

**Representations are (still) theoretical posits**

Zoe Drayson

DOI: 10.1387/theoria.27209

Received: 14/01/2025

Final version: 16/05/2025



This is a manuscript accepted for publication in *THEORIA. An International Journal for Theory, History and Foundations of Science*. Please note that this version will undergo additional copyediting and typesetting during the production process.

## REPRESENTATIONS ARE (STILL) THEORETICAL POSITS

*Las representaciones son (aún) postulados teóricos*

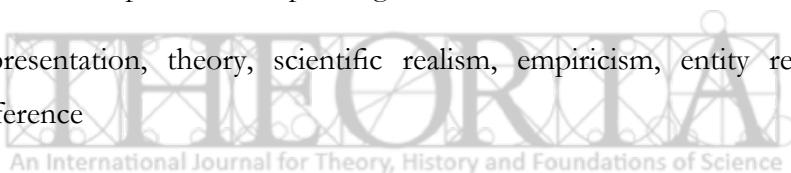
Zoe Drayson

University of California

---

**ABSTRACT:** The debate over whether cognitive science is committed to the existence of neural representations is usually taken to hinge on the status of representations as *theoretical posits*: it depends on whether or not our best-supported scientific theories commit us to the existence of representations. Thomson and Piccinini (2018) and Nanay (2022) seek to reframe this debate to focus more on scientific experimentation than on scientific theorizing. They appeal to arguments from observation and manipulation to propose that experimental cognitive neuroscience gives us non-theoretical reasons to be ontologically committed to representations. In this paper, I challenge their claims about observation and manipulation, and I argue that the question of whether we are ontologically committed to representations is still best understood as a question about the level of support we have for our representation-positing scientific theories.

**Keywords:** Representation, theory, scientific realism, empiricism, entity realism, cognitive neuroscience, inference



**RESUMEN:** El debate sobre si la ciencia cognitiva está comprometida con la existencia de representaciones neuronales suele centrarse en el estatus de las representaciones como *postulados teóricos*: depende de si nuestras teorías científicas mejor fundamentadas nos comprometen o no con la existencia de representaciones. Thomson y Piccinini (2018) y Nanay (2022) buscan reformular este debate y enfocarlo más en la experimentación científica que en la teorización científica. Apelando a argumentos basados en la observación y la manipulación, sugieren que la neurociencia cognitiva experimental nos proporciona razones no teóricas para comprometernos ontológicamente con las representaciones. En este artículo cuestiono sus afirmaciones sobre la observación y la manipulación, y sostengo que la cuestión de si estamos ontológicamente comprometidos con las representaciones debería interpretarse como una cuestión sobre el grado de apoyo que tienen nuestras teorías científicas que postulan representaciones.

**Palabras clave:** Representación, teoría, realismo científico, empirismo, realismo de entidades, neurociencia cognitiva, inferencia

**Short Summary:** This paper argues that the debate over our ontological commitments to representations in cognitive science is still focused on whether representations are explanatory posits in our best-supported scientific theories. It considers and rejects arguments from Thomson and Piccinini (2018) and Nanay (2022) which aim to show that we have theory-independent reasons for being committed to representations.

---

Is cognitive science committed to the existence of representations? The answer to this question is usually taken to hinge on the status of representations as *theoretical posits*: standard scientific realism says that we are ontologically committed to representations if and only if they are explanatory postulates in our best-supported scientific theories. There have been two recent attempts, however, to reframe this issue. Thomson and Piccinini (2018) and Nanay (2022) suggest that if we turn our attention away from scientific theorizing and towards scientific experimentation, we find theory-independent reasons to be ontologically committed to representations. In this paper, I challenge their claims and argue that representations are still best understood as theoretical posits.

In Section 1, I introduce the position which Thomson and Piccinini (2018) and Nanay (2022) want to reject: the claim that the ontological status of representations depends on their role as theoretical posits. They claim instead that cognitive neuroscience gives us experimental access to neural representations in a theory-independent way. I distinguish two separate arguments which they give for this conclusion: the *argument from observation* and the *argument from manipulation*.

In Section 2, I focus on the argument from observation, as presented by Thomson and Piccinini (2018). They argue that certain kinds of neural representations have been observed by experimental cognitive neuroscientists, and that we now have reasons to be ontologically committed to these observed entities which we lacked when we were considering them as merely theoretical posits. Thomson and Piccinini's emphasis on the theory/observation distinction, however, commits them to an empiricist notion of observability which rules out instrumental or conceptual mediation. Since their experimental claims are both instrumentally and conceptually mediated, I argue that Thomson and Piccinini have not provided us with theory-independent reasons to be ontologically committed to neural representations.

In Section 3, I turn to the argument from manipulation. This argument focuses on the ways in which cognitive neuroscientists have interacted with representations by using them as tools for experimental interventions. Nanay (2022) uses the argument from manipulation to suggest that we

have an ontological commitment to neural representations which is logically independent of the role they play in scientific theories, i.e. from their status as theoretical posits. A version of the argument from manipulation can be also attributed to Thomson and Piccinini (2018). Both draw on Hacking's (1983) entity realism. I argue that neither Nanay's argument nor Thomson and Piccinini's argument involves the sort of theory-independent manipulations required by entity realism. Insofar as they support an ontological commitment to neural representations, it is via their status as theoretical posits.

I conclude in Section 4 that Thomson and Piccinini (2018) and Nanay (2022) have provided us, at best, with supporting evidence for representational theories in cognitive neuroscience: they have provided reasons to prefer a representational rather than non-representational theory of human cognition, and thus reasons to be ontologically committed to the unobserved representations posited by such theories. This debate over the status of representational theories, however, was precisely what both parties wanted to avoid. Given that their professed aim was to shift the debate about the existence of representations away from claims about their status *qua* theoretical posits, they have not succeeded.



## 1. *Representations as theoretical posits*

Theoretical posits, sometimes called theoretical constructs, are the postulated theoretical entities named by our theoretical terms.<sup>1</sup> To claim that *representations* are theoretical posits is therefore to claim that the concept of a representation plays a role in scientific theories. Representations are the theoretical posits which distinguish cognitive science from psychological behaviorism: since the cognitive revolution, scientists have theorized that there are physical states of the brain which *stand in for* and *carry information about* aspects of the external world.

