**What really lives in the swamp? Kinds and the illustration of scientific reasoning**

Andrew Richmond

**Abstract**

It’s not clear what philosophers of science can learn from thought experiments. Consider Swampman: a physical duplicate of Donald Davidson that arises by chance after lightning strikes a swamp. Swampman is a popular counterexample to teleosemantics: he appears to have representation, but no selection history. So, apparently, it’s a mistake to define representation in selectional terms. Teleosemanticists respond that Swampman can’t tell us anything about representation because he’s simply not real, or even realistic: representation is a scientific kind, and if we take scientific kinds seriously, we can’t say that just because some imagined creature *looks like* typical representational systems, or could be explained in representational terms, it *is* a representational system. So Swampman isn’t a counter-example to the teleosemantic account of representational systems, because he isn’t an *example* of a representational system in the first place. I endorse this response to the Swampman counterexample, and especially its motivation: to take the scientific role of representational concepts seriously. But this motivation supports another way of understanding Swampman, according to which he is an *illustration* of scientific explanation, rather than an *example* of a representational system. I draw out the logic of this kind of illustration, compare it to some experimental paradigms in science, and argue that it provides a better way of understanding Swampman and other thought experiments in philosophy of science.

*If a random quantum fluctuation somehow created an iPhone*

*SE out of thin air it would still cost $579. Checkmate Marxists*

— Jonathan Weisberg, on Twitter[[1]](#footnote-1)

**1. Introduction**

The philosophical bestiary is a surreal place. You can’t see it all in one go, but on any given trip you’re liable to visit, among other things: disembodied minds (Descartes, 1984) and disembodied brains (Putnam, 1992); regretful vampires (Paul, 2014) and unconscious humanoids (Chalmers, 1996); perfectly choreographed simulations of brain activity performed by entire nations (Block, 1978) or by single, dedicated individuals (Searle, 1980); intergalactic duplicates of both earthly creatures and the Psych 101 textbooks describing them (Egan, 2014); and odd creatures called ‘Kimus’ making confused pilgrimage to the sunset, leaving their (just as adorably-named) predators behind (Pietroski, 1992).

 It is, for philosophers of science, not at all clear what we’re supposed to learn from the kind of creatures that populate this menagerie. It’s easy to think they are just too unrealistic or strange to shed any light on how science actually works. But we can’t seem to stop talking about them. Consider how often Swampman sightings are still reported in respectable journals, and by intelligent people — most of whom (I assume) are not believers in garden-variety monsters, ghouls, or cryptids (Kim, 2021; Peters, 2014; Porter, 2020; Schulte, 2020; Sebastián, 2017; Tolly, 2021). Swampman is an especially odd case, for a couple reasons. First, many of us think Millikan killed Swampman dead pretty much the moment he was conceived (Millikan, 1984). And second, her argument seemed to undermine not only Swampman, but *any* fantastical thought experiments like him, by showing exactly whythese kinds of thought experiments are irrelevant to our understanding of science.

 This paper will develop an account of thought experiments that shows how creations like Swampman *can* be relevant to our understanding of science. In §2 I’ll re-cap Millikan’s argument, and in §3 I’ll draw a distinction between the way Millikan conceives Swampman — as an *example* of a scientific kind — and another way we might think of thought experiments — as *illustrations* of scientific reasoning. I’ll show that illustrations don’t hinge on the assumptions Millikan criticizes, and I’ll reinforce this by comparing the logic of illustrations with some experimental paradigms in science. In §4 I’ll argue that Swampman can be fruitfully understood as an illustration. So understood, he is untouched by Millikan’s argument, and might have interesting things to tell us about cognitive science. §5 will conclude with some upshots of this way of thinking about thought experiments, for both philosophy of cognitive science and philosophy of science more broadly.

**2. Teleosemantics and the Swampman counterexample**

Teleosemantics is easy enough to summarize at a high level: *representation* is a natural kind, and what makes something a member of that kind is that it has a certain sort of selection history (Millikan, 1984; Neander, 2017; Shea, 2018).[[2]](#footnote-2) The view is applied to representation as it appears in folk psychology, philosophy, and cognitive science, but I’m focused on the cognitive scientific case, where authors like Neander (2017) and Shea (2018) use it to understand the scientific role of representational concepts. It’s worth mentioning that there are similar views of *computation* in philosophy of cognitive science (Milkowski, 2013; Piccinini, 2015), and of course selectional notions of *function* that are important for cognitive science, even if they are mostly discussed in the context of biology(Garson, 2019; Neander, 1991; Wright, 1973). My discussion will bear on these views as well, but I’ll focus on teleosemantics.

 Swampman, too, is fairly easy to describe: imagine a creature physically identical to Donald Davidson, but with no selection history. E.g., imagine he’s created by a freak chemical reaction when lightning strikes a swamp. Of course (the argument goes) this creature would have representations. He would see, he would plan, and he would wonder how he ended up in the swamp (being identical to Davidson at the moment of his creation, he will come into existence with the thought that just a moment ago he was very much *not* standing in the middle of a swamp). We can ask him to deliberate over a career change: keep prowling swamps for a stable wage and good benefits, or try for a riskier but potentially more fulfilling career as a philosopher? He can weigh the pros and cons, the state of the job market, the possibility of finding himself redundant if swamp-prowling is automated, and so on. And, since he is a perfect copy of Davidson, he will give thoughtful and interesting answers to all these questions. The Swampman argument points to this paradigmatically representational activity and concludes that Swampman is a counterexample to teleosemantics: he has representations but no selection history.[[3]](#footnote-3)

This is tidy argument, but it was met pretty much immediately with what many of us see as a conclusive response. Scientific kinds are more sophisticated than the Swampman argument supposes — they are what Millikan calls *real kinds*:

*Real kinds* I define as groups over which a variety of relatively reliable inductions can successfully be run *not accidentally but for good reason.* The essence of a real kind is whatever accounts for its instances being alike. (Millikan, 1996)

Why should we think of scientific kinds this way? Because this is what science needs; mere similarity doesn’t serve its purposes (Neander, 1996, p. 120; and Shea, 2018, pp. 22,  28-29, are especially clear about this motivation). And creatures as fantastical as Swampman don’t count as members of the same real kinds as actual organisms: Swampman is similar to representational systems, just like he is similar to members of the kind *homo sapiens* — but not for good reasons(Millikan, 2010; Neander, 1996). There is a good reason that *real* representational systems like you, me, and other organisms, are similar: our similar evolutionary histories. And there is *a* reason, but not *the same* one, that Swampman is similar to those organisms: philosophers wanted to describe something that looked as superficially similar to them as possible. So Swampman has some of our features, but not for the same reasonswe have them, and therefore he does not belong to the same kinds as us — at least the kinds that are held together by something deeper than mere similarity. Representation is such a kind, so Swampman is not a member of the kind *representational system* any more than he is a member of the kind *homo sapiens*. And in case the idea of kinds as induction-generators seems outdated, that’s fine: we don’t need to adopt Millikan’s particular account of scientific kinds. Her deeper point, elaborated especially by Shea (2018, pp. 28–29), is that our intuitions about which kinds Swampman belongs to should be overridden by facts about *what scientific kinds must be, to serve their scientific roles*. If we want to claim that Swampman represents, out intuitions about him are next to worthless.

