Nicholas Shea, *Representation in Cognitive Science*. Oxford: Oxford University Press. Pp. xi, 292.

Nicholas Shea's writings are required reading for philosophers of mind, especially those interested in the mind sciences. *Representation in Cognitive Science* cements his position as one of the field's foremost practitioners. The book mobilizes his broad and deep knowledge of cognitive science, his formidable philosophical craft, and his admirably clear writing style to powerful effect.

When a mental state represents the world as being a certain way, we can ask whether the mental state *correctly* represents how the world is. Shea aims to provide naturalistically sufficient conditions for mental states to have correctness-conditions. Thus, his discussion contributes to the literature on "naturalizing intentionality." Unlike most participants in that literature, Shea illustrates his approach with detailed scientific examples, including research on perception, motor control, animal navigation, and causal reasoning. The generous buffet of scientific case studies is a welcome contrast with previous naturalizing efforts, which all too often read like exercises in make-believe.

Shea calls his approach *varitel semantics*. The basic idea is that representation arises from "exploitable relations" between the representing system and the represented domain. A paradigmatic exploitable relation is *correlation*: states of the representing system systematically correlate with states of the represented domain. For example, firing activity in a population of neurons may correlate with the direction in which a visual stimulus moves. Most philosophers agree that not just any correlation suffices for representational content to be present. What must be added to mere correlation for genuine representation to arise? Shea's answer, briefly, is that

correlations generate content when they are exploited in furtherance of *task functions*, such as procuring food or moving to a desirable location. Shea develops this answer in vivid detail, articulating a naturalistic account of task functions (Chapter 3) and a systematic theory describing how exploited correlations generate content (Chapter 4). His framework applies both to biological systems and to artifacts (e.g. robots).

Chapter 5 investigates a second exploitable relation that can generate representation: *structural correspondence*. Structural correspondence occurs when a relation over the representing system mirrors a relation over the represented domain: for instance, distances on a map mirror distances in physical space. According to Shea, representational content arises when a system exploits a structural correspondence in furtherance of a task function. His main example is spatial representation. Citing extensive neuroscientific evidence, he argues that rats navigate by exploiting a structural correspondence between neural activity and space. Exploitation of the correspondence ensures that the rat's neural activity represents proximity relations among locations in physical space.[1]

Correlation and structural correspondence are the only two exploitable relations that Shea discusses, but he leaves open that other exploitable relations may generate content (43). The "vari" in "varitel semantics" reflects the variety of exploitable relations that can potentially generate content (along with the variety of possible task functions). The "tel" reflects the teleological cast of Shea's theory, as embodied by his appeal to naturalized functions.

Varitel semantics resembles teleosemantics as espoused by Millikan (1984) and others. However, varitel semantics improves considerably upon previous teleosemantic theories. For example, Millikan's theory entails that representation is present even in very primitive systems, such magnetotactic bacteria. Many critics regard this consequence as implausible, because one

[1] I note a significant typo that occurs in this chapter: fn. 23 reads "surjective" in place of "injective."

can satisfactorily explain magnetotaxis without any attribution of representational content (Burge, 2010, 300). Shea avoids attributing representational content to magnetotactic bacteria by imposing a *robustness* constraint on task functions, meaning roughly that the task function is achieved across a diverse range of proximal and distal conditions (55, 65). Magnetotactic bacteria do not satisfy the robustness constraint (203), so varitel semantics does not count them as representational.

Still, varitel semantics is very permissive regarding which systems it classifies as representational. According to Shea, a plant whose flowers open in the day and close in the evening could count as representing time of day, so long as the plant has *two* ways of detecting time of day and thereby satisfies the robustness constraint (214). Shea also says that "it would not be at all surprising" if the immune system turned out to employ representations (214). This permissive attitude extends to artifacts: Shea says that a simple thermostat can count as representing room temperature if it measures temperature in *two* ways and thereby satisfies the robustness constraint (211).

I am skeptical of these verdicts. Representational attribution does not seem to shed any light on plants, the immune system, or the simple thermostat considered by Shea. We might informally describe a plant as "representing" that evening has arrived, but this seems at best a *façon de parler* rather than the start of a compelling representational explanation. Similarly, I see no explanatory value in saying that Shea's simple thermostat represents room temperature, or that the thermostat enters into an internal state that is correct iff the room has a certain temperature. As Shea acknowledges (211), we can offer a satisfying explanation for the thermostat's behavior merely by noting how its internal states correlate with temperature, how its outputs are caused by those internal states, and so on, without any mention of representational

content or correctness-conditions. Representational attribution is absent from rigorous scientific description of a plant, the immune system, or a simple thermostat. We have no grounds independent of Shea's own theory for attributing representational content to these systems.

Shea (205) intimates that the dispute is a purely verbal one about how to use the word "representation." However, I think that more substantive matters are at stake. Most fundamentally, the issue is what mode of explanation we should employ.