It has become almost a cliché to say that the most important explanatory posit in cognitive research is the concept of representation. Like most clichés, it also happens to be true. Since the collapse of behaviorism in the 1950s, there has been no single theoretical construct that has played such a central role in the scientific disciplines of cognitive psychology, social psychology, linguistics, artificial intelligence, and the cognitive neurosciences. (Ramsey, 2007, p. xi)<sup>2</sup>

---

<sup>1</sup> Theoreticity is standardly taken to be a property which can be applied to expressions in scientific language as well as to the referents and concepts of these expressions (Andreas 2021).

<sup>2</sup> There is arguably a notion of neurological representation which pre-dates the cognitive revolution (Chirimuuta, 2019).

There is, of course, a concept of mental representation which pre-dates the cognitive revolution. Philosophers including Descartes, Locke, and Hume write about ‘ideas’: mental states which are individuated semantically by their contents, where those contents are attributed through first-person introspection rather than third-person scientific theorizing. The representations posited by cognitive science, on the other hand, are physically-implemented vehicles of content which can be individuated non-semantically by the formal, functional, or physical properties which allow them to be manipulated by cognitive mechanisms:

The entities that vindicate the cognitive revolution’s first and defining commitment – to an internal mechanism involving vehicles of content which are individuable non-semantically – deserve to be called representations. (Shea, 2007, pp. 247-248)<sup>3</sup>

Cognitive science gives us several different ways to characterize such representations: as syntactically-structured symbols in a ‘language of thought’, as activation patterns distributed across networks, or as electrochemical signals of neural populations, for example.<sup>4</sup> For the purposes of this paper, my focus will be on neurally-implemented representations in organisms (primarily humans) rather than in artificial systems. The debate in which Thomson and Piccinini (2018) and Nanay (2022) are interested concerns whether we are ontologically committed to such neural representations. This debate is standardly characterized in terms of scientific theorizing: do our best-supported scientific theories commit us to the existence of neural representations? Thomson and Piccinini (2018) and Nanay (2022) seek to reframe this debate to focus more on scientific experimentation than on scientific theorizing. Following Hacking (1983), they propose that experimental cognitive neuroscience gives us theory-independent reasons to be ontologically committed to neural representations:

Since Ian Hacking’s (1983) groundbreaking work, philosophers of science have pointed out that experimental science often has a life of its own: through observation and manipulation, experimentalists can establish that an entity exists [...] While neural representations began as theoretical posits, neuroscientists have long reached the point where they routinely observe and manipulate representations using multiple methods and techniques in multiple model systems, just as they observe and manipulate neurons and action potentials. The techniques and procedures they use are validated independently of any debates about neuronal representation. (Thomson & Piccinini, 2018, pp. 192-193)

---

<sup>3</sup> This is compatible with thinking that representations can also be individuated semantically. When we individuate them *qua* vehicles, however, we are characterizing them in terms of the properties via which they participate in cognitive mechanisms. See Section 3 for further discussion.

<sup>4</sup> I am using the term ‘cognitive science’ broadly to include cognitive neuroscience.

Ian Hacking's view is [that] if we can manipulate  $x$  in such a way that this has direct influence on observable phenomena, we have reason to endorse realism about entity  $x$ . [...] the main claim of this paper [is] that we should be entity realist about mental representations, while being noncommittal about whether realism about representationalist theories of mind is correct. (Nanay, 2022, pp. 78-79)

In what follows, I will be exploring two proposals presented by Thomson and Piccinini (2018) and Nanay (2022) for thinking of neural representations, like electrons, as experimental entities rather than mere theoretical posits. I will argue that neither proposal gives us reason to reject the idea that neural representations are theoretical posits, and thus that neither proposal results in a substantial reframing of the existing debate over the ontological status of neural representations.

## 2. *The argument from observation*

The argument from observation, as presented by Thomson and Piccinini (2018), proposes that certain kinds of neural representations should not be understood merely as theoretical posits, on the grounds that they have been “routinely observed” by experimental neuroscientists.<sup>5</sup> They conclude that these observations significantly alter the debate over our ontological commitment to neural representations.

The historical debate on representation in cognitive science and neuroscience construes representations as theoretical posits and discusses the degree to which we have reason to posit them. We reject the premise of that debate. (Thomson & Piccinini, 2018, p. 191)

Once we recognize that neural representations [...] are routinely observed and manipulated experimentally, the long-standing debate over representations should finally be settled. Representations are no longer mere theoretical posits. (Thomson & Piccinini, 2018, p. 223)

The argument from observation involves two claims. The first of these is that (i) neural representations (or at least some kinds of them) have been observed by experimental cognitive neuroscientists. The second claim is that (ii) these observations give us new and superior epistemic access to neural representations which significantly alters the debate over our ontological commitments: in virtue of having been observed, neural representations are now something more

<sup>5</sup> Thomson and Piccinini (2018) claim at several points that neuroscientists “routinely observe and manipulate representations” without distinguishing between observation and manipulation. The 2018 paper is titled ‘Neural representations *observed*’ (my italics), and the authors themselves focus on the observation claim. I focus in Section 2 on observation and return to the issue of manipulation in Section 3.

than mere theoretical posits. In what follows, I will argue that there is no sense of ‘observation’ which would allow Thomson and Piccinini to simultaneously defend both claims.

### **(i) Have neural representations been observed?**

To endorse the argument from observation, we first need to establish that certain kinds of representations have in fact been observed. Thomson and Piccinini dedicate much of their paper to presenting evidence that “experimental neuroscientists routinely observe [...] neural representations in their laboratory” (Thomson & Piccinini, 2018, p. 191). They focus on three kinds of representations for which they believe we have the most observational data: sensory representations, motor representations, and uncoupled representations.

Thomson and Piccinini first consider sensory representations, understood as “activation patterns in the nervous system that carry information about the current environment” (Thomson & Piccinini, 2018, p. 198). They describe how neuroscientists have observed retinal ganglion cells, for example, operating as feature detectors: individual neurons which fire in response to a specific feature of the visual field, such as spatial contrast or left-to-right motion. They also describe how spatially adjacent populations of sensory neurons have been observed to respond selectively to spatially adjacent stimuli, creating a retinotopic map in the visual cortex. Thomson and Piccinini report that similar observations have been made of movement representations: individual neurons in the motor cortex seem to be tuned to specific kinds of bodily movement, firing more action potentials before a hand movement in a particular direction, for example. They also report that somatotopic maps have been observed in the motor cortex: spatially adjacent populations of motor neurons respond selectively to stimuli presented at spatially adjacent bodily locations. Thomson and Piccinini also survey the observation of ‘uncoupled’ representations, which are neither sensory nor motor. They propose that neuroscientists have observed neural representations which are involved in working memory, for example: neurons in the prefrontal cortex have been observed, which, once activated by a stimulus, will sustain their activation response once the stimulus is removed.