Is there a better reason to think Swampman represents? One candidate is that the best explanations of Swampman would attribute representations to him (Neander, 1996, p. 123). But the same idea can be applied here: our theory of a scientific kind is beholden to its scientific role. Like other scientific kinds, the kind *representation* has a role in *explanations of real worldly patterns*. And Swampman isn’t an instance of the worldly patterns cognitive science tries to explain, any more than he’s an instance of the pattern biologists are trying to explain when they study the kind *homo sapiens*. It might be useful for doctors to treat Swampman as a human being (Antony, 1996), and, for the same sort of reason, it might be useful for cognitive scientists to treat him as a representational system. It might help them give good explanations. But that doesn’t make Swampman a part of the worldly pattern that cognitive science is aiming to capture and explain when it uses the kind *representation*. And, therefore, he is still an inappropriate source of evidence aboutthat pattern or kind (cf Neander, 1996).[[4]](#footnote-4)

To sum up, Swampman is not an example of a representational system, so he isn’t a *counter-*example either: his lack of a selection history doesn’t tell us anything about whether representational systems can lack a selection history. The deeper problem is that the Swampman argument doesn’t take scientific kinds seriously enough. Scientific kinds are properly defined by their roles in actual science, and by the features they must have to play those roles. Therefore, *if we want to derive the nature of a scientific kind from examples of it, we need real examples of it* (or at least realistic ones). That, I take it, is the enduring lesson of Millikan’s response to Swampman. As she put it: representation, like any other *real kind*,is “a case where the a priori method of example and counter-example fails” (Millikan, 2010, p. 79; see also Papineau, 2001).

I think Millikan is right: trying to derive the nature of representation from Swampman, understood as an example of a representational system, is a mistake. And I think she’s right about why: that approach does not take scientific kinds seriously enough. Our intuitions about Swampman don’t tell us much, if anything, about the features that the kind *representation* must have in order to play its scientific role. But I want to reconsider the kind of argument Swampman is supposed to be. I’ll argue that Swampman should not be understood as an attempt to derive the nature of a scientific kind from an example of it, but as an attempt to do exactly what the teleosemanticist urges: to probe scientific explanations, the role that the kind *representation* plays in them, and the sort of features that kind must have to play this role. We can use Swampman to probe scientific explanations just like we use cardboard cutouts to probe anuran prey capture, without any commitment to those cutouts being *real*, or even realistic, members of the kind *frog prey* (Wachowitz & Ewert, 1996). The next section will spell this out in detail.

**3. Examples vs illustrations**

The *real kinds* response assumes that the Swampman argument depends on an intuition that Swampman is a representational system. In §3.1 I’ll develop a simple thought experiment, analogous to Swampman, that works differently. I’ll call this sort of thought experiment an *illustration*. In §3.2 I’ll compare the logic of illustration with the logic of some experimental paradigms in cognitive science. Partly, this is building to the argument in §4 that Swampman is best understood as an illustration. But it also stands on its own as an account of the way that thought experiments — even fantastical ones — can serve philosophy of science.

 *3.1 An example of illustration*

Imagine a student in a physics class who expresses the following misunderstanding: “Explanations in quantum mechanics appeal to observers, so quantum mechanical effects rely on conscious observation. Therefore [some woo].” Though this is wrong, observation is a difficult and contested notion in physics, and it might be impossible to point the student to a passage in the textbook with a clear definition of the term. A better tactic would be to show the student that quantum mechanical (QM) explanations work, predict, and explain experimental results even if any potential conscious observers are, say, looking away from the experiments. If the student’s misunderstanding persisted (“But we set up the device that does the observation, and *we’re* conscious!”) you might respond with a more extreme hypothetical: a universe without conscious observers at all. We can imagine a physics lab popping into existence in that universe, arranged so that it sets into motion a classic QM experiment. Our QM explanations, models, and predictions, including everything they say about observation, would still succeed there, and in all the same ways they succeed in the actual world.

All we’re doing here is paring away an irrelevant feature (consciousness) to show that it is, in fact, irrelevant. We’re hoping the student follows us in an inference that goes something like this: if QM explanations go through just fine, and for all the same reasons, whether consciousness does or doesn’t exist, then those explanations can’t rely, for their success, on consciousness. Therefore consciousness cannot be an essential feature of any kind the explanations invoke, like *observation*. Now, does it matter that consciousness actuallyexists? That the universe we’ve described, with a physics lab but no consciousness, is wildly unlikely? It’s surely not an instance of the same worldly pattern as our own world, which developed on a large scale according to dynamical laws — not by the random coming-into-existence of immensely complex and apparently goal-directed things like physics laboratories. A world with physics labs but no consciousness might even be *impossible*, on certain (niche) views of consciousness (Goff et al., 2022; Tononi et al., 2016). But none of this matters because you’re not using the thought experiment, in the first place, as an example of the kind *observation*, and then deriving the nature of the kind from that example. In the terminology I’ll use, you’re using the thought experiment to *illustrate* a bit of scientific reasoning, by paring away a feature to show that it is irrelevant to the success of the explanations.[[5]](#footnote-5) To return to the main lesson of Millikan’s response to Swampman, note that in the QM case we are starting with actual scientific explanations, and we are probing how they work in their actual contexts. We’re doing that by paring away features of those contexts and checking the consequences for the explanations’ success. That does involve a hypothetical non-conscious world, but we’re not starting from an intuition that the non-conscious world is a member of some kind, and nor are we saying that, because we would explain it in terms of some kind, it must be a member of that kind. We’re probing the role of consciousness in the actual explanations by asking how the presence or absence of consciousness would affect the success of the explanations. Kind-membership only comes in as a further step, once we’ve shown that consciousness is not relevant to the success of QM explanations: *then*, if we’re paying attention to the teleosemanticist’s stricture that the philosophical account of a scientific kind should be determined by its scientific role, then the kind *observation* should not be defined in terms of consciousness.

*3.2 The logic of illustration*

The logic of illustration is fundamentally no different than the logic you might use in trying to understand how any system performs a task — it’s just that in this case the system is science, and the task is explanation. We pare away parts and ask whether the system is still able to perform the task, to the same degree and in the same way. If it *is*, then the part you removed wasn’t contributing to the task in the first place. E.g., if you want to understand how a can-opener works, you can determine that its cosmetic properties are irrelevant by paring them away (either really or imaginatively), and seeing that this has no impact on its ability to open cans.

But I’ve aligned myself with the side of this debate that demands we take scientific explanation more seriously; I don’t think anyone will be satisfied by a comparison between explanations and can-openers. So I want to flesh out the logic of illustration by comparing it to two experimental paradigms in cognitive science. These comparisons will reinforce the point that illustrations don’t make any problematic assumptions about kind-membership. And they will support the *legitimacy* of illustration — at least, insofar as the paradigms I compare it to are legitimate — and allow me to introduce some complications and challenges.

To set the stage for these comparisons, we can break the logic of illustration down into three components.

1. There is a *broad explanatory target*: we’re trying to understand how a form of explanation (e.g., QM explanation) successfully explains events and patterns.
2. We *narrow this target* to ask about a feature of the explanations (e.g., the consciousness of observing devices) — does that resource contribute to the explanations’ success?
3. And we *probe* this question by paring that resource away (e.g., imagining an unconscious lab) — what would happen to the success of QM explanations when that resource is missing?

How does this compare to standard cases of scientific reasoning? On a pretty uncontroversial characterization, cognitive science is “the study of how agents perform tasks” (Mekik & Galang, 2022, p. 2), and specifically how they use different resources to perform them — both internal resources (like neural structures or activity patterns) and external ones (like features of the environment).[[6]](#footnote-6) And to answer these questions, cognitive science routinely performs experiments thatremove resources and examines the effect on task performance. I’ll discuss one paradigm that removes internal resources (specifically, brain areas) and one that removes external resources (features of the environment).

