



# Swampman, teleosemantics and kind essences

David Papineau<sup>1</sup> 

Received: 4 January 2022 / Accepted: 4 November 2022

© The Author(s) 2022

## Abstract

One powerful and influential approach to mental representation analyses representation in terms of biological functions, and biological functions in terms of histories of natural selection. This “teleosemantic” package, however, faces a familiar challenge. Surely representation depends only on the present-day structures of cognitive systems, and not on their historical provenance. “Swampman” drives the point home. Suppose a bolt of lightning creates an intrinsic duplicate of a human being in a steamy tropic swamp; will not this creature be representing its surroundings, despite its lack of any selectional history? In this paper I shall answer this challenge by showing how a proper appreciation of the structure of natural kinds in general, and of mental representation in particular, implies that selectional histories are indeed built into the nature of mental representation. In particular, I shall address a recent argument by Peter Schulte against this general line of argument.

**Keywords** Representation · Teleosemantics · Natural kinds · Essences · Swampman

## 1 Introduction

One powerful and influential approach to mental representation analyses representation in terms of biological functions, and biological functions in terms of histories of natural selection (Dretske, 1986; Millikan, 1984; Neander & Schulte, 2021; Papineau, 1984). This “teleosemantic” package, however, faces a familiar challenge. Surely representation depends only on the present-day structures of cognitive systems, and not on their historical provenance. “Swampman” drives the point home (Davidson, 1987). Suppose a bolt of lightning creates an intrinsic duplicate of a human being in a steamy tropic swamp; will not this creature be representing its surroundings, despite its lack of any selectional history? In this paper I shall answer this challenge by showing how a proper appreciation of the structure of natural kinds in general, and of mental

---

✉ David Papineau  
david.papineau@kcl.ac.uk

<sup>1</sup> Department of Philosophy, King’s College London, Strand, London WC2R 2LS, UK

representation in particular, implies that selectional histories are indeed built into the nature of mental representation. In particular, I shall address a recent argument by Peter Schulte against this general line of argument (Schulte, 2020).

## 2 Teleosemantics

It will be useful to begin with a brief outline of the teleosemantic approach to mental representation.

Let me start with the abstract structure of the approach. Suppose we have a biological system composed of a “producer” that generates some state *R* and a “consumer” that reacts to this state with some behaviour *B*. Then, at first pass, the vehicle *R* will have the indicative content *C* if and only if *C* is the condition under which the behaviour *B* will fulfil the consumer’s biological function. (See in particular Millikan, 1984.)

Putting it in more intuitive terms, such a system treats the state *R* as *standing proxy* for *C*, in the sense that it behaves in a way that is appropriate to its function *given C*. In this sense the system *interprets* *R* as signifying *C*, by *behaving* in a way that is functional *given C*.<sup>1</sup>

This is a very general framework that can be applied across a wide range of cases. For example, the producer and the consumer could be different organisms, as when vervet monkeys issue alarm calls and their conspecifics respond in ways that are variously appropriate to snakes, eagles and leopards (Seyfarth et al., 1980). More commonly, the producer and the consumer can be within one organism, as when signals from the frog’s optic tectum direct its tongue-snapping system to fire in a certain direction (Lettvin et al., 1959); or when signals from the primate dorsal visual stream guide arm and hand movements (Goodale & Milner, 1992); or indeed when the products of human belief-forming mechanisms inform our conscious planning.

Note that the functions served by the consumers of a teleosemantic signal can be more fine-grained than mere survival and reproduction. Biological systems decompose into subsystems each with their own more specific functions (Neander, 1995). The function of the frog’s tongue-snapping system is to catch insects, the function of dorsally guided movements is to grasp nearby objects; the function of conscious human planning is to satisfy desires (Papineau, 2016).

