**What really lives in the swamp? A new monster for etiologists**

Andrew Richmond

**Abstract:**Well, ok,it’s an old monster, but I propose to rebuild and reanimate it. Swampman — a molecule-for-molecule copy of Donald Davidson that arises randomly after lightning strikes a swamp — is a popular counterexample to etiological teleosemantics, and to the use of etiological kinds in cognitive science more generally. He appears to have functional, representational, and computational properties but no (relevant) etiology; it is thus a mistake to define function, representation, and computation in etiological terms. Teleosemanticists have responded by rejecting the absurd and unscientific nature of the counterexample. Millikan, Papineau, Neander, and others argue that Swampman is not in the domain of the theory of functions (representations, computations) in the first place. He is not a member of the *real kinds* (Millikan’s term) that we are investigating when we invoke functional (etc.) notions, despite superficially looking like one. By adopting a scientifically-motivated approach to natural kinds, these philosophers avoid accepting Swampman as an example of a functional (etc.) system. So, a fortiori, they do not accept him as a counterexample to any theory of functions (etc.). I concede this as a response to the Swampman Counterexample, but I argue that it was a mistake to treat Swampman as a counterexample in the first place. Rather than an example of the kinds of system cognitive science treats, Swampman should be understood as a means to *illustrate* certain cognitive scientific explanations and the resources they rely on. In this role he does not flinch from the etiologists demand for a serious treatment of science. He is, in fact, a way of carefully probing the explanatory resources involved in science. I argue that functional (etc.) explanations would apply just as successfully, and for the same reasons, to Swampman as to his human counterpart (by any reasonable metric: prediction, explanation, systematization, …). That means that functional (etc.) explanations could work equally well, and for the same reasons, if their target systems had no etiological properties. And to define functional (etc.) kinds, insofar as they figure into cognitive scientific explanations, in etiological terms would immediately contradict this fact. So functional (etc.) kinds should not be defined in etiological terms.

*The use of representations is an engineering principle, like the use of levers or gears.*

– “Neuroscience and teleosemantics”, Ruth Millikan

*You live between three trees. One is behind you. The Lote — the tree of life for your Persian ancestors. The tree at the boundary of the seventh heaven, that none may pass. Ah, but engineers have no use for the past, do they?*

– Ma Shouying in *The Overstory*, Richard Powers

**1 The Swampman Counterexample**

Biology, we are told, is *shot through* with functions, *saturated* with them, *swimming* in them, *slathered* *head-to-toe* in functions and functional language. The same holds for cognitive science, the focus of this paper. It is common to hear, e.g., that circuits in the visual system have the *function* to represent edges in the environment, *so that* they can derive a representation of shapes and objects. Parts of the retina are *supposed to*respond to colors in the environment. Parts of the hippocampus are *meant to* respond to the animal's location in its environment, the *job*of other parts is tokeep track of the animal's movement or visual flow, and yet other parts are *there to* register the direction the animal is facing. But the nature of functions is not investigated in biology or cognitive science themselves. So *function* is an important but unanalyzed scientific concept. And it is a potentially fruitful one: functions are often suggested as the grounds for representational (Millikan, 1984; Neander, 2017; Shea, 2018) and computational properties (Milkowski, 2013; Piccinini, 2015). According to teleosemanticists, “[t]hings are called ‘representations’ if they have a certain kind of function or telos and perform it in a certain kind of way” (Millikan, 2020, p. 1). The teleological view of computation (gaining support only recently) tends to rely on a similar notion of function, especially in Piccinini’s version (2015).

 So, to support and understand the use of functional, representational, and computational notions, philosophers have given various accounts of functions. The main two are the *etiological* account and the *causal role* account. The much-maligned *causal role* account (Cummins, 1975) says that something has the function to X if and only if it’s Xing figures into a *causal decomposition* of its host system’s perform capacities. The etiological account (Millikan, 1984; Neander, 1991) — by far the majority view — says that something has the function to X if and only if it exists because previous members of its kind were *selected for*, and those members were selected for because they Xed. Selection is understood to include evolutionary and intentional selection, among other processes.

The main problem for the causal role account has been that it counts too many things as functions. The main problems for the etiological account have been counterexamples, the most prominent of which is *Swampman*, a molecule-for-molecule copy of Donald Davidson who crawls out of a swamp, having been created by a chemical reaction when lightning struck the swamp. There are more straightforward counterexamples. Features of an organism when they appear for the first time (or for some other reason haven’t had a chance to be selected for) can seem to instantiate functions, representations, and computations (Porter, 2020; Walsh & Ariew, 1996). But Swampman brings out the problem most starkly. Swampman appears to have exactly the same functional organization that Davidson has, and he *really* seems to have the same sort of representational and computational properties. After all, he’s walking around and navigating, speaking English and whatever other languages Davidson spoke, processing visual scenes, etc. We can ask him to follow a recipe or to imagine himself as an artist. We can ask him to deliberate over a career change, deciding whether to leave his job haunting swamps to become a painter. Isn’t he going to do this the same way we do, and doesn’t that mean he’s doing it by virtue of his functional, representational, and computational properties? If that’s the case, Swampman’s lack of a selection history makes him a counterexample to the etiological account of those properties.