In what sense have these neural representations been *observed*? Some philosophers (e.g. Van Fraassen, 1980) propose that only unaided human perception qualifies as observation and thus deny that we observe stars through telescopes or cells through microscopes. Since all neuroscientific experimentation is mediated by instruments, these philosophers would similarly deny that neuroscientists have observed neural representations. Presumably, this cannot be the sense of observation which Thomson and Piccinini have in mind. On a more relaxed notion of observation, instrumentation is not an in-principle barrier to observation: many philosophers

maintain that we can make observations of the world through magnifying glasses and acoustic stethoscopes, for example, because the products of these instruments do not require interpretation.<sup>6</sup> But where we use scientific instruments to produce data which requires further interpretation, this is not considered to constitute observation of the world. Most philosophers deny that chlorine tanks allow us to *observe* solar neutrinos, for example, or that bubble chambers allow us to *observe* particle interactions (Boyd and Bogen, 2021). The observable data supplied by these instruments may allow us to *infer* the existence and properties of the phenomena in question, but we do not observe the phenomena themselves.

To establish whether neural representations have been observed, we therefore need to study the sorts of instruments which cognitive neuroscientists have used to detect them. Consider the topographic maps which allow us to observe sensory and motor neural representations, according to Thomson and Piccinini. Our data about these neural maps comes from neuroimaging techniques, and most commonly from functional magnetic resonance imaging (fMRI).<sup>7</sup> There are a several problems, however, with assuming that we can *observe* neural representations via fMRI. For one, fMRI does not allow us to observe neural activity, but only to infer neural activity from changes in the oxygenation of blood. Furthermore, the images produced by fMRI are created by subtractive algorithms which compare results from subjects performing different cognitive tasks, and they rely on theoretical assumptions about how these tasks are performed by functionally-decomposed areas of the brain.<sup>8</sup> Additionally, the images produced by fMRI rely on statistical techniques: each image is the average of multiple scans. The color intensities in the image show the likelihood of activity rather than its strength, and further statistical tests are employed to determine which of the effects in the final image are significant. The result is that the information needed to interpret fMRI images is not present in the images themselves:

We lack an intuitive understanding of how to interpret neuroimages because proper interpretation is highly theory laden; we have no antecedent knowledge of what the objects of our neuroimaging studies are like (the studies themselves constitute our understanding of the phenomena we investigate); and the image fails to contain information crucial to its correct interpretation. (Roskies, 2007, pp. 868-869)

<sup>6</sup> See Teller (2001), Boyd and Bogen (2021), and Monton and Mohler (2021). In Roskies' (2007) terminology, the products of such instruments are belief-transparent.

<sup>7</sup> Functional magnetic resonance imaging is the “dominant means” for identifying topographic maps in the human brain (Patel, Kaplan & Schneider, 2014).

<sup>8</sup> The use of fMRI to establish claims about topographic maps, for example, relies on fundamental principles about the relationships between topographic organization, anatomical structure, and function in the brain (Patel, Kaplan & Schneider, 2014).

The images which result from neuroimaging techniques therefore differ in important ways from the sorts of images produced by reflecting telescopes and optical microscopes. As a result, even those philosophers who work with a relaxed notion of observation reject the idea that fMRI images allow us to observe neural representations.<sup>9</sup> This is consistent, of course, with the view that fMRI is a reliable source of *evidence* about neural representations or that fMRI can be used to aid us in the *detection* of neural representations. But detecting a phenomenon from observational evidence is not sufficient for observing the phenomenon: detecting rainfall by observing puddles is not the same as observing rain, for example.

Some of the alleged observations of neural representations discussed by Thomson and Piccinini, however, do not rely on imaging technology and so may be less susceptible to the concerns described above. Recall that Thomson and Piccinini's examples of neural representations include individual neurons which selectively respond to features of the visual field or to specific bodily movements. The experimental data they reference draws on evidence from single-unit recordings, which use microelectrodes to measure the electrophysiological response of a single neuron. These microelectrodes measure the change in voltage potential across the cell membrane when an action-potential is evoked by the opening and closing of ion channels. Unlike in fMRI, where neural activity is inferred from metabolic properties, single-unit recordings directly measure the electrophysiological properties of a neuron. It could therefore be argued that single-unit recordings allow us to observe, rather than merely infer, neural activity.

Not all observations of neural *activity*, however, will be observations of neural *representations*, because there is no reason to think that all neural activity is representational. Thomson and Piccinini acknowledge this: they claim that observing neural representations requires establishing “that a neural signal fits the *criteria for representation*” (Thomson & Piccinini, 2018, 195, my emphasis). They propose the following criteria:

For something to count as a representation, it must have a *semantic content* (e.g., “there is yogurt in the fridge”) and an *appropriate functional role* (e.g., to guide behavior with respect to the yogurt in the fridge). (Thomson & Piccinini, 2018, p. 193, my emphasis)

---

<sup>9</sup> See Roskies (2007), Klein (2010), Boyd and Bogen (2021). Roskies (2021) makes the further argument that the data provided by neuroimaging experiments is in the form of voxel activation values, which are experimental constructions and do not meet the realist criteria for vehicles of representation.

They cash out the notion of functional role in terms of the neural activity serving as a stand-in for some property of the world, such that there should be a “functioning homomorphism [...] or, equivalently, exploitable similarity” between the representation and the content (Thomson and Piccinini, 2018, 194).<sup>10</sup> The notion of semantic content is characterized in teleosemantic terms, with different criteria for indicative sensory representations and imperative motor representations:

For sensory representations, the criteria are that (1) the signal carries information about some state external to the system, (2) there is a systematic mapping between a range of similar signals and a range of similar external states, and (3) the system uses these internal states to guide behavior. For motor representations, the criteria are that (4) the signal correlates with a future state of the environment (where the environment includes the body), (5) there is a systematic mapping between a range of similar signals and a range of similar future states of the environment, and (6) such signals actually cause movements that bring about the future states of the environment. (Thomson & Piccinini, 2018, pp. 195-196)

Thomson and Piccinini suggest that we are observing neural representations whenever we observe neural activity which meets these criteria. But to say that neural activity counts as representational when it meets functional and semantic criteria *just is* to have a prior theoretical concept of what it is for something to be a representation. To the extent that we observe neural representations, we must do so in a theory-dependent way: even if we observe neural activity and we observe its correlation with behaviors or stimuli, we still have to *infer* that the activity in question is representational based on our theoretical understanding of representations.<sup>11</sup> Perhaps when Thomson and Piccinini claim that neural representations are observable, they are merely committed to the idea that the existence of neural representations can be inferred from observational data. I will now argue, however, that this notion of observability does not allow them to conclude that observation gives us superior epistemic access to neural representations in a way which would significantly alter the debate over our ontological commitments.