As I said above, teleosemantics standardly understands biological functions “aetiologically” in terms of histories of natural selection. In the first instance, a trait *T* has function *F* just in case the past selection of *T* hinged on its causing *F*. Note that this schema too is flexible, and in particular need not be restricted to functions that depend directly on the intergenerational selection of genes (Macdonald & Papineau, 2006). For a start, phylogenetically novel functions can be *derived* from inherited mechanisms whose biological function is to produce traits that produce given effects (Millikan, 1984). Moreover, the intergenerational selection of *non-genetically but vertically inherited* cultural traits can also be a source of relevant aetiological functions (Jablonka & Lamb, 1999; Mameri, 2004). And finally *ontogenetic* neural selection is

<sup>1</sup> Note that I do not here suppose that the system interprets *R* by forming some further *representation* of its significance. That would of course be regressive in an analysis of representational content.

also arguably an independent source of relevant aetiological functions (Garson, 2019; Garson & Papineau, 2019).

Still, while the requirement that biological functions derive from histories of natural selection might be flexible in these ways, it is still a historical requirement. And so teleosemantics still faces the challenge of explaining why the current representational status of organisms should depend on their past history. Why should swampman's lack of a selectional provenance prevent it from representing anything? I turn now to this challenge.

### 3 Doing without history

Note first how the teleosemantic approach to representation has genuine explanatory content. It allows us to identify real patterns in nature. By assigning truth conditions to signals, it allows us to track, not just which behaviours will be prompted by the signals, but when those behaviours will lead to biological success.

The signal sent to the frog's brain by its eye prompts it to shoot out its tongue in a certain direction. That is one pattern. But now the ascription of a truth condition—*flying insect in a certain location*—points to a further pattern. In those cases where this condition is satisfied, not only will the frog shoot out its tongue in response to the signal, but it will succeed in catching an insect.

The point generalizes. Representation understood teleosemantically tells us not just how agents will behave proximally, but also when those behaviours will succeed in producing distal results.

Still, this point only highlights the challenge about history. What work is the historical dimension of teleosemantics doing in charting these explanatory patterns? Surely what matters for the success-involving patterns is whether we can identify circumstances under which certain animals systematically succeed in achieving certain distal ends in the contemporary world, not the historical provenance of the structures that allow them to do this.

Note in this connection that a number of other philosophical accounts of representation agree with teleosemantics about the way representation enters into contemporary explanatory patterns, and diverge only in not building selective history into the nature of representation. Thus “success semantics” and “non-aetiological teleosemantics” both analyse the truth conditions of mental states in terms of their role in the systematic achievement of distal ends, but deny that the relevant ends need to be understood as biological functions arising from histories of natural selection. Success semantics focuses on belief-desire psychology and takes the truth conditions of beliefs to be those circumstances in which the actions they prompt will satisfy desires—where the satisfaction conditions of desires are explained not in terms of aetiological functions but simply as those outcomes which typically result from desire states and serve to quench them (Ramsey, 1927; Whyte, 1990). Non-aetiological teleosemantics follows orthodox teleosemantics in not being restricted to systems with belief-desire psychological structures, and moreover in focusing on the way representations guide consumers in fulfilling their “biological functions”, but sees no need to understand “biological functions” in this context as meaning outcomes for which the consumers

have been historically selected, as opposed to outcomes that contribute systematically to the flourishing of organisms in the contemporary world (Abrams, 2005; Nanay, 2014).

By their nature, these alternative theories would seem to share all the explanatory success of aetiological teleosemantics without the extra historical commitment. Indeed they would seem to have an advantage over orthodox teleosemantics, in that they can offer representational explanations given any structures in which inner states facilitate the systematic achievement of distal results, even in cases where those structures are not the upshot of historical selection. (At the same time, they can of course invoke histories of selection to *explain* current psychological structures when such explanations are available—but they will then regard this history as the *cause* of the representational structures, and not as *constituting* them.)

Swampman offers graphic support to this challenge. A bolt of lightning in a tropical swamp by freakish chance creates a perfect molecule-for-molecule duplicate of me. At first pass, it seems uncontentious that this swampman would have representational powers. Perhaps there are doubts about it representing specific entities like *my wife Rose*, or *Nelson's Column*, despite the intrinsic match between its cognitive states and mine, given that it has never enjoyed any causal contact with those entities. But it would seem odd to deny that it can represent such things as the wetness of the water it is standing in and the attractions of sitting on the dry bank, and that these representations can guide its behaviour in ways appropriate to its ends. Yet by hypothesis swampman lacks any selectional history. Once more, it seems wrong to build selectional history into representation.