Let’s start with the former. The logic of *ablation studies* received a lot of attention in philosophy and psychology in the last century. Many of the lessons are now received wisdom (e.g., von Eckardt Klein, 1977). Ablation studies investigate the role of a brain area in some task by either ablating that area, or finding organisms in whom it has been ablated naturally, e.g., by strokes or railroad spikes (Damásio et al., 1994; Salvalaggio et al., 2020).[[7]](#footnote-7) Classical work found, e.g., that “bilateral lesions to lateral occipital–temporal cortex could lead to impairments in recognizing objects but no difficulty performing grasping and reaching movements to the same objects,” prompting the inference that the lesioned area was involved in one task, but not the other (Mahon & Hickok, 2016, p. 942). Current work uses tools that create more carefully targeted ablations (Liu et al., 2019; Zhang et al., 2021), or that temporarily disrupt activity in a brain area, assuming that the area is essentially incapacitated by the disruption (Weissman-Fogel & Granovsky, 2019). But the inferences are fundamentally of the same form as they’ve always been: if an area or its activity can be eliminated without affecting task performance, it must not have been used to perform the task in the first place.[[8]](#footnote-8) That logic comes with important caveats, which I’ll discuss shortly, but it bears comparison to the logic of illustration:

1. There is a *broad explanatory target*: we’re trying to understand how an organism (e.g., a human being) successfully performs certain tasks.
2. We *narrow this target* to ask about a feature of the organisms (e.g., a particular brain area) — does that feature contribute to the organism’s performance?
3. And we *probe* this question by paring that feature away (literally ablating it) — what happens to the organism’s performance when it doesn’t have that feature?

Recall that there were two things I wanted to do with these comparisons. First, I wanted to make it clear that this logic does not require any illicit assumptions about the kinds their target organisms are examples of. To see this, note what the experiments are trying to explain: just as an illustration asks what resources *real* scientific explanations use (not just what resources the *imaginary* one uses), ablation studies generally conclude that *intact* organisms of a certain species do or don’t use a brain area in performing a task (not just that the *ablated* organism does).[[9]](#footnote-9) So the conclusions aren’t based on an assumption that the ablated organisms are members of the kind we’re trying to understand. In fact, we know they aren’t. The ablated organisms aren’t intact, and they often aren’t even the same species as the organisms we’re trying to understand (e.g., Weiskrantz et al., 1974, p. 709). There is even talk of drawing conclusions about human brains from ablations in artificial neural networks (Lillian et al., 2018). So, clearly, the conclusions about intact organisms of a particular species do not derive from the ablated organism *also* being an intact organism of that species. The conclusions derive, instead, simply from evidence that the two organisms (the ablated one, and the intact ones that are our ultimate target) are performing the task the same way, i.e., using the same brain areas or structures. To see this more clearly, imagine trying to give the *real kinds* response here. You would be saying, *regardless of whether the ablated organism and the intact one perform the task the same way*, the experiment doesn’t tell us how the intact organism performs the task, because, in the end, we only studied a non-intact organism. If you were worried that scientists were sniggering about Swampman (Millikan, 1996, p. 115), imagine how they’d respond to an argument like that.[[10]](#footnote-10)

 The second thing I wanted to do is highlight challenges that to apply ablation experiments and maybe, by extension, illustrations. The main type of challenge to ablation experiments concerns the possibility that an organism learned to perform the task in a new way, or that its brain compensated for the missing area somehow. Maybe the ablated area *is* used in the task, but it’s redundant. Maybe another brain area took over its role. The point is that a *lack of change in task performance* does not entail a *lack of change in the resources used* (e.g., Barsalou, 2016, p. 1128). I won’t rehearse the many examples of this here. Suffice it to say, ablation studies have a difficult problem: they have to make the case that the organism under study really is doing the task the same way as their target (intact) organisms. In practice, this means ruling out plausible learning or compensatory mechanisms, e.g., by eliminating the necessary time for learning or plasticity, or showing that brain areas that might be expected to compensate don’t actually change their activity patterns after the ablation. Legitimacy, for an ablation study, means doing this convincingly. And it will mean the same for illustrations. If you use the illustration I described in the QM case, you need to rule out plausible ways the explanations might have succeeded even if they originally *did* rely on the consciousness of the observer. Perhaps another concept steps up and does extra explanatory work, or the explanation implicitly appeals to new facts/resources that it didn’t before.

The second type of experiment I promised to discuss is one that pares away an organism’s *external* resources, namely, features of its environment. One note before jumping in: these studies often pare away many features at a time. The perception of motion, e.g., might be studied by paring away almost every aspect of a typical moving environment, leaving only a simple shape displayed briefly in one location and then another (e.g., Ramachandran & Anstis, 1986). There are compelling objections to that kind of work (Gibson, 1972; but see Shepard, 1984 on remaining roles for these experiments), but they don’t apply when we’re just paring away a single feature of an organism’s environment, or a few.

 There is a particularly clear description of this experimental paradigm in a paper that teleosemanticists are well aware of, investigating prey-capture in frogs (Lettvin et al., 1959). The set-up of that paper tells us that a frog “will starve to death surrounded by food if [the food] is not moving” (p. 1940). Eliminate motion, and prey-capture is affected. So motion seems to be one of the environmental features prey-capture depends on. But the frog “will leap to capture any object the size of an insect or worm, providing it moves like one. He can be fooled easily not only by a bit of dangled meat but by any moving small object” (p. 1940). Eliminate any of the prey’s features aside from its physical dimensions and movement, and you won’t see an effect on prey-capture. So prey-capture does not depend on those other features. The logic is even more transparent here than it was in the case of ablation:

1. There is a *broad explanatory target*: we’re trying to understand how an organism (e.g., the frog) performs certain tasks.
2. We *narrow this target* to ask whether the organism relies on a particular resource (e.g., a particular feature of the object) — does that resource contribute to the organism’s performance?
3. And we *probe* this question by paring that resource away (either actually stripping the feature from the object or constructing an environment lacking that feature) — what happens to the organism’s performance when the object in question has or doesn’t have that feature?

 In the experiments Lettvin et. al. performed, a number of stimuli were chosen that had various features in common with flies (a typical prey item for the frog). And a number of nerve fibers, assumed to drive prey-capture behavior, were recorded to see which stimuli caused a change in their response and which didn’t. E.g., eliminating all the physical features of a ‘fly’ except its rough size and shape did not stop the fiber from responding, but eliminating its movement relative to a background *did* (p. 1945). The point is that, assuming that these nerve fibers’ response is what drives prey-capture, the physical features of a fly aside from its rough size, shape, and movement patterns do not contribute to the prey-capture task.[[11]](#footnote-11)

 To be fair, these experiments don’t exactly represent the state of the art. Neander (2017) brings out the increasing complexity of anuran prey capture, and gives a more current picture of the research. But much of that research follows the same logic, with the essential finding being that “the relevant visual discrimination in an unconditioned toad is largely unaffected by features not captured by [a small number of] dummy stimuli” — the infamous cardboard cutouts (Neander, 2017, p. 104). That finding is so important because, by telling us which features can be eliminated with no effect on prey capture (or prey detection, which Neander focuses on), it tells us which features are and aren’t used in that task.