Etiologists respond that Swampman is not a counterexample because he is not an *example* in the first place: he’s not in the domain of the etiological accounts of function, representation, or computation, so he’s not a system with functions (representations, computations) in the first place — he’s just not in a position to be a counterexample. This is generally cashed out in the following way. When we investigate functional (etc.) systems we are investigating a *real kind* (Millikan’s term, 1996, 2010), or an empirical kind. Those kinds are made up of things that more or less share a cluster of properties, and share them for good empirical reasons.[[1]](#footnote-1) And Swampman doesn’t satisfy the second condition. Imagine I presented Swampman to you as a counterexample to the historical definition of *species* (Millikan, 2010; Neander, 1996). That definition taxonomizes species partly cladistically: not just by the common properties that conspecifics share, but by the ancestry relations that explain why they share those properties. Taking those ancestry relations into account makes certain generalizations possible and enables a deep understanding of species change, at the minor cost of excluding some monsters in the philosophical bestiary from their intuitive species. Pointing out that Swampman looks like a human being is no objection to the definition of *human being* in historical terms because it’s still plausible that he’s not a human being in the first place. He doesn’t share the properties that bind human beings together into the real kind that is relevant for the purposes of scientific explanation. All the thought experiment shows is that freak accidents could (in a very permissive sense of “could”) create things that seem like human beings but aren’t. Or, more pessimistically, it shows that the human imagination can cook up things that seem like human beings but aren’t, when that imagination is motivated to take advantage of our naïve, scientifically uninformed intuitions about human beings (like our usually-useful heuristic that if something walks like a human and quacks like a human, it’s probably a human).

As with species, so with function, representation, and computation, according to the etiologist about those notions. Swampman plausibly isn’t a member of the empirical kinds *functional system*, *representational system*, or *computational system* because even if he seems similar to members of those kinds, he doesn’t fit the broader profile required for membership. Swampman’s ‘functions’ (etc.) look like real functions (etc.), but only for the same reason that Swampman looks like a real human being. On a deeper investigation, one sensitive to the relevant scientific context and explanations, etiological kinds provide the right basis for categorizations that serve functional (etc.) explanation. The broader lesson is that here, as generally in the investigation of empirical kinds, we have “a case where the a priori method of example and counter-example fails” (Millikan, 2010, p. 79; see also Papineau, 2001), because there is no reason to believe our a priori intuitions about kind membership successfully track empirical kinds. This does not *rule it out* that Swampman is a member of the relevant kinds — that’s the job for a positive theory of function (etc.). The response to the Swampman Counterexample just shows that if Swampman really is home to functions (etc.), that has to be established by argument, not by surface intuition. Teleosemanticists have been waiting for such an argument, so far to no avail.

That’s about the consensus view on Swampman, especially in recent work on teleosemantics, which tends to discuss Swampman briefly if at all, only noting the standard response (e.g. Neander, 2017). And dissenters from the etiological consensus on functions (particularly dissenters from teleosemantics) tend not to deploy the Swampman Counterexample much anymore. They’re right not to. It’s not a good argument, for exactly the reasons the response above identifies. But, to come to the point of this paper, I think we can reinvent Swampman. We have to collect up his remains, replace some limbs, rearrange others, and stich everything back together, but we can reanimate him. The resulting argument — the *Swampenstein Argument*, of course — is not a counterexample, but a direct challenge to the explanatory relevance of etiological properties in the cognitive scientific explanations that functional (etc.) notions figure into. It uses Swampman only to illustrate the way those explanations work and the resources they rely on — not as an example of the systems that a theory of functions (etc.) targets. I take the Swampenstein Argument to be a good representation of what philosophers are *trying* to say when they talk about Swampman, even if they fall into the mistake of using Swampman as a counterexample.

Recall that our philosophical project is motivated, in the first place, by the use of functional (etc.) notions in cognitive scientific explanations. That’s why we’re investigating the notion of a *function*, and not the notion of an *Andy Kaufmanesque talk show appearance*. What we’re after is an account of functions (etc.) that shows us how they support the cognitive scientific explanations they figure into (Neander, 2004, pp. 30–31). This is how teleosemanticists especially motivate their project, but the same holds in the less widespread discussions of etiological notions of computation and function itself in cognitive science. Shea is a good illustration of this trend. In the space of a couple pages he makes these claims to prepare the way for his teleosemantic view:

A theory of content is answerable not to intuition, but to the role representations play in explaining behavior. (2018, p. 28)

When a scientific explanation points to representational content to explain behaviour, we need to get inside that explanation to see how it works. That means getting into the details of a behavioral profile and its causal underpinnings. Only with the details in view can we properly address the question: what kind of thing is representational content, to enable behavior to be explained in that way? (2018, pp. 28–29)