#### 4 An a posteriori reduction

Despite the strength of the case against, I would like to defend the way orthodox teleosemantics builds selectional history into the nature of representation. In a way, this point is perhaps peripheral to a philosophical understanding of representation. The first thing to grasp about representation is the way that some mental states stand proxy for external circumstances in prompting behaviour that will produce certain results if those circumstances obtain. That much is agreed between orthodox teleosemantics and the non-historical alternatives outlined in the last section. Moreover, given this, I am happy to allow that it can be perfectly productive to analyse representation in terms of biological functions without any commitment to those functions having a historical basis.

Still, I think that it is no accident that the original proponents of teleosemantics (let us drop the “orthodox” and read “teleosemantics” as aetiological from here on) all proposed analyses that did make selectional history essential to representation. I shall argue that this commitment flows from the way representational systems are constituted as a unified natural kind. Aetiological histories might not add much to our understanding of the way representation operates in the contemporary world, but they are crucial to the status of representation as a natural kind.

The initial moves in this kind of teleosemantic response to swampman are familiar enough. Teleosemantics is intended as an a posteriori reduction of the property

of representation itself, not as an a priori analysis of our *concept* of representation. Swampman might have what it takes to satisfy the everyday pre-theoretical concept of representation. (He talks like a representer, he walks like a representer...) But that does not establish that he has the underlying properties that a posteriori investigation has revealed to constitute the real nature of representation. (Millikan, 1996; Neander, 1996; Papineau, 2001.)

Locke distinguished real essences from nominal essences. We might think of the signs by which we initially pick out instances of representation as its nominal essence. (Does the agent have states which prompt behaviour in ways appropriate to given external circumstances?) However, the real essence of representation, according to teleosemanticists, involves selectional histories. A posteriori scientific investigation has shown that crucial underlying property shared by all representers is that the relevant states feature in producer–consumer systems that have been *selected to achieve certain ends*.

It's like water and H<sub>2</sub>O, say the teleosemanticists. The nominal essence of water comprises the surface properties by which ordinary people recognize instances—odourless, colourless, tasteless, flows in rivers, and so on. But science has moved beyond these properties and identified the real essence of water, namely, that it is composed of H<sub>2</sub>O molecules. Similarly, science has shown us that the underlying essence of representational systems is their selectional aetiology.

So at first pass it seems open to teleosemanticists to argue that swampman is no more an objection to their theory than Hilary Putnam's XYZ is an object to atomic chemistry (Putnam, 1973). Putnam posited a world in which the local odourless, colourless, and tasteless liquid had some alien composition—"XYZ"—rather than being made of H<sub>2</sub>O. But Putnam didn't of course intend this as a refutation of chemistry. On the contrary, he wanted to bring home the point that a possible substance satisfying the nominal essence of water could lack the real essence that science has revealed to constitute genuine water. XYZ would be fake water, not real water. Similarly, the teleosemanticists can argue, swampman would be a fake representer, not a real one.

Of course, all this hinges on swampman being purely imaginary, rather than real (Papineau, 2001). If there were lots of swampcreatures in the actual world, satisfying the nominal essence of representation, but lacking any selectional histories, then science wouldn't have been able to maintain that selection is the real essence of representation. Sectional histories wouldn't have been an underlying property shared by all representers in the first place. But that's all right. If there were lots of odourless, colourless, tasteless XYZ in the actual world, then the nominal kind water wouldn't have turned out to be H<sub>2</sub>O either. But of course neither swampman nor XYZ are real. They are both merely imaginary constructs, and as such do not threaten a posteriori theses about the real essence possessed by representation and water in this world.

## 5 Schulte's objection

So far, so good. Still, it is scarcely enough for teleosemanticists simply to assert that selectional histories turn out to provide the real nature of representation, in the way that H<sub>2</sub>O turned out to provide the real nature of water. This is surely something to be

shown, not assumed. On the face of it, the two cases look very different. It is by no means obvious that selectional histories stand to representation in the way that H<sub>2</sub>O stands to water.