Folk psychology attributes representational content to mental states, especially propositional attitudes such as belief and desire. These representational attributions draw explanatory and predictive power from their affiliation with a network of informal norms and generalizations, such as norms linking belief and desire to action. Cognitive scientists refine folk psychological explanation by articulating systematized successors to the informal norms and generalizations. A good illustration is *Bayesian cognitive science*, which uses Bayesian decision theory to model perception, motor control, and many other psychological domains. For example, Bayesian perceptual psychology posits that the perceptual system approximately executes unconscious Bayesian inferences (Rescorla, 2015). Bayesian perceptual psychology assigns a central role to representational mental states (e.g. prior probabilities regarding distal conditions) when explaining why we perceive the world as we do. These representational attributions draw their explanatory power from a theoretical framework that refines the norms and generalizations found in folk psychology. Likewise for representational explanation of motor control, mammalian navigation, and numerous other psychological domains.

In contrast, nothing resembling folk psychological explanation applies in an illuminating way to plants, the immune system, or the simple thermostat. Scientific explanation of these systems does not invoke norms or generalizations reminiscent of folk psychology.

Representational attribution is explanatorily idle here, because it floats free from the norms and generalizations that typically lend it explanatory value. I worry that Shea blurs important distinctions by extending representational attribution to so many cases where its distinctive explanatory benefits go missing.

In Chapter 8, Shea explores what he sees as the benefits afforded by representational attribution. He says that representational discourse has a characteristic "explanatory grammar": correct representation explains successful achievement of a task function; incorrect representation explains failure to achieve the task function (199). It seems to me that Shea's analysis does not fully capture the power of representational explanation, for two reasons. First, there are many cases where we adequately explain success (or failure) in achieving a task function without invoking correctness (or incorrectness). For example, we can explain why a simple thermostat succeeds or fails at maintaining constant room temperature without attributing correctness-conditions to the thermostat's internal states. We can instead adduce how well the thermostat's internal states correlate with room temperature. Second, representational explanation offers notable benefits quite independent of the "explanatory grammar" emphasized by Shea. We can explain an agent's actions by citing her beliefs and desires *regardless of whether her beliefs are correct or whether her actions lead to achievement of any overall goal*. We can cite a rat's spatial representations to explain why the rat chooses a given route through space *regardless of whether the spatial representations are correct or whether the rat's chosen route leads to a desired destination*. In both cases, and in numerous others, we invoke representational mental states to explain a mental or behavioral outcome without considering whether the states are correct and without considering whether the outcome promotes achievement of any task function. Our explanation hinges upon norms and generalizations that

refine those found in folk psychology, norms and generalizations that apply whether or not the representational state is correct and whether or not any task function is achieved.

Notably, Shea focuses almost exclusively upon fairly low-level mental processes that are not consciously accessible, such as perception, navigation, and motor control. He remains agnostic as to whether varitel semantics can accommodate belief, desire, and other high-level propositional attitudes (222-226). Propositional attitudes have served for millennia as the main wellspring of philosophical inquiry into mental representation (rivaled only by perception). Given that varitel semantics may not apply to propositional attitudes, and given that nothing resembling folk psychological explanation applies to many systems classified by varitel semantics as representational, I question how well varitel semantics captures the traditional notion of representational content.

That being said, varitel semantics fits beautifully with lots of theorizing in cognitive science, especially cognitive neuroscience. Neuroscientists often reason along the following lines: neural firing pattern *X* correlates with distal state *Y*; therefore, *X* represents *Y*. In some cases, these representational attributions serve merely as florid evocations of the correlation between *X* and *Y*. In other cases, though, they seem to reflect an enthymematic attunement to something like Shea's apparatus of task functions and exploitable relations. Thus, I believe that Shea has identified *one* scientifically important notion of "representation."

I have argued that this notion differs significantly from the traditional philosophical notion. Shea himself acknowledges that it may well do so (31). Accordingly, I recommend that we recognize two distinct notions of representation: the one that Shea emphasizes, which underlies some cognitive science discourse; and the one that figures in folk psychology, traditional philosophical inquiry, Bayesian cognitive science, and many other areas of cognitive

science. (Cf. Burge, 2010, 292-308.) There are cases where the two notions overlap (e.g. mammalian navigation), cases where Shea's notion applies but the traditional one does not (e.g. the simple thermostat), and perhaps cases where the traditional notion applies but Shea's does not (e.g. high-level propositional attitudes).

Overall, then, Shea successfully isolates a notion of representation that figures prominently in some areas of cognitive science (especially cognitive neuroscience), and he offers a compelling analysis of that notion. Even if I am right that Shea's favored notion of representation does not align with the traditional notion, he has done us a huge service by analyzing his favored notion with such care and rigor. I urge you to read this marvelous book so as to experience its intellectual riches first-hand.

Michael Rescorla

Department of Philosophy

University of California, Los Angeles

Works Cited

Burge, Tyler. 2010. *Origins of Objectivity*. Oxford: Oxford University Press.

Millikan, Ruth Garrett. 1984. *Language, Thought, and Other Biological Categories*. Cambridge: MIT Press.

Rescorla, Michael. 2015. "Bayesian Perceptual Psychology." In *The Oxford Handbook of Philosophy of Perception*, edited by Mohan Matthen, 694-716. Oxford: Oxford University Press.