Rather than intuition, our theorizing is constrained by a desideratum, something we want to explain. What we want to explain is how adverting to representational content allows us to explain behavior. We want an account of content which shows why content plays that proprietary explanatory role. (2018, p. 29)

This point is especially important on the ‘real kinds’ response above, which makes it clear that the taxonomy excluding Swampman is motivated by its importance to cognitive scientific explanation. Etiologists about function, representation, and computation don’t do much to flesh out this proposal — they don’t give much detail about the explanatory goals that etiological notions serve. But the focus on how etiological kinds serve cognitive scientific explanations is important, and the Swampenstein Argument is an attempt to take this focus as seriously as teleosemanticists do, and to probe the question of whether functions (etc.) are etiological properties by asking whether functional (etc.) explanations really depend on etiological kinds. I’ll go ahead and make the argument in the next section, returning to potential controversies in the following one.

**2 The Swampenstein Argument**

Let Swampman find his way into a psychology or neuroscience lab and sign up to participate in a study. He is a molecule-for-molecule copy of Donald Davidson, so he will display all of Davidson’s cognitive abilities insofar as those are described without reference to history, to environmental facts that differ between the two, or to any properties that rely on one’s being the *real* Davidson. So, e.g., the capacity *to* *remember who one’s mother is* will have to be set aside, since Davidson will have such an ability and Swampman will not (or will systematically err in it). But this is not the kind of ability cognitive scientists concern themselves with. The phenomena cognitive scientists are interested in are things like the ability to store information and retrieve it, to navigate one’s environment, to speak, to discriminate between objects or colors — or, for that matter, the tendency to display biases or to exhibit certain patterns of irrationality. Louise Antony makes a similar point about biology insofar as it’s motivated by medical purposes: a doctor would be just as interested in treating a cancer patient from the swamp as one from Dallas (Antony, 1996, p. 72). The introduction of etiology does not happen at the level of explananda; if etiology is involved in cognitive science, it is as explanans or explanatory resource.[[2]](#footnote-2)

The first major step of the argument is a simple disjunction: either Swampman’s display of phenomena and capacities of interest to cognitive science is explicable, or it is inexplicable. It’s not his history that’s in question — we’re not asking how Swampman *came by* his capacities. It was never at issue that an explanation of *that* would appeal to history (if not necessarily historically-defined kinds). Our topic is the kind of thing that cognitive scientists, rather than evolutionary biologists or physicists or ecologists, would want to explain. We either can or can’t show how Swampman’s brain supports his cognitive capacities. We either can or can’t reverse-engineer Swampman’s cognitive capacities, as the project of cognitive science is often characterized (Dennett, 1998).

So, which way do we go? What would it be, exactly, to deny that Swampman is explicable? Would we sit Swampenstein down in the laboratory and simply be stymied by how his brain supports his capacities? Would we not be able to model the causal structure of his brain at the levels of grain that allow us to build models of his behavior? Would this project at least be less successful than it is with Davidson? Swampman is a physical system just like Davidson or anything else; he has no philosophical tricks up his sleeve. I don’t see any way of denying that his various capacities and abilities would be as explicable as the capacities of any system.

The next step is to ask how we would explain those 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. We would observe Swampman the same way we observe organisms of whom we don’t know their evolutionary history. 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, neural network models and optogenetic imaging, and so on. We would apply precisely the same probes and the same explanatory resources that we do in Davidson’s case. And we would take the same approach to building process models (Samuels, 2019) — including functional, representational, and computational approaches. We would apply these to both subjects, seemingly with the same degree of success, *even if we didn’t know which was Davidson and which was Swampman*. The models we would construct would equally predictive of each’s behavior, as revealing of its neural basis, as useful in medical interventions, and so on. Nothing about the creature or its brain would impede our understanding or interfere with those projects, any more than we are impeded by our lack of knowledge of organisms’ evolutionary history elsewhere in cognitive science.[[3]](#footnote-3)

It’s possible that we would not meet other explanatory goals 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? I’ll discuss a few possibilities in the next section, and I do think there may be room to push back here, but (as Papineau, 2001 notes) it is entirely the burden of the teleosemanticist to describe the ways that etiological kinds are required by scientific explanation — I can’t rule out every possibility here. For now I am happy to redirect attention to this debate and away from the use of Swampman as a counterexample, and I’m especially happy if it appears somewhat mysterious what goal our functional (etc.) explanations of Swampman might not meet.[[4]](#footnote-4)

To conclude the argument: I’ve argued that for cognitive scientific purposes, Swampman is just as explicable as Davidson is, and by the same forms of explanation. But the explanations in Swampman’s case cannot rely on etiological kinds. He does not instantiate those kinds, and that would make explanations relying on them less successful in his case than Davidson’s. So, since we have every reason to believe that these forms of explanation would be just as successful in Swampman’s case as in Davidson’s, the explanations do not rely on etiological kinds. Since the etiological theory of functions and the theories of representation and computation based on it claim that those explanations use etiological notions of function, representation, and computation, these theories are wrong, at least about the notions of function, representation, and computation that cognitive science uses.[[5]](#footnote-5)