---

<sup>10</sup> Notice that on Thomson and Piccinini’s own homomorphism criterion, it is unclear how single unit recordings would qualify as representations: they seem to lack the structure to be homomorphic to anything. See Facchin (2024) for further discussion of this point.

<sup>11</sup> Notice that this applies as much to iconic representations as it does to symbolic representations. As Facchin (2024) correctly points out, even if cognitive neuroscience relies on structural representations (neural vehicles whose physical shape resembles their targets in such a way that their semantic contents can play a causal role in mechanistic explanations of the brain), this does not make their status as representations any more observable.

## (ii) Does observation bring new ontological commitments?

Thomson and Piccinini (2018) are not merely trying to establish that experimental neuroscientists have observed neural representations. The key claim of their paper is that the observation of neural representations significantly alters how we frame debates about the ontological status of representations:

We argue that experimental neuroscientists routinely observe and manipulate neural representations in their laboratory. Therefore, neural representations are as real as neurons, action potentials, or any other well-established entities in our ontology. (Thomson & Piccinini, 2018, p.191)

Neural representations are observable [...] Therefore, neural representations are real [...] Representations are no longer mere theoretical posits: they are as established a part of our ontology as anything that can be empirically discovered. (Thomson & Piccinini, 2018, p. 223)

Notice that Thomson and Piccinini seem committed to the importance of a distinction between observed empirical entities, on one hand, and unobserved theoretical entities, on the other hand. They are proposing that empirically-observed neural phenomena are in some sense better-established or “more real” than neural phenomena which are posited by our neuroscientific theories but remain unobserved. This demonstrates a clear commitment to some form of empiricism over scientific realism. The scientific realist proposes that we are ontologically committed to the posits of our best-supported scientific theories, whether or not those posits are observable. The empiricist denies this, either by denying that theories which posit unobservables have a truth value (logical empiricism) or by denying that we are justified in believing theories which posit unobservables (constructive empiricism). Either variety of empiricism leads to the conclusion that we have an ontological commitment to observable entities which we lack to unobservable entities, even where they are posited by our best-supported scientific theories.<sup>12</sup>

Endorsing empiricism, however, relies on having a clear distinction between observation and theory, such that inferences drawn from observable evidence to the existence of some entity or property do not count as observations of that entity or property. I have argued above that Thomson and Piccinini have only established that neural representations have been observed in the sense that we can infer their existence from observational data by relying on certain theoretical

<sup>12</sup> It is unclear whether Thomson and Piccinini (2018) should be read as claiming that there is a metaphysical difference between observables and unobservables, or merely that we should take a different epistemic attitude toward claims which posit observables and claims which posit unobservables. The latter is akin to van Fraassen’s (1980) constructive empiricism, on which we can remain agnostic about the truth of claims which posit unobservables.

commitments. But this is the position adopted by the scientific realist about representations, which is denied by the empiricist.<sup>13</sup> The scientific realist argues that we are ontologically committed to neural representations because they are posited by scientific theories which are well-confirmed by empirical evidence, whether or not those neural representations are themselves observable, because our best-supported scientific theories are at least approximately true and their theoretical terms successfully refer to mind-independent entities (Chakravarthy, 2017). Thomson and Piccinini's claim that we are ontologically committed to the existence of neural representations seems to rely on an inference to the best explanation: that the observational data from cognitive neuroscience are better explained by a theory which posits neural representations than by a theory which doesn't posit neural representations.<sup>14</sup> But if they are merely arguing that we have reason to be ontologically committed to neural representations because we can infer them from our observational data combined with our best theories, then there is no sense in which they have rejected or reframed the debate over neural representations.

*An International Journal for Theory, History and Foundations of Science*

I have argued that Thomson and Piccinini's attempt to reframe the debate over neural representations seems to rely on endorsing an empiricist approach to ontological commitment, on which observable scientific entities are in some sense “more real” (or at least that our belief in them is more warranted) than the unobserved entities posited by our best scientific theories. For this to work, however, neural representations would need to have been observed in a strongly empiricist sense, rather than merely inferred from observed data on the basis of our best-supported scientific theories. Thomson and Piccinini have not established that neural representations have been observed in this stronger sense, so they have not succeeded in reframing the debate: whether

---

<sup>13</sup> Notice that Piccinini self-identifies as a scientific realist rather than an empiricist in the 2020 PhilPapers survey (<https://survey2020.philpeople.org/>). The survey asks “Science: scientific realism or scientific anti-realism?” with the following options: accept or lean towards either view, reject or lean against either view, accept an alternative view, remain agnostic or undecided, claim there is no fact of the matter, declare the question too unclear to answer, or give some other response. Piccinini answers “accept scientific realism” (<https://philpeople.org/profiles/gualtiero-piccinini/views>). He confirmed in personal correspondence (April 2025) that he is a scientific realist.

<sup>14</sup> Otherwise it seems hard to make sense of claims like the following: “It has been clear at least since Descartes that anomalous perceptual phenomena (hallucinations, dreams, illusions) are extremely difficult to explain if perception is just skilled engagement with the world, without any need for representational intermediaries” (Thomson and Piccinini, 2018, 203). While I interpret Thomson and Piccinini (2018) as relying on an inference to the best explanation, my point is the same if they are performing a deductive inference which relies on a premise about the theoretical properties of representations (e.g. their functional role, the fact that they have semantic contents). I am grateful to Edouard Machery for encouraging me to make this clearer.

or not we are ontologically committed to neural representations still relies on neural representations being theoretical posits.