In his paper “Why Mental Content is not Like Water: Reconsidering the Reductive Claims of Teleosemantics” (2020), Peter Schulte objects on these grounds to my (2001) defence of teleosemantics as an a posteriori reduction of representation. Schulte argues that I fail to show that selectional histories constitute the same kind of essence as H<sub>2</sub>O.

Schulte starts by observing that those of us who run the a posteriori defence of teleosemantics standardly assume that our pre-theoretical concept of representation picks out representational states as states that play a certain *causal-dispositional role*. Our idea is that everyday thought views representational states as states that are typically produced by certain external causes, typically generate certain forms of behaviour, and typically lead to distal success when their truth conditions obtain. Now it is true, grants Schulte, that in the actual world all states that fit this general specification are also states that have been designed by selectional histories to satisfy this pre-theoretical role. But, he observes, a posteriori identity is by no means sufficient for reduction. After all, the properties *water* and *liquid that covers 71% of the earth* are a posteriori identical, but this certainly doesn't suffice to show that *covering 71% of the earth* is the real essence of *water*.

The reason *liquid composed of H<sub>2</sub>O molecules* offers a reduction of water, Schulte continues, is not just that the two properties are a posteriori identical. Rather it is that the former property *explains* the causal-dispositional properties which enter into our pre-theoretical concept of water. Schulte shows in some detail how the chemical composition of water accounts for a number of the causal-dispositional features by which we initially recognize water.

Schulte draws a general moral. If one property is to provide an a posteriori reduction of another, it must explain the nominal surface features by which we pre-theoretically recognize the latter. But the standard a posteriori reduction defence of teleosemantics cannot satisfy this demand, he argues. This standard teleosemantic defence, as observed above, starts from the assumption that our pre-theoretical concept of representation picks out states that play a certain *causal-dispositional role*. But selectional histories seem quite unsuited to explain such roles, he argues. After all, causal-dispositional roles are displayed in the present, and as such seem to demand explanation in terms of intrinsic properties of their bearers. That is why H<sub>2</sub>O composition is suited to explain the causal-dispositional role of water. It is an intrinsic property of water samples. But having been *selected for certain effects* is a highly relational feature of representational states, and as such is not suited to explain the current causal-dispositional features of those states. Those causal-dispositional features are due to the intrinsic properties of representational states, not their historical properties. Swampman once more emphasises the point. Swampman's states have the same causal-dispositional features as the representational states of ordinary humans, courtesy of swampman's intrinsic duplication of an ordinary human. Given this, an

explanation of those causal-dispositional features should invoke some intrinsic property common to swampman and humans, not some historical feature that differentiates them. Or so Schulte argues.<sup>2</sup>

In the remainder of this paper, I shall respond to this argument of Schulte's. I shall not contest the idea that our pre-theoretical concept identifies representation in terms of a causal-dispositional role. But I shall deny that this means that representation can only be reduced to some intrinsic property rather than to a relational historical one. My argument will hinge on a general analysis of what it takes for some underlying property to constitute the nature of a natural kind. I shall be arguing that Schulte addresses this issue with an insufficiently wide focus.

Before proceeding, however, I would like to comment briefly on an alternative suggestion that Schulte makes on behalf of reductive teleosemantics. He suggests that teleosemanticists will do better to argue that our pre-theoretical concept of representation is not exhausted by causal-dispositional features. Rather they should hold that this nominal concept also involves a *normative* element. On this suggestion, the everyday concept commits us to the normative thought that by judging truly we meet a certain standard, and by judging falsely we fall short of that standard. And this normative dimension of representation, suggests Schulte, is indeed happily explained by reference to histories of natural selection—which will then deliver the required teleosemantic conclusion that a selective aetiology contributes to the underlying nature that explains the nominal features of representation (Schulte, 2020 Sect. 5).

I have some sympathy with this line of thought, but I don't think it will serve in the current argumentative context. The difficulty is to find a form of normativity that is both explained by selective history and plausibly built into the everyday thinking about representation.