Note that this is not just an objection to etiology in the definition of representation, computation, or function, but to etiology in cognitive science at all. And it does not succumb to the “real kinds” objection because nothing in the argument depends on Swampman belonging to the kind *functional (etc.) system*, or any other kind for that matter. It pivots (as the “real kinds” objection suggests we do) to focus on the nature of the explanations cognitive science gives, and it uses Swampman only as an illustration of those explanations and of the resources they require, an illustration in particular of the fact that they do not require their target systems to belong to any etiological kinds. If the argument is right, cognitive scientific explanations are non-etiological, and insofar as those explanations include functional, representational, and computational notions, those notions are non-etiological too.

**3 Discussion**

I’ll now discuss some points of controversy in the argument above, by addressing potential objections. First, though, a caveat. The argument excludes etiologically-defined functions from cognitive science, but not other kinds of function, e.g., forward-looking functions (Walsh & Ariew, 1996), functions as the causal role account defines them, or something else.[[6]](#footnote-6) That shouldn’t incline us towards any particular view of functions — the thought experiment barely begins the investigation of the explanatory resources that figure into cognitive science. But it is notable that such a bare beginning already seems to exclude etiological functions.

 *3.1 What is Swampman supposed to be?*

A natural objection to the Swampenstein Argument would claim that it still uses Swampman as a counterexample — are we sure I’m not surreptitiously treating Swampman as an example of a functional (etc.) system, perhaps by assuming functional (etc.) explanations of him would succeed? And even if I’m not, is the case *just too ridiculous* to tell us anything about actual cognitive scientific practice, as Dennett seems to argue (2007)? First, note the difference between an example and an illustration. An example is a member of the kind it exemplifies. I haven’t depended on any assumption that Swampman belongs or doesn’t belong to any kind, except for the stipulation he doesn’t belong to kinds like *evolved system*. He may belong kinds like *representational system* or not; the point is that whatever kinds he belongs to, our explanations of him don’t rely on his membership in *etiological* kinds.

To be sure, it’s possible that Swampman belongs to the kind *representational system*, depending on its definition. Perhaps a deflationist like Egan (Egan, 2010, 2014, 2019) would be sympathetic to an argument saying he belongs to that kind simply because representational explanations are applicable to him. In that case the argument against teleosemantics is precisely the Swampman Counterexample. But teleosemanticists are generally operating with a more restrictive notion of a kind than this, and I have not assumed anything like the deflationary view. I’ve left it open that Swampman doesn’t belong to the kind *representational system*, in which case the Swampenstein Argument shows (in addition to the irrelevance of etiological properties) that a target system’s belonging to the kind *representational system* would be irrelevant to the success of representational explanations. This just reflects the variety of definitions we might have of representation (and function, and computation). For *some* definitions of representation (etc.) Swampman certainly represents; the point is that he cannot represent in the etiological sense, and that this means representing in the etiological sense is not a property relied on by representational explanation. The argument makes no commitments to the definition of representation (etc.), except that excludes the simplest route to a counterexample by taking a non-deflationary point of view on representation.[[7]](#footnote-7) The argument, rather than casting him as an example of a representational (etc.) system, uses him to illustrate the workings of representational (etc.) explanations.[[8]](#footnote-8)

The illustration can be understood in two ways. First, it could be the hypothetical *inflation* of an existing type of case: an organism with an original, or evolutionarily new, trait. The inflation gives Swampman all of his traits originally, rather than just some. Second, it could be the hypothetical *removal* of some irrelevant properties of a system (etiological ones) to show that they are, in fact, irrelevant. On either understanding, we’re illustrating the way cognitive scientific explanation actually works by considering how it could be applied to systems that are just inflated or pared down versions of the systems it really applies to. Consider a common misunderstanding of physics expressed by people who say things like, “Explanations in quantum mechanics appeal to observers, so quantum mechanical effects rely on conscious observation”. If I raised this worry, you would probably respond by showing me how those explanations actually work without a conscious observer — just some non-conscious measuring device. But this misunderstanding (like many involving quantum mechanics) is often intractable, and if I persisted in my misunderstanding — “But even where it’s a non-conscious measuring device, isn’t the thing you’re measuring in some kind of superposition until we check the measurement?” *—* you could respond with a more extreme hypothetical, showing me how quantum mechanical explanations still make sense in a world without conscious observers at all. That is, you would exclude an irrelevant feature (conscious observation) to show me that it is, in fact, irrelevant. It does not matter that consciousness actuallyexists; what matters is that if it hadn’t, the relevant explanations would be unchanged. This is an illustration of scientific reasoning, not an example of it. It doesn’t matter that the world without human observers is non-actual, and it wouldn’t even matter if that world was highly unlikely given our understanding of physics. None of this matters because the world so described is not intended as an example of situations in which quantum mechanical explanations work, but as a way of illustrating how they work. And, recall, this attention to the detailed workings of scientific explanation is exactly what teleosemanticists were after in the first place. It motivates their project and furnishes the ‘real kinds’ response.