 How would the *real kinds* response fare against this body of work? About as well as it did against ablation studies. Here we are using non-naturalistic environments, including some wildly unrealistic ones — e.g., cardboard cutouts standing in for worms — to probe the way that anuran prey capture works in real, natural, environments. We might object that prey capture works differently in different environments, but these objections would rely on *facts about how prey capture works in those environments*. They would not go, “Well, these are very nice experiments, but unfortunately you’ve made a terrible error: the non-naturalistic environment isn’t of a kind with the naturalistic one! Shame you spent so much time on this study; it’s worthless.” The reason this is so ridiculous is that these studies put no evidential or epistemic weight put on the experimentalstimuli (cardboard cutouts) being *of a kind* with the stimuli we’re trying to draw conclusions about (worms), but only on the fact that the frog is using the same resources in the two environments. I’m not trying to mock Millikan here. She didn’t give this response to this sort of experiment. The point is that she was right not to: it would have been a mistake to think that the logic of these studies — and, by extension, the logic of illustration — relied on shared kind-hood in the relevant way.

 To come to the complications this comparison raises: it *is* possible to argue that prey capture works differently in naturalistic and non-naturalistic environments. Ecological psychologists have been making points like this for decades: in a highly impoverished (experimental) environment, an organism may perform its tasks differently than it does normally, when it has more environmental information to take advantage of. This is a harder case to make when we’re only eliminating one or a few features of the environment, but in any case, the task for the frog researcher is the same: to show that the *potential* differences in how prey-capture works are not *actual*. For example, consider Hartle & Wilcox’s (2016) study of stereopsis. They were investigating the role of different binocular cues in depth perception, and, chasing down of some unexpected patterns in the data, they realized that participants had (unknowingly) discovered *monocular* cues in the images, which are absent in normal environments. If Hartle & Wilcox hadn’t discovered this, they might have found that they could eliminate certain binocular cues without affecting depth perception. But this is only because their experimental stimuli hadn’t just removed a feature of the environment: they had *added another*. So the inference from *how depth perception works given the experimental stimuli* to *how it works generally* would be undermined. This is just to say that when an experiment is paring away feature of the environment to check the effects on task performance, it requires a plausible argument that the environmental variations introduced don’t induce the organism to perform its task in a new way, e.g., by unintentionally introducing new resources for it to take advantage of. Likewise, illustrations will need to show that *their* manipulations don’t have unintended effects on the way scientific explanations work or the resources they have access to.

**4. Swampman redux**

In §4.1 I’ll apply this understanding of thought experiments to Swampman. Understood as an illustration, Swampman escapes Millikan’s objection for the same reason as the two experimental paradigms I just discussed. In §4.2 I’ll talk about remaining objections to Swampman.

 *4.1 Swampman as an illustration*

This argument will sound familiar. It’s supposed to. I’m trying to capture the argument that philosophers ought to be making when they talk about Swampman, which is different only in small, but essential, ways from the argument that they actually make, and that teleosemanticists respond to. Let Swampman be generated in a swamp just as normal. We can imagine him wandering into a university building and seeing posters advertising Calls for Participants. We can imagine that he signs up for a study and lands in a psychology or neuroscience lab. Since he is a molecule-for-molecule copy of Donald Davidson, he will display all of Davidson’s cognitive capacities. Or, at least, he will display all those capacities insofar as they are described without reference to history, or to any properties that rely on one’s being the ‘real’ Davidson. Cognitive scientists study, e.g., *memory* and *accurate memory*, but I don’t think anyone will argue that what they mean by *accurate* is such that Davidson has the capacity to accurately recognize his mother, but Swampman doesn’t, because she didn’t actually give birth to him. That kind of ‘capacity’ might be interesting, but it is outside the remit of cognitive science. I take it as given that, as far as cognitive science is concerned, Davidson’s and Swampman’s cognitive abilities, like memory, are the same, and that the interesting role for selection history is as a *resource* to explain those abilities.[[12]](#footnote-12) Louise Antony makes a similar point about biology, insofar as it’s motivated by medical purposes: an oncologist would be just as interested in treating a patient whether they came from the Swamp or from Dallas (1996, p. 72). The question is about what resources she would rely on, and, specifically, whether those resources would include selection history.

The first major step of the Swampman illustration is a simple disjunction: Swampman’s display of the capacities of interest to cognitive science is either explicable or inexplicable. It’s not his history that’s up for explanation: we’re not asking whether the way Swampman *came by* his capacities is explicable. Our topic is the kind of capacity that cognitive scientists aim to explain. We either can or can’t reverse-engineer Swampman’s cognitive capacities. We either can or can’t show how his physical organization supports his behavior. What would it be, exactly, to deny that Swampman is explicable? If we sat Swampman down in the laboratory, would we be stymied-in-principle by his behavior? Would it be impossible to model the causal structure of his brain at the levels of grain that allow us to predict and explain his actions? Or would this project at least be less successful than it is with Davidson? There are no tricks up Swampman’s sleeve, here — he’s just another physical system. I don’t see any way of denying that his various capacities would be as explicable as the capacities of any system. For anything he can do, there must be an explanation of how he does it.

The next step is to ask how we would explain Swampman’s capacities. Would we have to use explanatory methods, strategies, or resources different than the ones we use to explain Davidson? I don’t see any way of supporting that view either. We would observe Swampman the same way we observe organisms whose evolutionary history is unknown to us. We would see patterns in his behavior: a tendency to forage his environment in a certain way; an ability to learn patterns in sets of stimuli; and so on. And we would investigate those behavioral patterns with questionnaires and response-time measurements, black-box models and eye-tracking experiments, computer simulations and fMRI data, circuit diagrams and electrode recordings, and so on. We would apply the same probes and explanatory resources that we do in cognitive science more generally. And we would take the same approaches to building models of Swampman — including the use of representational notions. To illustrate this, imagine a ‘meta-experiment:’ a single-blind trial where the participants are *two cognitive neuroscience labs*. We send Davidson off to one lab, and Swampman to the other. But we’ve cleaned Swampman up and washed off the swamp goo, so the scientists can’t tell who is Swampman and who is Davidson. *If we think that the success of cognitive scientific explanations relies on their targets’ selection histories, we must think they would fail in some respect when applied to Swampman, but would be successful in that respect when applied to Davidson.* (Recall the logic of illustration and the scientific paradigms I compared it to in §3.2.) What respect could this be? The models we would construct would be equally predictive of each’s behavior, as revealing of its neural basis, as useful in medical interventions, and so on. Nothing about Swampman, his swamp-brain, or his swamp-engagement-with-his-environment would seem to impede those projects, any more than our typical ignorance of an organisms’ evolutionary history impedes us.

Of course, there is always *some* role for selection history: in Davidson’s case but not Swampman’s it may be possible to use evolutionary history to suggest hypotheses about him, including about what he represents, or to connect cognitive science’s understanding of him to that of evolutionary biology (e.g., Cisek & Hayden, 2022). But teleosemantics makes a stronger claim: that *the kinds used in scientific explanations* are defined in terms of selection history. We can accept that, with selection history ‘ablated,’ hypothesis-generation looks different. The question is whether *explanations* do.

And, of course, one class of explanation *do* look different: successful explanations of how Davidson *came by* his capacities would not succeed in Swampman’s case. But no one has doubted that history was relevant to that kind of explanation: the question itself is *about* history.[[13]](#footnote-13) That’s why I’ve stipulated that the labs are investigating Swampman’s current abilities: the teleosemanticist’s claim isn’t anything as banal as the idea that something’s history is relevant to explaining its history. The teleosemanticist’s claim is that explanations of current abilities like prey capture (Neander, 2017, Chapter 5), navigation (Shea, 2018, Chapter 5), and bee dances (Millikan, 1984, Chapter 5) rely on a notion of representation that is, itself, defined in terms of selection history. So it’s this kind of explanation that Swampman is meant to probe.