Notice that as I have interpreted Thomson and Piccinini's argument from observation, it sits uneasily with their claim to be drawing on Hacking's insights. It is true that Hacking (1983), like Thomson and Piccinini (2018), steers the debate about scientific entities away from theoretical considerations and towards experimental considerations. But Hacking is not trying to establish that experimental posits are observable: Hacking rejects the empiricist claim that we only have epistemic access to observables.<sup>15</sup> He also rejects the scientific realist's claim that our epistemic access to unobservables depends on inferring their existence from scientific theories, proposing instead that experimental interactions can give us epistemic access to entities independently of the theories which postulate them. Hacking describes his own position as 'entity realism', understood as the stance that "[o]ne can believe in some entities without believing in any particular theory in which they are embedded" (Hacking, 1983, p. 29). Hacking intends his entity realism to be an alternative to both empiricism and scientific realism: he argues that we can be ontologically committed to unobservables (*contra* empiricism) without being justified in believing the truth of a scientific theory (*contra* scientific realism).<sup>16</sup> If we are to make sense of Thomson and Piccinini as offering a Hacking-style argument, we might therefore be better off interpreting their position as the claim that experimental *manipulation* – rather than observation – can provide us with an ontological commitment to neural representations.

### ***3. The argument from manipulation***

The standard debate over our ontological commitment to neural representations is framed from the point of view of the scientific realist, who proposes that we are ontologically committed to representations if and only if they are posits in our best scientific theories. I have argued that Piccinini and Thomson's attempt to reframe the debate from the empiricist point of view, by arguing that neural representations have been observed, does not succeed. But might Hacking's entity realism provide an alternative way to reframe the debate? Entity realism is the view that

---

<sup>15</sup> Hacking proposes that experiments don't just allow us to test hypotheses about observable entities: experiments can show that "entities that in principle cannot be 'observed' are regularly manipulated to produce new phenomena and to investigate other aspects of nature" (Hacking, 1983, pp. 262).

<sup>16</sup> Hacking suggests that if our ontological commitment to an experimental entity relied on observing the entity in question, we would still be "stuck in the nineteenth-century rut of positivism-cum-phenomenology" (Hacking, 1983, p. 208).

scientific experiments can give us epistemic access to unobservable entities independently of our beliefs in the scientific theories which posit such entities. This epistemic access to entities results from our use of an entity as a tool or an instrument of inquiry, where our experimental interactions involve “*manipulating* an entity, in order to experiment on something else” (Hacking, 1983, p. 263, italics in original).

Hacking’s own key example is electrons. The fact that experimentalists have sprayed electrons on a supercooled niobium ball to directly manipulate its electrical charge is sufficient, he maintains, to demonstrate the existence of electrons:

By the time that we can use the electron to manipulate other parts of nature in a systematic way, the electron has ceased to be something hypothetical, something inferred. It has ceased to be theoretical and has become experimental. (Hacking, 1983, p. 262)

Hacking proposes that these sorts of experiments give us an ontological commitment to electrons which does not depend on inferring their existence from a scientific theory of electrons: “*So far as I’m concerned, if you can spray them, they are real*” (Hacking, 1983, p. 23, italics in original).

Can Hacking-style entity realism be applied to neural representations? This is the claim made Nanay (2022) and hinted at by Thomson and Piccinini (2018). The suggestion is that if we can manipulate neural representations experimentally, then we might have reasons to be committed to their existence which are independent of our scientific theories of neural representation. If this works, then it could allow us to reframe the debate over neural representations in a way which avoided their status as theoretical posits.

A major obstacle faced by entity realism, however, is its potential incoherence. Hacking’s position seems to rely on the controversial idea that we can have knowledge of an experimental entity without being committed to the truth of a scientific theory which posits that entity. In order to be manipulable, however, experimental entities must have causal properties. Many philosophers of science have simply denied that the causal properties of an entity can be detached from scientific theories describing the entity as Hacking requires:<sup>17</sup>

the experimental realist can only have knowledge about theoretical entities if she assumes that the theories which describe those entities are at least approximately true (Resnik, 1994, p. 395)

---

<sup>17</sup> Musgrave (1996) suggests that such a detachment would be akin to believing in hobgoblins without having any beliefs about what hobgoblins are like or what they do.

can we assert that [...] entities exist as part and parcel of the furniture of the world, without also asserting that they have some of the properties attributed to them by our best scientific theories? I take it that the two assertions stand or fall together. [...] The very same process is involved in accepting the reality of an entity, and in accepting the (approximate) correctness of its theoretical description. (Psillos, 1999, pp. 256-257)

Contemporary proponents of entity realism acknowledge the difficult in separating our knowledge of an entity from knowledge of the theories about the entity. They tend to draw a distinction between those properties of an entity which we can access experimentally, and those properties which we can only access via a theoretical understanding of the entity. Egg (2017), for example, distinguishes between the causal properties of an entity which are measurable with scientific instruments (its 'detectable' properties) and the abstract properties of an entity which are attributable to it only by a theoretical model (its 'auxiliary' properties). Ontological commitment to an entity is causally warranted (rather than merely theoretically warranted), he proposes, only when we can infer its existence from its detectable properties. An alternative approach, found in Nanay (2019), is to distinguish between singular causal claims about an entity's property tokens and general causal claims about an entity's property type. Nanay proposes that purely experimental manipulations can demonstrate the truth of singular causal claims, but that they cannot demonstrate the truth of general causal claims. As a result, we can have non-theoretical epistemic access to property tokens, but only theoretical epistemic access to property types.<sup>18</sup>

Let's assume that there is some coherent version of entity realism along these lines and explore how it might apply to neural representations. Can the experiments of cognitive neuroscience give us non-theoretical access to neural representations in the same way that experimental physics, according to Hacking, gives us non-theoretical access to electrons? Nanay proposes that "if we can manipulate mental representations in a way that would have direct influence on behavior, we would have a strong case for entity realism about mental representations." (Nanay, 2022, p. 82).

To make his case, Nanay appeals to experiments in the cognitive science of motor control which explore how motor representations interact with other representations and with behavior. He references some experiments which indicate that motor representations operate independently from higher-level perceptual representations (e.g. Pelisson et al, 1986), and other experiments which indicate that motor representations are influenced by top-down processing (e.g. McIntosh

---

<sup>18</sup> A further way to defend a coherent concept of entity realism is proposed by Eronen (2015), who relies on the notion of robustness. For a discussion of the relationship between the entity realist views of Egg, Nanay, and Eronen, see Khalili (2023).

and Lashley, 2008). These experiments focus on the role of motor representations in generating and altering the trajectories of our arm movements and the grip size of our hands. This leads Nanay to conclude that the experimentalists “manipulate the motor representations to bring about direct changes in our observable behavior” (Nanay, 2022, p. 84), providing us with the sort of experimental access which would justify entity realism about motor representations.<sup>19</sup> My concern with the experiments referenced by Nanay, however, is that the manipulations in question are alterations of the subject’s external environmental features rather than direct alterations of their neural representations. Experimenters are manipulating the location and size of objects, for example, and then measuring the changes in limb direction or grip size that result. But for the entity realist, ontological commitment to an entity comes via manipulations of the entity itself and not merely via manipulations of other features of the world which influence it. Hacking does not think that we get an ontological commitment to electrons, for example, by doing something which changes the electrical charge on the niobium ball and then inferring that this change is best explained by the intermediate effects of our actions on electrons. But in Nanay’s examples, changes in the subject’s behavior are being explained by changes in their motor representations, which are themselves being inferred from changes in the sensory input assumed to be caused by the manipulations of the environment. For Nanay, “we have good reasons for positing representations” (Nanay, 2022, p. 82) because the changes to the motor representations changes are inferred as the best explanation of the changes in behavior, given our manipulation of the environmental inputs to our sensory systems. This no longer looks like Hacking-style entity realism, but rather seems a lot closer to scientific realism.