But what about Dennett’s worry: is Swampman still just a bit too ridiculous? Well, he is ridiculous, especially given the particulars of his case and the intuition pumps surrounding him — lightning in Swamps and whatnot. But this doesn’t pose a problem given my purposes. As I said, Swampman is merely an inflation of an existing phenomenon (the existence of organisms with original traits), and as long as we’re relying on him only to illustrate the resources cognitive scientific explanations depend on, deviating from actual cases is to be expected. In extinguishing all consciousness, we move far from the real world in which quantum mechanical explanations are given, and if fundamental physics makes conscious life highly likely we might even move far from the physically plausible worlds, far from the ones that contain real kinds of scientific interest to us. But we do so only to show that the eliminated aspects of reality were not explanatorily relevant. The ridiculous hypothetical serves a perfectly serious point.[[9]](#footnote-9)

One last point on from the nature of Swampman himself. There can be a feeling of underhandedness when you bring up Swampman: it might seem that the explanations of Swampman only work because he’s *tailored* to be identical to Davidson and therefore to be explicable in the same ways (Neander, 1996). In the argument I do move explanations straight from Davidson to Swampman. But we could construct a similar hypothetical with a new ‘human’, different from any existing human. Or we could construct it with an animal that differed from *every* species. Still, if the animal triggers the interest of cognitive scientists — if they find interesting things to explain in its patterns of behavior — they will apply the same methods and explanatory strategies they apply elsewhere, and there is every reason to believe they’ll have the same success. If they won’t, we again need to be told what cognitive scientific goals the explanations would fail to meet and why. This type of story is what the etiologists owe us, and what I’ve suggested doesn’t exist.

*3.2 Farquaad’s revenge*

Some, including Shea (2018), get around Swampman-style objections by gerrymandering the notion of a function to include Swampman’s early interactions with his swamp. Shea has a “disjunctive” account of functions, arguing that not only evolutionary selection but also learning, and even the differential persistence of an organism, can engender functions (2018, pp. 56–63). So while Swampman has no functions at the moment he is created, he gains them after he crawls out of the swamp and starts getting around in it. Likewise for representations and possibly computations. This is a great response to the Swampman Counterexample: even if we can’t exclude Swampman from the domain of our theory (as the standard response does), we can encompass him without contradiction: our view *confirms* the intuition that the objection wanted to exploit — Swampman has functions (etc.) after all.

 But what about the Swampenstein Argument? Shea’s response is consistent with etiological kinds being an explanatory resource in cognitive science, given the rejiggered etiology that includes learning and so on. My argument depended on a hypothetical case where etiology was *not* present, so that that explanations successful in that case must not depend on etiological kinds. If “etiology” is defined as broadly as Shea defines it, it remains present in the hypothetical I’ve described, so I can’t reach the conclusion.

I’ll say two short things about this. First, it seems dubious to assume that Swampman doesn’t have any functions (etc.) until he starts learning or persisting. It would require a further argument to make this case, but I would not just assume that something like the Swampenstein Argument couldn’t also apply to Swampman in his very first moments. Second, and more importantly, as a response to the Swampenstein Argument this would have Shea admitting that the functions engendered by non-evolutionary etiology are *it* — natural selection is still excluded. That would be a substantial revision to the details of Shea’s view, and I think to its intentions. The result is a possible view, but it simply isn’t his view, or for that matter Millikan’s, Neander’s, Papineau’s, etc. There may be more to say about how far we can get with non-evolutionary etiology, but since the majority of etiological views, and especially the most influential ones, are diligently committed to an evolution-inclusive notion of etiology, the argument against that influential majority is not undercut.

 *3.3 How do I get off this airboat?*

One might worry about the other properties this style of argument excludes from explanations. Would it exclude etiological kinds too often? What about a *whole* *swamp* generated by a lightning strike? Would that hypothetical show that etiology is irrelevant to ecology?[[10]](#footnote-10) Which properties can be excluded with Swampenstein-style arguments — which properties can be *Swampensteined* — depends on the explanandum, along with context to do with the type of inquiry involved, the specific mode of explanation used, and so on. For example., imagine trying to explain why I’m wearing pants right now. You could probably construct an argument like the one above to show that this explanation doesn’t require reference to any history beyond this morning, let alone historically-defined kinds. But if you wanted to explain why I’m wearing pants *rather than* a skirt, and if you were giving explanations in the style of social science or gender studies, you’re going to have a hard time finding any useful mode of explanation that can be Swampensteined. Those modes of explanation tend to appeal directly and explicitly to historical patterns, and it is the reliance of current phenomena on historical trends that students of those fields are mostly out to explain in the first place. So explananda and explanans concern history in a way that would undercut a Swampenstein-style argument. (And it’s worth noting that reference to history may be present without the use of historically-defined kinds. The fact that I’m wearing pants rather than a skirt might be explained by reference to historical events without any assumption that the kind *pants* (or any other kind) is defined by its history.)