To come to the point of the Swampman illustration: the kind of explanations teleosemanticists are interested in — the kind that, they think, require a selectional notion of representation — actually do not rely on selection history, because selection history can be pared away or ‘ablated,’ and the explanations will still work just as well, and in all the same ways. So, if we agree with the teleosemanticist that our account of a scientific kind should capture its role in scientific explanation, our account of *representation* should not define it in terms of selection history, for the same reasons that our account of *observation* should not define it in terms of consciousness.

It’s certainly possible that cognitive science would not meet some explanatory goals when confronted with Swampman. I’ve mentioned prediction, modeling, understanding, and intervention. Is there another goal cognitive science has for its explanations that will expose a difference in explanatory success between Swampman and Davidson? If there is room to push back here, it is the burden of the teleosemanticist to show how our explanations of Swampman’s capacities would fail, and (therefore) how selection history is relevant to those explanations.

 Before I move on, what about the challenges we saw above for experimental paradigms with the same logic as illustration? First, we have to rule out *compensation*, where some other feature of Swampman or the lab explaining him does extra work when his selection history is absent. It’s not clear what this would be. Does the lab that receives Swampman have to put extra emphasis on, say, brain data? Will that lab make additional modeling assumptions? There doesn’t seem to be a plausible compensatory mechanism, especially since, in the hypothetical, the two labs don’t know who got Swampman and who got Davidson. Second, we have to rule out any unintended effects of removing Swampman’s selection history — especially any resources that might *introduce*. But there don’t seem to be any plausible candidates for this objection. Swampman will have had a very different day than Davidson, who woke up that morning in a bed, not a swamp. But that kind of thing wouldn’t seem to explain why the Swampman lab reaches the conclusions about their subject’s brain it does (the same conclusions the Davidson lab reaches). Regardless, I’m happy to leave the conclusion somewhat tentative, until teleosemanticists have had a chance to dissect the illustration themselves and propose whatever compensatory mechanisms or unintended effects they think might have made a difference.

*4.2 Objections to Swampman, even as an illustration*

There is a complication that I want to discuss briefly before moving on to some more concrete objections: teleosemanticists might accept that our explanations of Swampman work just as well as our explanations of Davidson, and for the same reasons, but argue that some *broader* explanatory goal would be undermined if, because of this fact, we defined representation in non-selectional terms. E.g., it is sometimes suggested — or, more accurately: stipulated — that cognitive science is in the business of generalizing over species-kinds (Millikan, 1996, p. 109), or maximizing the reliability of its inductions (Millikan, 1996, p. 108) or the generality of its theories (Neander, 1996, p. 123), and that this makes selectional kinds necessary. But these are dubious characterizations of cognitive science as a whole, and it is far from clear that non-selectional kinds would fail to support the necessary generalizations and inductions (Shea, 2018, p. 22). I’ve focused on explanations of Swampman’s particular capacities because these explanations can be probed in detail, and involve less hand-waving than claims about the explanatory goals of cognitive science as a whole. Regardless, if teleosemanticists want to make that sort of argument, they will have to confront illustrations much like the one here. (E.g., how would cognitive science as a whole be any different if some or all of its target systems lacked selection histories?).

 I want to close out the discussion of Swampman with some objections aside from Millikan’s *real kinds* response. First, consider the teleosemanticist who defines selection to include, e.g., learning, differential survival, and other short-term processes (e.g., Garson & Papineau, 2019; Millikan, 1984; Neander, 2017; Shea, 2018). These teleosemanticists say that Swampman would have a selection history (and therefore, potentially, representations) *almost* as soon as he emerged. That means they’re not even appealing to the ‘ablated’ resource — evolutionary selection history — and they can accept that even representation explanations of Swampman would be successful. But they *will* have to predict that the explanations of Swampman would succeed in a radically different way than they do for Davidson: the explanations of Davidson would successfully appeal to a larger and richer set of representations, grounded in his evolutionary history. And that runs afoul of the illustration I’ve given, where explanations of Swampman seem to work identically to explanations of Davidson. We *could* try restricting explanations of Davidson too, so that evolutionary selection doesn’t figure in there either — just learning, differential survival, etc. But I don’t know anyone who holds that view: a teleosemanticist who appeals *only* to short-term functions is almost as rare a sighting as Swampman.[[14]](#footnote-14)

 Next, isn’t Swampman (I don’t know how to put this delicately) just a bit ridiculous? Leave aside the philosophical niceties, aren’t the scientists laughing at us (e..g., Dennett, 1988, 2007; Millikan, 1996)? Shouldn’t our response to things like Swampman be simply, “Bah! We’re doing serious work here — leave us alone.” But, as the study of frog vision should make clear, legitimate scientific inquiries also put their target systems in ridiculous scenarios made up of fake organisms. (Or consider the virtual realities that laboratory mice and flies spend so much of their time in.) *What matters is that the deviations from reality are motivated.* Cardboard cutouts are far less similar to worms than Swampman is to Davidson. But the cutouts are relevant to real prey-detection (the detection of real worms) because they’re tailored to investigate hypotheses about real prey-detection. What I’ve been arguing is that Swampman is relevant to scientific explanation for the same reason: because he’s tailored to investigate hypotheses about it, even if he isn’t an example of it.[[15]](#footnote-15)

 A more ecumenical response to the ‘bah’ objection might try to bring Swampman down to earth, using real organisms to make a similar point. E.g., what about organisms with evolutionarily novel traits, which haven’t had a chance to be selected for yet (Peacocke, 2014; Peters, 2014; Porter, 2020; Walsh & Ariew, 1996; see also (Mendola, 2009)). The problem with these cases is that they can be nit-picked to death. Sure, the trait is evolutionarily new, but is there some broader selected mechanism that confers content on it? Might it have derived content? Could it have an evolutionary pre-cursor of a similar enough kind for it to count as selected for? So, e.g., Porter (2020) argues that brand new color vision capacities and the mechanisms supporting them — i.e., ones with no selection history yet — represent. But a teleosemanticist can suggest that it’s not particular color representations that are selected for, but a *general* capacity for color representation, and individual representations get their content from their place in the whole system of color representations (Neander, 2017). Or they can say that teleology endows *learning mechanisms* with a function-conferring ability (Garson & Papineau, 2019), which endow the color representations with functions. And now we are debating how exactly brain mechanisms evolved, and endlessly fiddling with ancillary hypotheses that might let us say certain novel mechanisms do have functions of some variety. The reason we use Swampman is the same reason scientists studying anuran prey capture use cardboard cut-outs rather than real worms in some sort of motion-restriction device: it affords them control over *exactly the features they want to manipulate*, and keeps the investigation from being swamped by irrelevant details.

 A related worry is that the explanations of Swampman are only successful because we made him up to be as similar to Davidson as possible: explanations of Swampman take advantage of *pre-existing* explanations of Davidson. But consider the frogs again. The cardboard cutouts are created precisely to take advantage of the way the frog’s visual system responds to real worms. They are *supposed* *to* be similar to real worms except in a particular feature; that is what allows us to infer, from the fact that they provoke a stereotyped response, that this response doesn’t rely on the missing features. Here again, Swampman is in no worse shape than your typical anuran vision experiment.