Recall that in cognitive science, a representation is not merely a disposition to alter one’s behavior in response to sensory input: it is a physically-implemented vehicle of content which is individuable in terms of its physical or functional properties. To manipulate a neural representation in the sense required by entity realism would require intervening on a specific property of the internal representational vehicle, rather than merely intervening on a feature of the environment which is inferred to have downstream effects on some property or another of the representational vehicle. For Hacking, electrons are real because you can spray them. In Nanay’s experimental examples, there is no sense of what we are doing to neural representations that is analogous to spraying.

---

<sup>19</sup> He makes a similar claim about pragmatic mental imagery: saying that “Manipulating pragmatic mental imagery leads to observable behavioral changes in the same way as manipulating motor representations leads to observable behavioral changes” (Nanay, 2022, p.86).

Notice that the non-representationalist can allow that Nanay's experiments demonstrate a causal connection between changes in sensory input and changes in behavioral output which is mediated by neural activity, while denying that this gives us reason to infer the existence of neural representations. I propose that the experiments referenced by Nanay are not sufficiently analogous to the sorts of experiments which the entity realist requires in order to warrant an ontological commitment.

This is not, of course, an in-principle objection to entity realism about neural representations. There may be other experiments in cognitive neuroscience which would result in epistemic access to neural representations which is independent of our theories and inferences about neural representations. If we return to Thomson and Piccinini (2018), it looks like we find cases where internal representations themselves, and not merely external aspects of the environment, are manipulated in the manner required by Hacking-style entity realism. When discussing sensory representations, for example, Thomson and Piccinini reference motion-detecting neurons which selectively respond to activity in a certain direction. They discuss experiments performed on rhesus monkeys in which clusters of these directionally-selective neurons were electrically stimulated, which significantly increased the monkeys' saccadic eye-movements toward the direction in question (Salzman et al., 1990). And when Thomson and Piccinini discuss somatosensory maps in the motor cortex, they reference experiments on rhesus monkeys in which surgical lesions were made to the area of the primary motor cortex which is active when the monkeys make arm movements; the monkeys' fine-grained motor performance when reaching and grasping for food decreased in proportion to the volume of white matter removed (Darling et al., 2011). Unlike the experiments referenced by Nanay (2022), these experiments seem to involve directly manipulating specific neural properties (e.g. electrically stimulating neurons, lesioning white matter) in ways which have specific and proportional effects on behavior (e.g. changing the direction of eye-saccades, decreasing accuracy of hand-movements). As Cao (2022) emphasizes, what makes some neural entity deployable as a representation is the degree to which changes in our manipulation of some vehicular property match the changes in content reflected by the resulting behavior.<sup>20</sup> Paraphrasing Hacking, we might say that *if you can electrically stimulate them or surgically lesion them, they are real.*

---

<sup>20</sup> Cao is not arguing for entity realism, however, but rather for a form of representational pragmatism on which our representational commitments are relative to both the probes used to detect them and our interests.

To be an entity realist about some neural state requires that there is a well-defined notion of what it means to intervene on it.<sup>21</sup> But to have a well-defined notion of what it means to intervene on a neural property such as firing-rate is not yet to have a well-defined notion of what it means to intervene on a neural representation. For this, we would presumably have to know how firing-rate functions to represent movement direction, for example – and this seems to require some theory of representation.<sup>22</sup> To what extent, then, are the experimental manipulations referenced by Thomson and Piccinini sufficiently theory-independent to satisfy the entity realist? Of course, contemporary entity realists tend to allow that no experimental manipulations will be entirely theory-independent, because the causal properties of an entity cannot be wholly detached from scientific theories describing the entity – as discussed above. They still think that entity realism is a coherent position, as long as the properties of the entity that we are manipulating are those properties which can be accessed experimentally without a theoretical understanding. Increases in electrical stimulation and the removal of white matter, for example, seem to be sufficiently detectable to qualify as causal properties for Egg (2017). And since individual experiments take place on token neural properties, the results should be describable by the sorts of true singular causal claims which Nanay (2019) thinks give us epistemic access to those property tokens.

Couldn't the non-representationalist, however, allow that lesioning and stimulating neural areas has the effects in question while denying that what is being manipulated are neural *representations*? What makes something a vehicle of representation (rather than some non-representational neural entity) is that it has a semantic content. And notoriously, the contents of representations cannot be specified purely causally – some theory of content is required. (For Thomson and Piccinini, for example, that theory of content is a teleosemantic one.) The semantic properties of neural representations, however, look like they would count as auxiliary properties rather than detectable properties under Egg's (2017) distinction: they are attributable to the neural vehicle only under a

---

<sup>21</sup> This is why Khalili (2023) denies that we can be entity realists about dark matter, for example. One promising line of research for the entity realist about neural representation involves work in optogenetics, a technique for using light to control the activity of genetically-defined classes of neurons which has been used to manipulate memory representations ("engrams"). As Robins (2023) points out, optogenetic experiments provide a new level of precision which allows experimenters to target and track individual engrams, to experimentally demonstrate a way to distinguish memory storage from memory retrieval, and to explore the precise nature of memory consolidation in the hippocampus and prefrontal cortex.

<sup>22</sup> Recall that Thomson and Piccinini propose that for some neural state to count as a representation, it must play "an appropriate functional role (e.g., to guide behavior with respect to the yogurt in the fridge)" (Thomson and Piccinini, 2018, p.193).

theoretical model.<sup>23</sup> So while these experiments might warrant entity realism about specific kinds of neural properties, they would not warrant entity realism about neural *representations* on at least one prominent formulation of the position.