 So, the exclusion of etiology is not as radical as it seems — etiology remains explanatorily relevant in many cases, particularly where the explananda is historical, or where the mode of explanation, for any number of reasons, relies on historical kinds or facts. To address the original example: if we were interested in how ecosystems develop, a whole swamp generated by a lightning strikewould not compose a Swampenstein-style argument because it would erase the explanandum. If we were interested in how equilibriums are maintained once they’re created, that particular explanandum may be Swampensteinable, so long as is nothing a historical explanation can uniquely accomplish, and no kinds involved in the explanation that defy non-historical definition. The moral is just that etiology cannot — any more than gender norms or fashion trends — be pulled into an explanation without a plausible story about what it contributes and how. And, a fortiori, etiologically-defined kinds will have to be justified in the same way.

*3.4 From ‘can’ to ‘should’*

Maybe you accept that our current modes of explanation in cognitive science canwork without using etiological kinds. That doesn’t mean they *should* work without using etiological kinds. Consider a different type of case: just because we can explain a phenomenon in a certain way, that doesn’t mean we should — our explanation might not be the only possible one, and others explanations may better suit our purposes. E.g., among other phenomena, human aggression can be explained in various ways depending on our interests, pragmatic concerns, or cognitive needs (Potochnik, 2017). Perhaps a similar point can be made in the case at hand: given a certain type of explanation, there may be different ways of understanding it, rationalizing it, or providing it philosophical grounding. And even if we can get by in these respects without etiological kinds, those kinds have some other benefit: they better suit our interests, pragmatic concerns, cognitive needs, or something else.

If that’s so, it remains to make the case for this purported benefit. Reverse-engineering isn’t generally aided by etiological kinds, even if it is buttressed by an understanding of the target system’s history (for inspiration or exploratory purposes). So it’s not obvious that details about our interests/pragmatics/cognition will support the use of etiological kinds. Papineau begins a description of the benefits he thinks etiological kinds have like so:

Perhaps, if we had a *fully* detailed grasp of all the current causal powers of psychological states, then we would have no need of classification by historical origins. But I take our knowledge of such causal powers to be at best fragmentary, and so I think it can help greatly to know about the histories that shaped psychological states. (Papineau, 2001, p. 288)

The attention to our pragmatic needs in imperfect explanatory contexts is laudable, but unfortunately Papineau never carries out the elucidation he proposes here except to point out that the etiological properties of a component can be informative about its causal role in a system. And that much is conceded by the non-teleosemanticist: the teleosemanticist doesn’t need to show that etiology is sometimes informative in some contexts. The etiologically-inclined need to show that *in the context of* *cognitive science*, etiology is informative in that it *defines the kinds we use* in our explanations of the brain. I’ll consider three ways of making something like this case, appealing to benefits that etiological kinds may confer on explanations.[[11]](#footnote-11)

First, what about the benefit of *unification*? We want cognitive science to be unified, or at least compatible, with evolutionary biology. At least as a long-term goal the two disciplines should shed light on each other, and there seems to be a lot to learn from their collaboration. Does that mean cognitive science should adopt etiological kinds because biology does? No. Unification doesn’t mean *kind-sharing*. Compare the unification of biology with chemistry: we want that unification for similar reasons that we want the unification of cognitive science and biology; we especially want a story about how biological systems are composed of and historically formed from more basic structures. But we don’t want biologists to adopt the classificatory scheme of chemistry, nor vice versa. Unification means understanding how subject matters relate, what bridges can be drawn between them, and what composition or emergence or causal relationships hold between their parts, among other things. It does not mean sharing a single classificatory scheme. Taxonomic pluralism rather than taxonomic monogamy is the norm in science, and is justified even in closely related inquiries (Kitcher, 1985).

What about a deeper concern with the possibility of cognitive science in the first place? Perhaps we must understand the brain in terms of etiological kinds because those kinds give us the only *real kinds* for cognitive science to target, generalize over, and so on. This is suggested by Millikan when she says, “real kinds have properties in common … by empirical connection, and there is no empirical connection between the properties of humans qua human and Swampman. *There is no way to run a univocal empirical science over both*” (1996, p. 109). This makes a familiar response to the Swampman Counterexample, but as a response to the Swampenstein Argument it fails. Generalizations can be run over systems with similar causal structures even if they do not belong to a real kind in Millikan’s sense. And anyways, areas of inquiry are not typically held together by their investigating a common real kind in her sense. What real kind (in Millikan’s sense) does philosophy, geology, or computational neuroscience target? There is no plausible candidate in any of those cases. Sciences target systems that are interesting for any number of different reasons (and often *because* they are artificial and do not belong the same real kind as that science’s traditional objects of study — computer scientists studying the internet, neuroscientists studying artificial neural networks, sociologists studying cellular automata, etc.). They target whatever systems are of interest to them and are epistemically tractable by their tools and methodology, pitch-able to their funding sources, and so on. There is no reason to think that cognitive science, to be a genuine science (or to make fruitful generalizations or what-have-you) needs philosophers to define a real kind to subsume its target systems, in reference to which real kind its explanatory concepts need to be defined.