 Another objection might be that illustration can get out of hand. Imagine *Swampswamp*: a whole swamp generated by a lightning strike. Would that hypothetical show that etiology is irrelevant to ecology, since ecology would still be able to explain the Swampswamp?[[16]](#footnote-16) Which properties can be ‘Swampmanned away’ depends, of course, on the scientific context: the explanandum, the type of inquiry involved, the specific mode of explanation used, and so on. So the Swampman strategy for excluding etiology is not as radical as it might seem — etiology remains explanatorily relevant where the explananda are historical, or where the mode of explanation, for any number of reasons, relies on historical kinds or facts. If we were interested in how ecosystems develop, a whole swamp generated by a lightning strikewould not compose a Swampman-style argument because it would erase the explanandum. If we were just interested in how equilibriums are maintained once they’re created, then historical kinds in the particular explanations of that particular explanandum may be targeted by a Swampman-style argument, so long as they really do not invoke historical kinds or trends. The moral is, really, not radical at all. It is just that etiology is, like everything else, a genuine part of an explanation only when it *contributes* something to that explanation. And we can see whether/what it contributes to an explanation by taking it away, and examining the effects on explanations’ success.

**5. Upshots**

Let me issue a brief reminder, which will also give me a chance to recap the argument: I haven’t been arguing that *since Swampman is explicable in representational terms, he has representations*. As I discussed in §1, the teleosemanticist would respond (I think correctly) that this argument doesn’t take the scientific role of the kind *representation* seriously. It’s analogous to arguing that since we explain misinformation as a sort of virus, biology needs a new definition of the kind *virus* that subsumes misinformation. My argument has an intermediary step: showing that cognitive scientific explanations of Swampman would be *just as successful* as they are of cognitive science’s paradigmatic targets, and *in all the same ways*. If we could make a parallel argument about virus-based explanations — that they explain misinformation just as well, and in all the same ways, as they explain biological viruses — then we would have shown that the scientific role of the concept virus doesn’t have much to do with biology. (That kind of argument is not forthcoming, for obvious reasons.)

This extra step is what makes the Swampman illustration a way of probing actual scientific explanations and the resources they rely on. As I put Millikan’s point above, she was saying that (1) to derive the nature of a scientific kind from examples, we need *real* examples of it, or at least *realistic* ones, because (2) we need to probe scientific kinds’ role in actual science. But (1) I’m not deriving the nature of a scientific kind from examples of it, but from a consideration of *precisely* (2) its role in actual scientific explanations, and specifically the resources those explanations rely on. And I’m deriving its role in actual scientific explanation using the same logic that scientists often use to study the role of different resources in cognition. The point is that Swampman — as strange a beast as he is — can be informative about scientific explanation as long as our *use* of Swampman is grounded in a legitimate investigation scientific explanation.

By now, I have taken up a considerable amount of your time trying to rescue a member of philosophy’s ridiculous bestiary from extinction. Surely, I owe you some implications. I’ll draw out two implications for philosophy of cognitive science, and two for philosophy of science more generally.[[17]](#footnote-17)

Let’s start with philosophy of science generally. The first implication is that, when we look at the special sciences, we should embrace the task of *illustrating their explanatory strategies*, rather than just, or first of all, *characterizing their kinds*. In fact, in illustration-style arguments, the nature of scientific kinds shows up as a sort of afterthought — the nature of the kind might not have very much to tell us once we understand the explanation and the resources it relies on (cf. Richmond, 2022, 2023). This is in sharp contrast with the way philosophers tend to investigate the sciences, which is to understand their explanatory strategies by *directly* characterizing their kinds, without first drawing out the structure and logic of the explanations those kinds figure into. This is currently the dominant approach whether the kind in question is *function* (Garson, 2019), *gene* (Griffiths & Stotz, 2008), *representation* (Shea, 2018), *computation* (Shagrir, 2022), or something else.

Second, there is a whole bestiary of strange and wonderful creatures that can be revisited in light of the distinction between examples and illustration. As long as they are used carefully, thought experiments have the same legitimate role to play in philosophy of science that ablation studies or environmental manipulations do in neuroscience and psychology. This is not to say the whole bestiary is welcome! Philosophical zombies, e.g., don’t seem to have anything to say about scientific explanation, except perhaps to make the obvious point that when the science of consciousness relies on self-report it is relying on a *proxy* for consciousness (scientists would treat the zombie’s self-reports just like they treat ours). But we should remain open to thought experiments understood in the way I’ve described. We should dismiss them if we find they have nothing to tell us about scientific explanation or practice, but we should not dismiss them out of hand. I think teleosemanticists in particular should appreciate this, since, to my knowledge at least, they have not attempted to defend Kimus and Snorfs (Pietroski, 1992) from Millikan’s criticisms of unrealistic thought experiments, or shown how those creatures are informative representation, while Swampman is not.

For philosophy of cognitive science specifically, there are further implications. First, there is a whole new Swampman sighting for teleosemanticists to debunk — and this one, I think, is much clearer about Swampman and his role than the blurry and partial photos that previous expeditions have come back with. If teleosemanticists are right, they should be able to show what the Swampman illustration gets wrong about cognitive scientific practice. To do this, it is not enough to build a notion of representation that *could* serve scientific practice (Shea, 2018); it also means showing that this notion *is* *used* in scientific practice, and that means confronting cases like Swampman, where the relevant scientific practices seem to work just as well in the absence of everything that the teleosemanticist builds into the notion of representation.

Second, and finally, the Swampman argument seems to show that plenty of things in biology make quite a bit of sense *without* the light of evolution (to mangle the famous quote from Dobzhansky, 1973). Arguments for teleosemantics are often prefaced with remarks about how essential it is to take evolution seriously (e.g., Garson & Papineau, 2019, Chapter 12), and it is worth asking what happens to the role of evolution in cognitive science when it isn’t defining scientific kinds. Evolution doesn’t become irrelevant, of course. But the murky causal origins of the brain are only veryweakly informative about its current structure (de Sousa et al., 2023). Any further role for evolutionary considerations has to be demonstrated by an investigation of cognitive scientific explanation (whether by illustration or otherwise), and the way its explananda relate to evolutionary theory (e.g., Cisek & Hayden, 2022). At the very least, that kind of investigation promises to be more revealing than simply defining cognitive science’s kinds in evolutionary terms.

**References**

Antony, L. (1996). Equal Rights for Swamp-persons. *Mind & Language*, *11*(1), 70–75.

Barsalou, L. W. (2016). On Staying Grounded and Avoiding Quixotic Dead Ends. *Psychonomic Bulletin & Review*, *23*, 1122–1142. https://doi.org/10.3758/s13423-016-1028-3

Block, N. (1978). Troubles with Functionalism. In C. W. Savage (Ed.), *Perception and Cognition* (pp. 261–325). University of Minnesota Press.

Boorse, C. (1976). Wright on Functions. *The Philosophical Review*, *85*(1), 70–86. https://www.jstor.org/stable/2184255

Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.

Cisek, P., & Hayden, B. Y. (2022). Neuroscience needs evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1844). https://doi.org/10.1098/rstb.2020.0518

Cummins, R. (1975). Functional Analysis. *The Journal of Philosophy*, *72*, 741–765.

