.1523/jneurosci.3275-12.2013>

Greenbaum, S., & Quirk, R. (1970). *Elicitation experiments in English: Linguistic studies in use and attitude*. Coral Gables, FL: University of Miami Press.

Griffiths, P. E. (1997). *What emotions really are*. Chicago, IL: University of Chicago Press.

Griffiths, P., Machery, E., & Linquist, S. (2009). The vernacular concept of innateness. *Mind & Language*, 24(5), 605-630.

Haugeland, J. (1998). *Having thought: Essays in the metaphysics of mind*. Cambridge, MA: Harvard University Press.

Honey, C. J., & Sporns, O. (2008). Dynamical consequences of lesions in cortical networks. *Human Brain Mapping*, 29(7), 802-809.

Hull, D. L. (1988). *Science as a process: An evolutionary account of the social and conceptual development of science*. Chicago, IL: University of Chicago Press.

Hutto, D. D., & Myin, E. (2013). *Radicalizing enactivism: Basic minds without content*. Cambridge, MA: The MIT Press.

Izhikevich, E. (2007). *Dynamical systems in neuroscience: The geometry of excitability and bursting*. Cambridge, MA: MIT Press.

Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex*, 48(7), 805-825.

- Kriegeskorte, N., & Diedrichsen, J. (2019). Peeling the onion of brain representations. *Annual Review of Neuroscience*, 42, 407-432. <https://doi.org/10.1146/annurev-neuro-080317-061906>
- Machery, E. (2009). *Doing without concepts*. New York, NY: Oxford University Press.
- Machery, E. (2017). *Philosophy within its proper bounds*. Oxford, UK: Oxford University Press.
- Machery, E., Griffiths, P., Linquist, S., & Stotz, K. (2019). Scientists' concepts of innateness: Evolution or attraction? In D. A. Wilkenfeld & R. Samuels (Eds.), *Advances in experimental philosophy of science* (pp. 172-201). New York, NY: Bloomsbury.
- Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology-Paris*, 102(1-3), 59-70.
- Planer, R. J., & Godfrey-Smith, P. (2021). Communication and representation understood as sender-receiver coordination. *Mind & Language*, 36(5), 750-770. <https://doi.org/10.1111/mila.12293>
- Poldrack, R. A. (2021). The physics of representation. *Synthese*, 199, 1307-1325. <https://doi.org/10.1007/s11229-020-02793-y>
- Ramsey, W. M. (2007). *Representation reconsidered*. New York, NY: Cambridge University Press.
- Reid, R. C., & Usrey, W. M. (2013). Vision. In L. R. Squire, D. Berg, F. E. Bloom, S. Du Lac, A. Ghosh, & N. C. Spitzer (Eds.), *Fundamentals of neuroscience* (4th ed.; pp. 577 - 598). Waltham, MA: Academic Press.

- Rheinberger, H.-J. (2000). Fragments from the perspective of molecular biology. In E. F. Keller, P. Beurton, R. Falk, & H.-J. Rheinberger (Eds.), *Decoding the genetic program* (pp. 219-239). Cambridge, MA: Cambridge University Press.
- Richardson, M. J., Shockley, K., Fajen, B. R., Riley, M. R., & Turvey, M. T. (2008). Ecological psychology: Six principles for an embodied-embedded approach to behavior. In R. Calvo & T. Gomila (Eds.), *Handbook of cognitive science: An embodied approach* (pp. 161-187). Amsterdam: Elsevier Science.
- Roskies, A. L. (2021). Representational similarity analysis in neuroimaging: Proxy vehicles and provisional representations. *Synthese*, 199, 5917-5935. <https://doi.org/10.1007/s11229-021-03052-4>
- Schnotz, W., Hauck, G., & Schwartz, N. H. (2021). Multiple mental representations in picture processing. *Psychological Research*. doi:10.1007/s00426-021-01541-2
- Shea, N. (2018). *Representation in cognitive science*. Oxford, UK: Oxford University Press.
- Shenoy, K. V., Sahani, M., & Churchland, M. M. (2013). Cortical control of arm movements: A dynamical systems perspective. *Annual Review of Neuroscience*, 36, 337-359. <https://doi.org/10.1146/annurev-neuro-062111-150509>
- Stotz, K., Griffiths, P. E., & Knight, R. (2004). How biologists conceptualize genes: An empirical study. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 35, 647-673.
- Sussillo, D., & Barak, O. (2013). Opening the black box: Low-dimensional dynamics in high-dimensional recurrent neural networks. *Neural Computation*, 25(3), 626-649.
- Trappenberg, T. P. (2014). *Fundamentals of computational neuroscience* (2nd ed.). New York, NY: Oxford University Press.

Voelker, A. R., Blouw, P., Choo, X., Dumont, N. S. Y., Stewart, T. C., & Eliasmith, C. (2021).

Simulating and predicting dynamical systems with spatial semantic pointers. *Neural*

Computation, 33(8), 2033-2067. https://doi.org/10.1162/neco_a_01410

Zhang, M., Kalies, W. D., Kelso, J. S., & Tognoli, E. (2020). Topological portraits of multiscale coordination dynamics. *Journal of Neuroscience Methods*, 339, 108672.

<https://doi.org/10.1016/j.jneumeth.2020.108672>