Lastly, how about another sort of benefit altogether: a philosophicalone. Some etiologically-enamored philosophers think that teleosemantics is the only naturalistic way of explaining *real* intentionality, as opposed to mere as-if or derived intentionality, because it’s the only naturalistic path to the distinction between *correct* and *incorrect*, *representation* and *misrepresentation*, *function* and *malfunction*, and so on.This marks a significant shift — we’re accepting that cognitive science can get by without historical kinds,[[12]](#footnote-12) and instead arguing for the *philosophical* gains teleosemantics provides.[[13]](#footnote-13) There may be something to say for this view of representation in philosophy, but there is no reason or justification for imposing it on cognitive science. Grant that philosophers, for philosophical purposes, need to add etiological components to the kinds used in cognitive science. To support teleosemanticists in cognitive science, this addition would have to have role in the understanding of cognitive scientific explanations themselves. And that’s precisely what the Swampenstein Argument was an argument against. If philosophers, for philosophical purposes, need there to be such a thing as *real* intentionality and need to define it etiologically, they can borrow scientific resources to define it. But that does not mean scientists need to be using that notion of intentionality themselves. (The points about taxonomic pluralism and unification apply here as well.)

**4 Conclusion**

Philosophers have become sympathetic to a methodological view that the philosophy of any area of science must be done with a sophisticated understanding of that area of science in mind: its context, goals, methods, and so on. The teleosemantic response to the Swampman Counterexample is a laudatory example of this careful treatment of science. But, in turn, that response suggests a more fruitful use of Swampman — one that engages yet more seriously with the scientific explanations that he is intended to illuminate. I’ve begun that project here; I haven’t made a conclusive argument that the illustration of cognitive scientific explanation that Swampman provides is one on which etiological properties (and especially etiologically-defined kinds) play no role. But I’ve set the grounds for a more fruitful debate over Swampman, and shown the significant challenge that Swampman poses for teleosemantics.

**Works Cited**

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

Cao, R. (2012). A teleosemantic approach to information in the brain. *Biology and Philosophy*, *27*, 49–71.

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

Dennett, D. C. (1998). *Brainchildren*. MIT Press.

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

Egan, F. (2010). Computational models: a modest role for content. *Studies in History and Philosophy of Science*, *41*, 253–259.

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

Egan, F. (2019). The nature and function of content in computational models. In M. Sprevak & M. Colombo (Eds.), *The Routledge Handbook of the Computational Mind* (pp. 247–258). Routledge.

Kitcher, P. (1985). Narrow Taxonomy and Wide Functionalism. *Philosophy of Science*, *52*(1), 78–97.

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

Millikan, R. G. (2020). Neuroscience and teleosemantics. *Synthese*. https://doi.org/10.1007/s11229-020-02893-9

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. (2004). Teleological Theories of Mental Content (from the Stanford Encyclopedia of Philosophy). In *Stanford Encyclopedia of Philosophy*. papers2://publication/uuid/FB24D0D8-99D4-4B73-B339-94D540532B41

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

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

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

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

Potochnik, A. (2017). *Idealization and the Aims of Science*. University of Chicago Press.

Samuels, R. (2019). Classical Computational Models. In M. Sprevak & M. Colombo (Eds.), *The Routledge Handbook of the Computational Mind* (pp. 103–119). Routledge.

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

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

1. “*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, p. 108). [↑](#footnote-ref-1)
2. Shea seems to think that cognitive scientists would have nothing to explain about Swampman: they are interested in explaining *successful* behavior, and without etiological functions Swampman can have no ends to be successful with respect to (Shea, 2018, p. 22). This isn’t an attractive characterization of cognitive science’s explananda. Cognitive science is interested in explaining various 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*. The patterns are still there, and are still interesting to human inquirers, so they will remain scientific explananda. And they will still be most fruitfully approached with the tools and methods of cognitive science, so they will remain *cognitive* scientific explananda.