Damásio, H., Grabowski, T., Frank, R., Galaburda, A. M., & Damásio, A. R. (1994). The return of phineas gage: Clues about the brain from the skull of a famous patient. *Science*, *264*(5162), 1102–1105. https://doi.org/10.4324/9780203496190

Davidson, D. (1987). Knowing One’s Own Mind. *Proceedings and Addresses of the American Philosophical Association*, 441–458.

de Sousa, A. A., Beaudet, A., Calvey, T., Bardo, A., Benoit, J., Charvet, C. J., Dehay, C., Gómez-Robles, A., Gunz, P., Heuer, K., van den Heuvel, M. P., Hurst, S., Lauters, P., Reed, D., Salagnon, M., Sherwood, C. C., Ströckens, F., Tawane, M., Todorov, O. S., … Wei, Y. (2023). From fossils to mind. *Communications Biology*, *6*(1), 1–21. https://doi.org/10.1038/s42003-023-04803-4

Dennett, D. C. (1988). Precis of The Intentional Stance. *Behavioral and Brain Sciences*, *11*, 495–546.

Dennett, D. C. (2007). Cow‐sharks, Magnets, and Swampman. *Mind & Language*.

Descartes, R. (1984). Meditations on First Philosophy. In J. Cottingham, R. Stoothoff, & D. Murdoch (Eds.), *The Philosophical Writings of Descartes, Volume 2* (pp. 1–62). Cambridge University Press.

Dobzhansky, T. (1973). Nothing in Biology Makes Sense except in the Light of Evolution. *The American Biology Teacher*, *35*(3), 125–129.

Egan, F. (2014). How to think about mental content. *Philosophical Studies*, *170*, 115–135.

Garson, J. (2019). *What Biological Functions are and Why They Matter*. Cambridge University Press.

Garson, J., & Papineau, D. (2019). Teleosemantics, selection and novel contents. *Biology and Philosophy*, *34*(3), 1–20. https://doi.org/10.1007/s10539-019-9689-8

Gendler, T. S. (2000). *Thought Experiment: On the Powers and Limits of Imaginary Cases*. Garland Publishing, Inc.

Gibson, J. J. (1972). A theory of direct visual perception. In J. R. Royce & W. W. Rozeboom (Eds.), *The Psychology of Knowing* (pp. 215–240). Gordon & Breach.

Goff, P., Seager, W., & Allen-Hermanson, S. (2022). Panpsychism. In *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/sum2022/entries/panpsychism

Griffiths, P. E., & Stotz, K. (2008). Experimental philosophy of science. *Philosophy Compass*, *3*(3), 507–521.

Hartle, B., & Wilcox, L. M. (2016). Depth magnitude from stereopsis: Assessment techniques and the role of experience. *Vision Research*, *125*, 64–75. https://doi.org/10.1016/j.visres.2016.05.006

Kim, D. (2021). Explanation and modality: on why the Swampman is still worrisome to teleosemanticists. *Synthese*, *199*(1–2), 2817–2839. https://doi.org/10.1007/s11229-020-02913-8

Kucharski, A. (2016). Post-truth: Study epidemiology of fake news. *Nature*, *540*(7634), 525. https://doi.org/10.1038/540525a

Lettvin, J. V., Maturana, H. R., McCulloch, W. S., & Pitts, W. H. (1959). What the Frog’s Eye Tells the Frog’s Brain. *Proceedings of IRE*, 1940–1951.

Lillian, P. E., Meyes, R., & Meisen, T. (2018). Ablation of a Robot’s Brain: Neural Networks Under a Knife. *31st Conference on Neural Information Processing Systems*. http://arxiv.org/abs/1812.05687

Liu, F., Dai, S., Feng, D., Peng, X., Qin, Z., Kearns, A. C., Huang, W., Chen, Y., Ergün, S., Wang, H., Rappaport, J., Bryda, E. C., Chandrasekhar, A., Aktas, B., Hu, H., Chang, S. L., Gao, B., & Qin, X. (2019). Versatile cell ablation tools and their applications to study loss of cell functions. *Cellular and Molecular Life Sciences*, *76*(23), 4725–4743. https://doi.org/10.1007/s00018-019-03243-w

Mahon, B. Z., & Hickok, G. (2016). Arguments about the nature of concepts : Symbols , embodiment , and beyond. *Psychonomic Bulletin & Review*, *23*, 941–958. https://doi.org/10.3758/s13423-016-1045-2

Mekik, C. S., & Galang, C. M. (2022). Cognitive Science in a Nutshell. *Cognitive Science*, *46*(8). https://doi.org/10.1111/cogs.13179

Mendola, J. (2009). Against Etiological Semantics. In *Anti-Externalism* (pp. 103–142). https://doi.org/10.1093/acprof:oso/9780199534999.003.0005

Milkowski, M. (2013). *Explaining the Computational Mind*. MIT Press.

Millikan, R. G. (1984). *Language, Thought, and Other Biological Categories*. MIT Press.

Millikan, R. G. (1996). On Swampkinds. *Mind & Language*, *11*(1), 103–117.

Millikan, R. G. (2010). On Knowing the Meaning ; With a Coda on Swampman. *Mind*, *119*(473), 43–81. https://doi.org/10.1093/mind/fzpl57

Nanay, B. (2014). Teleosemantics without etiology. *Philosophy of Science*, *81*(5), 798–810. https://doi.org/10.1086/677684

Neander, K. (1991). The teleological notion of “function.” *Australasian Journal of Philosophy*, *69*(4), 454–468.

Neander, K. (1996). Swampman Meets Swampcow. *Mind and Language*, *11*(1), 118–129. https://doi.org/10.1111/j.1468-0017.1996.tb00036.x

Neander, K. (2017). *A Mark of the Mental*. MIT Press.

Nersessian, N. J. (1992). In the Theoretician’s Laboratory: Thought Experimenting as Mental Modeling Author. *Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 291–301.

Panchal, R., & Jack, A. (2022). The contagiousness of memes: containing the spread of COVID-19 conspiracy theories in a forensic psychiatric hospital. *BJPsych Bulletin*, *46*(1), 36–42. https://doi.org/10.1192/bjb.2020.120

Papineau, D. (2001). The status of teleosemantics, or how to stop worrying about swampman. *Australasian Journal of Philosophy*, *79*(2), 279–289. https://doi.org/10.1080/713659227

Paul, L. A. (2014). *Transformative Experience*. Oxford University Press.

Peacocke, C. (2014). Perception, biology, action, and knowledge. *Philosophy and Phenomenological Research*, *88*(2), 477–484. https://doi.org/10.1111/phpr.12092

Peters, U. (2014). Teleosemantics, Swampman, and Strong Representationalism. *Grazer Philosophische Studien*, *90*, 273–288.

Piccinini, G. (2015). *Physical Computation: A Mechanistic Account*. Oxford University Press.

Pietroski, P. M. (1992). Intentionality and teleological error. *Pacific Philosophical Quarterly*, *73*(3), 267–282.

Porter, B. (2020). Teleosemantics and tetrachromacy. *Biology and Philosophy*, *35*(1), 1–22. https://doi.org/10.1007/s10539-019-9732-9

Putnam, H. (1992). Brains in a Vat. In K. DeRose & T. A. Warfield (Eds.), *Skepticism: a Contemporary Reader* (pp. 187–204). Oxford University Press.

Ramachandran, V. S., & Anstis, S. M. (1986). The perception of apparent motion. *Scientific American*, *254*(6), 102–109. https://doi.org/10.1038/scientificamerican0686-102

Richmond, A. (2022). How Computation Explains. *ArXiv*. http://philsci-archive.pitt.edu/id/eprint/21184

Richmond, A. (2023). What is a Theory of Neural Representation For? *Arxiv*. http://philsci-archive.pitt.edu/id/eprint/22401

Salvalaggio, A., de Filippo De Grazia, M., Zorzi, M., de Schotten, M. T., & Corbetta, M. (2020). Post-stroke deficit prediction from lesion and indirect structural and functional disconnection. *Brain*, *143*(7), 2173–2188. https://doi.org/10.1093/brain/awaa156

Schulte, P. (2020). Why mental content is not like water: reconsidering the reductive claims of teleosemantics. *Synthese*, *197*(5), 2271–2290. https://doi.org/10.1007/s11229-018-1808-6

Searle, J. R. (1980). Minds , brains , and programs. *The Behavioral and Brain Sciences*, *3*, 417–457.