As Schulte realizes, "strong" prescriptive normativity—normativity that per se provides reason for action—does not fit this bill. Maybe some such strong notion is indeed part of the everyday notion of representation. You really ought not to judge falsely. But if so this is not something that can be explained by selective provenance. That something has been designed by evolution to serve a given purpose by no means shows it really *ought* to be so deployed. (Our knuckles have arguably been designed for hitting, aspects of the male brain have arguably been designed to foster sexual predatoriness,...)

In response to this point, Schulte suggests that his proposed teleosemantic strategy will do better to appeal to "weak" normativity, the kind of normativity that comes with the contingent observance of some standard, as when in certain circles it is viewed as "correct" to wear ties with suits and "incorrect" to wear socks with sandals (2020, p. 2284). But the trouble now is that selective histories seem quite unnecessary to account for representation displaying this kind of weak normativity. After all, non-aetiological accounts of representation, of the kind gestured at in Sect. 3 above, also ascribe truth conditions to mental states, and so distinguish between true representations and false misrepresentations. And this would already seem to provide a perfectly good non-historical standard against which the former representations can be counted

<sup>2</sup> For some related objections to Papineau 2001, see also Kim (2021).

as “correct” and the latter as “incorrect”, where these evaluative terms are understood as merely conveying “weak” normativity.

Then there is the normativity of *design*. You are “supposed” to judge truly because that is what your representational abilities have been designed for, just as you are “supposed” to grasp the tennis racquet by the handle because that is what the handle was designed for. This kind of representational normativity does indeed call for explanation in terms of selectional histories. What else could fill out the idea that our representational mechanisms have been designed for certain purposes except an appeal to the way natural selection favours traits that produce those results? Still, is the idea that our representational abilities have been *designed* for certain purposes part of the everyday pre-theoretical idea of representation? Maybe so, maybe not. However, in the present argumentative context, teleosemanticists are in no position to assume this. After all, their current opponents are precisely those non-historical thinkers who hold that everyday representational thinking does not incorporate any idea of historical design origin. The contention of these non-historicists was specifically that representation’s role in everyday explanations of behaviour and success hinges solely on contemporary structures, and they backed up this point by observing that our pre-theoretical concept of representation applies happily to swampman, despite its lack of any design history.

It is true that we will need to bring in selectional history if we want to explain the sense in which representational mechanisms have been designed. But this only shows that selection is essential to representation if it is already granted that design is part of our pre-theoretical concept of representation—and this is precisely what the teleosemanticists’ non-historical opponents deny.

## 6 Natural kinds

There is another way for teleosemanticists to respond to Schulte’s challenge. They can query Schulte’s assumption that an intrinsic property is needed to explain the causal-dispositional nominal essence of representation, and argue that that selectional histories are in fact precisely what we need for this explanatory task. The key here will be a more refined understanding of the general way in which the real essences of natural kinds explain their nominal characteristics. I shall argue that, contra Schulte’s assumption, what kind essences offer is not a case-by-case explanation of why the nominal characteristics are displayed *within* each instance of the kind. Rather they explain why those nominal characteristics cluster in the way they do *across* instances of the kind—that is, they explain why things that display some of those characteristics generally display all the others (Godman et al. 2020). In the case of chemical substances like water, it is indeed an intrinsic property that explains this co-incidence. But with other sorts of kinds, as we shall see, it is precisely relational historical properties that explain the relevant co-incidences.

At this point, it will be useful to make some general points about the structure of natural kinds. I shall take it that the defining characteristic of natural kinds is that their instances share a *range* of properties. Chemical substances provide a paradigm illustration. All samples of water under the same conditions have the same density,

boiling point, melting point, heat conductivity, electrical conductivity, proportional dispositions to combine with other substances, and so on. All samples of gold, or sulphuric acid, or any other chemical substance, display a similar range of shared properties.

Chemical substances are not the only kinds. Biological species also have instances that share many properties. For example, all horses are alike in eating grass, growing manes, having uncloven hooves, and many other behavioural, anatomical and physiological features.

The same applies to higher biological taxa. They too share many properties. For example, all mammals share fur, sweat glands, milk glands, and other anatomical features. Note how the features common to a higher taxon will be a subclass of those common to its subordinate taxa. The features common to all horses include those common to all mammals.