Even setting aside issues about semantic content, the very idea of representational *vehicles* causes problems for entity realism. To classify a neural state as a vehicle of representation is to individuate it in terms of the causal properties to which neurocognitive mechanisms are sensitive: this is what enables cognitive neuroscience to capture generalizations which are not present in lower-level neurophysiology. Individuating neural states as representational vehicles allows us to say whether two token neurophysiological states are tokens of the same or different representational types.<sup>24</sup> But recall that Nanay's (2019) defense of entity realism distinguishes between property tokens (which license entity realism) and property types (which don't license entity realism). If experiments only give us epistemic access to neural states in virtue of being manipulations of property tokens, as Nanay proposes, then we are not warranted in being *realist* about those neural states *qua* representational vehicles: this would require making general causal statements about property types. Nanay denies that experimentation can give us theoretically-independent epistemic access to property types.

The upshot of this is that neither Nanay (2022) nor Piccinini and Thomson (2018) have provided a version of the argument from manipulation which establishes entity realism about neural representations independently of these neural representations understood as theoretical posits of our best scientific theories.

#### 4. Conclusion

Philosophers of cognitive science have spent decades fighting the 'representation wars'.<sup>25</sup> On one side are those who propose that our best scientific theories of intelligent behavior appeal only to non-representational mechanisms like neural oscillators or dynamical systems, and who argue that we thus have no grounds for an ontological commitment to representational entities. On the other

<sup>23</sup> This point seems to apply even to the concept of *structural* representation, on which the physical properties of neural vehicles resemble their targets in a such a way that we can understand their semantic contents as playing a causal role. The notion of resemblance required is the *abstract* relationship of second-order resemblance, and thus semantic content would still count as an auxiliary property rather than a detectable property.

<sup>24</sup> See Drayson (2018) for further discussion of the distinction between representational vehicle types and their token realizers.

<sup>25</sup> For an overview of the 'representation wars', see Constant, Clark, and Friston (2021).

side are those who propose that we need to appeal to internal stand-ins with semantic properties in order to adequately explain the same behavior, and who argue that we are thus ontologically committed to representational entities. What both sides have in common is an apparent commitment to scientific realism, understood as a claim about our scientific *theories*: to the extent that we are justified in taking our best scientific theories to be approximately true, we are ontologically committed to their theoretical posits, whether observable or unobservable. One way to reframe the debate over our ontological commitment to neural representations, therefore, would be to reject the presumption of scientific realism on which it seems to stand. If we adopt an empiricist stance instead, however, then we seem to lack ontological commitments to neural representations because they are not observable in the appropriate sense.

Thomson and Piccinini (2018) boldly attempt to use the empiricist stance to justify an ontological commitment to neural representations, by claiming that the representations in question have in fact been observed by experimental cognitive neuroscientists. In Section 2, I argued that Thomson and Piccinini's argument does not succeed, because they establish at most that the existence of neural representations can be inferred from observational evidence in combination with certain theoretical commitments about functional role and semantic content. They are, in effect, simply offering evidence that supports the scientific realist in the existing debate: they are arguing that the best explanation of our observational data is a theory which posits neural representations, and they are concluding that we should be ontologically committed to the posits of our best theory. So they have neither rejected nor reframed the existing debate.

If we focus on Thomson and Piccinini's claims about the manipulability rather than the observability of neural representations, then we can interpret them as offering an argument for Hacking-style entity realism about neural representations, similar to the one offered by Nanay (2022). In principle, the entity realist could argue that we can be ontologically committed to the existence of unobservable neural representations without having to accept the truth of theories which posit neural representations. This approach would offer an alternative to both scientific realism and empiricism. Even if we can prevent entity realism *in general* from collapsing back into scientific realism, however, entity realism *about neural representations* seems incredibly difficult to salvage. In Section 3, I argued that insofar as the concept of neural representation in cognitive neuroscience is committed to either some notion of semantic content or to claims about the individuation of representational vehicles, there seems to be no form of entity realism which would allow us to have appropriately theory-independent epistemic access to neural representations.

I have suggested that the only way in which either the argument from observation or the argument from manipulation justifies us in being ontologically committed to neural representations is via inference. Thomson and Piccinini come close to acknowledging this point when they discuss the uncoupled representations involved in working memory:

well-crafted behavioral experiments, coupled with the general fact that maintenance of stimulus-specific information is required to solve working memory tasks, lets us *infer* that working memory is representational. [...] *Inferring* representations from behavior and generic informational considerations in this way is a useful, and relatively ubiquitous, first move toward representation observed. (Thomson and Piccinini 2018, pp. 207-208, my emphasis)

Similarly, as I mentioned in Section 3, Nanay seems to concede that our ontological commitment to neural representations is the result of inference to the best explanation:<sup>26</sup>

Stephen Stich said in 1984 that: ‘we now have an enormous collection of experimental data which, it would seem, simply cannot be made sense of unless we postulate something like [representations]’ (Stich 1984, p. 649). I am not sure that this was in fact true in 1984 [...] But with the advances of the cognitive neuroscience of action, Stich’s claim is definitely true now. (Nanay, 2022, p. 88)

Postulating representations to account for our experimental data, however, is simply a return to scientific realism. Nanay appears at one point to think that this position is compatible with entity realism, because it results in *realism about the entities* in question. But realism about unobservable entities is a commitment shared by both standard scientific realism and entity realism. What distinguishes entity realism is the claim that we can be ontologically committed to unobservable entities *without* being justified in believing the truth of a scientific theory which posits those entities.<sup>27</sup> If Nanay ultimately takes himself only to have established realism about neural representations by inference, then he has not succeeded in his professed aim to “shift the emphasis from the debate concerning realism about theories to the one concerning realism about entities” (Nanay, 2022, p.75).

---

<sup>26</sup> Further evidence of Nanay’s inference to the best explanation can be found in his discussion of his ‘Criterion B’ and his appeal to Morgan’s Canon (Nanay, 2022, pp. 82-83).

<sup>27</sup> This point is also made by Khalili, who emphasizes that “*theory* realism—the realist attitude toward most aspects of scientific theories and models that rely somehow on empirical evidence—is not included under the definition of entity realism” (Khalili, 2023, p. 902). Nanay proposes at one point that “the core commitment of entity realism is that the more evidence we have about the causal powers of x, the more reason we have to be realist about x” (Nanay, 2022, p. 78), which also fails to adequately distinguish entity realism from standard scientific realism. Elsewhere, however, Nanay (2019) himself seems to endorse a view of entity realism as excluding theory realism.