It is also possible to include etiological questions in the domain of cognitive science by fiat. Scientific disciplines are not divided in an especially clean or rigorous way, especially ones like cognitive science. We could just stipulate that cognitive science includes the question, “How did the mind evolve?” But this replaces the philosophical problem with a districting dilemma, and clearly doesn’t get at the heart of the matter. (See also the discussion of unification and taxonomic pluralism in section 3.) [↑](#footnote-ref-2)
3. Of course, we might want to use evolutionary history to *suggest* possible representations or to explain why Davidson ended up the way he did. But that’s not using evolutionary history to *individuate* Davidson’s representations. No one says we should pretend Davidson doesn’t have an evolutionary history; the idea is that that evolutionary history doesn’t ground his representational states or constitutively determine their content. [↑](#footnote-ref-3)
4. It is easy to slip into circular arguments here, like some more sophisticated version of a claim that our explanations of Swampman would be less successful or unsuccessful because they would attribute properties (functions etc.) that he does not have. This is why it’s important (as the “real kinds” response emphasizes) to focus strictly on the science: the goals it sets itself, and whether its methods could meet those goals in the case of Swampman. This is also why I’ve discussed cognitive scientific *models* and *predictions* and so on rather than sticking to the philosophically more familiar notions like *generalization* and *inference*. It would be a mistake to lapse into the philosophical fiction of science as stating propositions and assessing their truth, or making generalizations and testing them for universality. Science — cognitive science at least — spends its time instead building models and assessing their predictive utility, interpretability, biological realism, computational tractability, the acceptability of their idealizing assumptions, and so on. It is with respect to cognitive science’s actual explanatory goals that we should assess our theories of its notions. [↑](#footnote-ref-4)
5. All this may be why Millikan herself casts teleosemantics, in recent work, as irrelevant to cognitive science: “But the cognitive scientist need not care about the etiology of the mechanisms that she/he studies. Cognitive science is the study of how the brains of healthy organisms contribute to management of their environments. The history of these mechanisms is irrelevant to studies of this kind. (Engineering principles are relevant of course.) Teleosemantics, as such, does not offer anything to the cognitive scientist” [Draft, permission to cite?]. [↑](#footnote-ref-5)
6. If we are sufficiently cold-blooded, we might try cutting off Swampman’s future as well as his past. It remains to be seen whether this would support a case against forward-looking accounts of functions as well. For some examples of teleosemantic and functional treatments of cognitive science that do not appeal to etiology, see Nanay (2014) and Cao (2012). [↑](#footnote-ref-6)
7. Is the deflationist stuck using the Swampman Counterexample? Perhaps, but there is also more for them to say here. Even for the deflationist there are important questions about *how representational notions are used in explanations, and why they are successful*. Those are the questions that deflationists take largely to *supplant* questions of the metaphysics of representation. And for the deflationist it is possible to answer those question by appealing to etiological properties. The conclusion here is that they should not. [↑](#footnote-ref-7)
8. Someone opposed to teleosemantics might press a non-deflationary version of the inference from *Swampman is explicable in representational terms* to *Swampman has representations*, namely through an understanding of representational explanations as *positing* representations in the brain: if explanations that posit representations in a system are explanatory of that system, this is good evidence that the representations are really there. I don’t consider this in the text because such a philosopher would already agree with my conclusions, but would be pressing them in a way much less likely to convince etiologists, who can again advert to the contrived nature of the counterexample and make familiar points about real kinds. The argument we’re considering attempts to define the kind *representational system* by considering the systems it explains: representations are something those systems share, and ex hypothesi they do not share in the possession of etiological properties. But the etiologist will want to limit our understanding of the systems representational explanation explains to the ones it *actually does* in the real empirical world, because those are the ones for which there is some *reason* it explains them, and therefore some reason to expect there is a natural kind at play. We can avoid all this by simply not casting Swampman as a counterexample. (Thanks to David Rosenthal for convincing me of the importance of this point to the overall dialectic.) [↑](#footnote-ref-8)
9. This is not to say we should stick with the ridiculous hypothetical. Millikan talks of biologists sniggering about the philosophical menagerie (1996, p. 115), and many of us have witnessed this (reasonable) response from cognitive scientists as well. Our interdisciplinary goals would be better served with less fantastical ways of getting at these issues, or with more scientifically-motivated ways of framing the strange bestiary we keep. [↑](#footnote-ref-9)
10. Thanks to Devin Morse for the illuminating example. [↑](#footnote-ref-10)
11. In connection with this: I said that Swampman is just as successfully explicable as Davidson, by all the same explanations. It remains open to say that when those explanations are applied to Swampman they are applied differently — perhaps they require an *idealization* of Swampman as having a certain kind of selection history. This may be a promising route for the teleosemantist, but the pressure is on to say what this idealization does for us. The following points can all be understood in this connection. [↑](#footnote-ref-11)
12. If our explanations of Swampman are successful, as the Swampenstein Argument has it, then we cannot say that *cognitive science* *itself* needs etiological kinds because it needs *real* intentionality. That would amount to claiming that the explanations of Swampman are not successful, undermined as they are by a lack of *real* intentionality. And then we return to an earlier point: we need to see some cognitive scientific goal that the explanations fail to serve in the case of Swampman because he lacks *real* intentionality — understood as requiring etiological kinds — that they do serve in the case of Davidson. [↑](#footnote-ref-12)
13. The Millikan quote from footnote 5 suggests this sort of retreat. [↑](#footnote-ref-13)