Sebastián, M. Á. (2017). Functions and mental representation: the theoretical role of representations and its real nature. *Phenomenology and the Cognitive Sciences*, *16*(2), 317–336. https://doi.org/10.1007/s11097-015-9452-9

Shagrir, O. (2022). *The Nature of Physical Computation*. Oxford University Press.

Shea, N. (2018). *Representation in Cognitive Science*. Oxford University Press.

Shepard, R. N. (1984). Ecological Constraints on Internal Representation: Resonant Kinematics of Perceiving, Imagining, Thinking, and Dreaming. *Psychological Review*, *91*(4), 417–447. https://doi.org/10.1093/mind/xxv.3.415-b

Stich, S. (1978). Autonomous Psychology and the Belief-Desire Thesis. *The Philosophy and Psychology of Cognition*, 573–591.

Tolly, J. (2021). Swampman: a dilemma for proper functionalism. *Synthese*, *198*(7), 1725–1750. https://doi.org/10.1007/s11229-018-1684-0

Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: From consciousness to its physical substrate. *Nature Reviews Neuroscience*, *17*(7), 450–461. https://doi.org/10.1038/nrn.2016.44

von Eckardt Klein, B. (1977). Inferring Functional Localization from Neurological Evidence. In E. Walker (Ed.), *Explorations in the biology of language* (pp. 27–66). MIT Press.

Wachowitz, S., & Ewert, J. P. (1996). A key by which the toad’s visual system gets access to the domain of prey. *Physiology and Behavior*, *60*(3), 877–887. https://doi.org/10.1016/S0031-9384(96)00070-4

Walsh, D. M., & Ariew, A. (1996). A Taxonomy of Functions. *Canadian Journal of Philosophy*, *26*, 493–514.

Weiskrantz, L., Warrington, E. K., Sanders, M. D., & Marshall, J. (1974). Visual Capacity in the Hemianopic Field Following a Restricted Occipital Ablation. *Brain*, *97*(1), 709–728. https://doi.org/10.1093/brain/97.1.709

Weissman-Fogel, I., & Granovsky, Y. (2019). The virtual lesion approach to transcranial magnetic stimulation: Studying the brain-behavioral relationships in experimental pain. *Pain Reports*, *4*(4), 1–12. https://doi.org/10.1097/PR9.0000000000000760

Wright, L. (1973). Functions. *Philosophical Review*, *82*, 139–168.

Zhang, C., Kaye, J. A., Cai, Z., Wang, Y., Prescott, S. L., & Liberles, S. D. (2021). Area Postrema Cell Types that Mediate Nausea-Associated Behaviors. *Neuron*, *109*(3), 461-472.e5. https://doi.org/10.1016/j.neuron.2020.11.010

1. (He’s joking.) [↑](#footnote-ref-1)
2. I’m ignoring versions of teleosemantics that appeal to non-selectional functions, like forward-looking (Nanay, 2014) or causal-role functions (Cummins, 1975), to define representation. They are not Swampman’s target, or mine. [↑](#footnote-ref-2)
3. Swampman is usually attributed to Davidson (1987), but there is an earlier version of him in Millikan (1984, p. 93), and precursors in Stich (1978) and Boorse (1976). [↑](#footnote-ref-3)
4. Not to mention that systems can be explained in terms of kinds they don’t instantiate. We can explain misinformation as a virus (Kucharski, 2016; Panchal & Jack, 2022), but (thankfully) philosophers aren’t demanding that biologists redefine *virus* to include misinformation. [↑](#footnote-ref-4)
5. N.B., the idea isn’t that we would explain the situation using the notion of observation. It’s that the explanations that invoke observation would be *just as successful*, and *in all the same ways*. [↑](#footnote-ref-5)
6. *Use*, here, is not an intentional notion. Agents can use resources intentionally, but when they’re using a part of their hippocampus to perform the task of navigation, we’re dealing with a purely functional notion of *use*. [↑](#footnote-ref-6)
7. This work is not necessarily tied to strong ‘localization’ assumptions, to the effect that *all a brain area does* is to support the task in question. [↑](#footnote-ref-7)
8. Many, ablation studies do find a change, and conclude that the ablated area *was* contributing to the task. But I’m focused on the subset of ablation studies that find no change in task performance because the *illustrations* I discuss are ones where removing a feature has no effect on explanations. [↑](#footnote-ref-8)
9. They also, generally, make some conclusions about *how* it does so, but nothing is lost for my purposes if we stick to the less sophisticated inferences. [↑](#footnote-ref-9)
10. Of course the ablated organism has *some* kinds in common with the intact one, just as Swampman does with human beings. The point is that the arguments don’t rely on the two sharing the kind we’re investigating: *representational system*, or *intact member of a certain species*. [↑](#footnote-ref-10)
11. This is not the only kind of reasoning in the study, or course. Much of the work couldn’t be described as paring away features of a known stimulus, but as trying out lots of simple stimuli, like edges (pp. 1946-7). And the majority of the paper isn’t even about prey capture. But the parts that *are* do involve the kind of reasoning I’ve described [↑](#footnote-ref-11)
12. To be fair, some teleosemanticists might say cognitive scientists would have nothing to explain about Swampman: they want to explain *successful* behavior, and without a selection history Swampman can have no ends to be successful with respect to (cf. Shea, 2018, p. 22). But this isn’t a plausible characterization of cognitive science, which aims to explain patterns of behavior, like the fact that you forage in your environment in a certain way, or that you are able to make it from one place to another more often than chance would have it if you are given certain cues. These explananda do not disappear if we stop characterizing them as successes-in-a-selectional-sense. [↑](#footnote-ref-12)
13. Thanks to Charles Bakker for a discussion about these alternative, explicitly historical explananda. [↑](#footnote-ref-13)
14. There is also no reason to doubt that if we go back to whatever slice of time in which Swampman doesn’t yet have functions, a similar illustration-based argument would apply. [↑](#footnote-ref-14)
15. Maybe the problem is just that Swampman is imaginary: he’s a *thought* experiment rather than an *actual* experiment. This would be a problem if it meant he only probed our intuitions, but the reasoning that an illustration asks you to undertake is more like prediction than intuition. *What would happen if Swampman walked into the lab* isn’t mere intuition-mongering any more than *What would Mom do if she found out I got in a fight at school* or *How would this organism behave in its Normal environment (Millikan, 1984)?* As long as we can be reasonably confident in our predictions, these are entirely unproblematic questions, even if Mom *doesn’t* find out I got in a fight, and the situation remains imaginary. [↑](#footnote-ref-15)
16. Thanks to Devin Morse for the example. (The name is my own fault.) [↑](#footnote-ref-16)
17. There are also implications for the philosophical literature on thought experiments, which I will only note briefly here. That literature often understands thought experiments as the removal of a target system’s features, but the consequences drawn generally concern *kind-hood* (is it still an *X* without that feature?) (Gendler, 2000), or the imagined system’s role as a *model* (Nersessian, 1992). Neither provide quite the role for thought experiments the illustration account does. [↑](#footnote-ref-17)