In conclusion, I propose that neither Thomson and Piccinini (2018) nor Nanay (2022) have succeeded in reframing the existing debate over the existence of neural representations. Insofar as either party has given us reasons to be ontologically committed to neural representations, these are reasons which rely on the status of neural representations as *theoretical posits*: entities which are characterized by their role in our best-supported scientific theories.

## Acknowledgements

Earlier versions of this paper were presented at a workshop on Representations in Cognitive Science at the University of Barcelona in September 2002 and at a Philosophy, Psychology and Neuroscience research seminar at the University of Glasgow in February 2023. I am grateful to the audiences at both these events for their feedback. Rohan French and Sarah Robins provided helpful suggestions, and Gabe Dupre and Dan Weiskopf both gave extensive comments on a written draft. I have also benefitted from discussing the issues in this paper with faculty and graduate students in the MENTaL reading group at UC Davis, and with the Neural Mechanisms Online community.



## References

Andreas, H. (2021). *Theoretical terms in science*. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2021 Edition). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/fall2021/entries/theoretical-terms-science/>

Boyd, N., & Bogen, J. (2021). *Theory and observation in science*. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2021 Edition). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/win2021/entries/science-theory-observation/>

Cao, R. (2022). Putting representations to use. *Synthese*, 200(2), 151. <https://doi.org/10.1007/s11229-022-03522-3>

Chakravartty, A. (2017). *Scientific realism*. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/sum2017/entries/scientific-realism/>

Chirimutu, M. (2019). Synthesis of contraries: Hughlings Jackson on sensory-motor representation in the brain. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 75, 34–44. <https://doi.org/10.1016/j.shpsc.2019.01.007>

Darling, W. G., Pizzimenti, M. A., Hynes, S. M., Rotella, D. M., Headley, G., Ge, J., Stilwell-Morecraft, K. S., McNeal, D. W., Solon-Cline, K. M., & Morecraft, R. J. (2011). Volumetric effects of motor cortex injury on recovery of ipsilesional dexterous movements. *Experimental Neurology*, 231(1), 56–71. <https://doi.org/10.1016/j.expneurol.2011.05.015>

Drayson, Z. (2018). The realizers and vehicles of mental representation. *Studies in History and Philosophy of Science*, 68, 80–87.

Egg, M. (2017). Entity realism. In J. Saatsi (Ed.), *The Routledge handbook of scientific realism* (pp. 20–32). Routledge.

Eronen, M. (2015). Robustness and reality. *Synthese*, 192(12), 3961–3977.

Facchin, M. (2024). Neural representations unobserved—or: A dilemma for the cognitive neuroscience revolution. *Synthese*, 203(7). <https://doi.org/10.1007/s11229-023-04418-6>

Hacking, I. (1983). *Representing and intervening: Introductory topics in the philosophy of natural science*. Cambridge University Press.

Khalili, M. (2023). A dialogue among recent views of entity realism. *Philosophy of Science*, 90(4), 901–921. <https://doi.org/10.1017/psa.2023.18>

McIntosh, R. D., & Lashley, G. (2008). Matching boxes: Familiar size influences action programming. *Neuropsychologia*, 46(9), 2441–2444. <https://doi.org/10.1016/j.neuropsychologia.2008.03.003>

Monton, B., & Mohler, C. (2021). *Constructive empiricism*. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2021 Edition). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/win2021/entries/constructive-empiricism/>

Musgrave, A. (1996). Realism, truth and objectivity. In R. S. Cohen, R. Hilpinen, & Q. Renzong (Eds.), *Realism and anti-realism in the philosophy of science: Beijing International Conference, 1992* (pp. 19–44). Springer. [https://doi.org/10.1007/978-94-015-8638-2\\_3](https://doi.org/10.1007/978-94-015-8638-2_3)

Nanay, B. (2019). Entity realism and singularist semirealism. *Synthese*, 196(2), 499–517. <https://doi.org/10.1007/s11229-016-1179-9>

Nanay, B. (2022). Entity realism about mental representations. *Erkenntnis*, 87(1), 75–91. <https://doi.org/10.1007/s10670-019-00185-4>

Pélisson, D., Prablanc, C., Goodale, M. A., & Jeannerod, M. (1986). Visual control of reaching movements without vision of the limb. *Experimental Brain Research*, 62(2), 303–311. <https://doi.org/10.1007/BF00238849>

Psillos, S. (1999). *Scientific realism: How science tracks truth*. Routledge.

Resnik, D. B. (1994). Hacking's experimental realism. *Canadian Journal of Philosophy*, 24(3), 395–411.

Robins, S. (2023). The 21st century engram. *WIREs Cognitive Science*, 14(5), e1653. <https://doi.org/10.1002/wcs.1653>

Roskies, A. L. (2007). Are neuroimages like photographs of the brain? *Philosophy of Science*, 74(5), 860–872. <https://doi.org/10.1086/525627>

Roskies, A. L. (2021). Representational similarity analysis in neuroimaging: Proxy vehicles and provisional representations. *Synthese*, 199(3), 5917–5935. <https://doi.org/10.1007/s11229-021-03052-4>

Salzman, C. D., Britten, K. H., & Newsome, W. T. (1990). Cortical microstimulation influences perceptual judgements of motion direction. *Nature*, 346(6280), 174–177. <https://doi.org/10.1038/346174a0>

Shea, N. (2007). Content and its vehicles in connectionist systems. *Mind and Language*, 22(3), 246–269. <https://doi.org/10.1111/j.1468-0017.2007.00308.x>

Teller, P. (2001). Whither constructive empiricism? *Philosophical Studies*, 106(1), 123–150. <https://doi.org/10.1023/A:1013170506726>

Thomson, C., & Piccinini, G. (2018). Neural representations observed. *Minds and Machines*, 28(1), 191–235. <https://doi.org/10.1007/s11023-018-9459-4>

Van Fraassen, B. C. (1980). *The scientific image*. Oxford University Press. <https://doi.org/10.1093/0198244274.001.0001>



**Zoe Drayson** is Associate Professor of Philosophy and Director of Cognitive Science at the University of California, Davis. She works in the philosophical foundations of cognitive science, with a particular interest in the nature of explanation in the mind-brain sciences.

**Address:** Philosophy Department, University of California, Davis, 1 Shields Avenue, Davis, CA 95616, USA. E-mail: [zdrayson@ucdavis.edu](mailto:zdrayson@ucdavis.edu) – ORCID: 0000-0002-1665-3319