Again, astronomical objects form kinds. All main sequence stars (those powered by hydrogen fusion) are spherical, in hydrostatic equilibrium, radiate energy of certain wavelengths, and so on. Other astronomical kinds include red giant stars, white dwarf stars, neutron stars, and supernovae.

In due course I shall also consider “functional kinds”. I shall be arguing that representational systems constitute a functional kind.

With most kinds, there is a single property, possessed by each instance, that explains why all the other common properties also appear across the instances. With chemical substances, this will be their molecular constitution. The molecular composition of water, and gold, and sulphuric acid, and so on, explains why the instances of these substances share so many other features. With astronomical kinds, it is their common internal physical constitutions that explain the further shared properties of their instances. We shall return in a moment to the question of what plays this role for other sorts of kinds.

In what follows I shall talk about these single properties as “super-explanatory properties” because of the way that they explain all the other features shared by instances of the relevant kind (Godman et al. 2020).

We can usefully view super-explanatory properties as consequences of the principle of the common cause. This principle asserts that if some A is correlated with B, then either A causes B, or B causes A, or they have a common cause. The defining feature of natural kinds, if you think about it, is that they involve a multiplicity of correlations. For any kind, there will be a bunch of properties F, G, H... such that the presence of any increases the probability of all the others. But at the same time these F, G, H... do not standardly cause each other. The electrical conductivity of water does not cause its density nor vice versa, the manes of horses do not cause their hooves nor vice versa, and so on. So the principle of the common causes tell us that these correlated properties must have a common cause. The instances of the kind must possess some underlying super-explanatory property that accounts for all the other properties displayed by the kind instances (Godman et al. 2020, pp. 319–320).

On Richard Boyd’s influential account, natural kinds are “homeostatic property clusters” (Boyd, 1991). I agree entirely that kinds are “property clusters”. Indeed, as indicated above, I take the sharing of multiple properties to be definitive of natural kinds. But I disagree with Boyd’s emphasis on homeostasis, in the sense of some

self-regulating feedback mechanism. Perhaps there are some few natural kinds which owe their profile to such homeostatic processes. These would be cases where the F, G, H, and so on characteristic of the kind all cause each other, rather than being independent effects of a common cause. Meteorological phenomena like cyclones might be a case in point. Again, some psychological syndromes like depression might plausibly fit the bill. But these are the exception rather than the rule (Godman et al. 2020, pp. 320–321). In the great majority of kinds, the clustering of properties is due to one super-explanatory common cause, rather than any homeostatic process of reciprocal causation. We have already seen how the common features of chemical and astronomical kinds stem from super-explanatory core properties. And we shall see shortly how the common features of biological taxa and “functional kinds” stem similarly from super-explanatory common causes.

When a kind does have a super-explanatory property, it is natural to regard this property as providing the “real essence” of the kind. The essence of water is its chemical structure. The essence of supernovae is their runaway nuclear fusion.

Philosophers of science who work on natural kinds often emphasize their complexity and lack of sharp boundaries, and in consequence are suspicious of any suggestion that they have “essences” (e.g., Häggqvist & Wikforss, 2018; Hull, 1965; Sober, 1980; Wilson et al., 2007). As they see it, the whole idea of essences smacks of some outmoded scholastic metaphysics. However, while it is true that kinds are often complex and can allow borderline cases, this is no reason to deny that they are standardly held together, so to speak, by their super-explanatory properties. It is because the instances of a typical kind all have a single underlying property that they share their other common properties. This gives us good reason to view this super-explanatory property as providing the essence of the kind, and the kind as therefore reducing to this unifying property.

Does not talk of “essences” imply modal consequences which go beyond anything ensured by structures of causal relations in the actual world? As it happens, talk of “essences” carries no modal implications that are not already implicit in the idea of a one property reducing to another, as I shall explain in my final section. Still, these modal niceties can wait until later. Until then, let us simply focus on the idea that standard natural kinds are held together in the actual world by super-explanatory properties.

## 7 Functional and historical kinds

What about biological species and other biological taxa? The idea that biological species have essences is a particular target of those philosophers who regard talk of kind essences as outmoded. As they see it, talk of biological essences is discredited by the population thinking that is central to the modern understanding of species as evolving under selective pressures. But this is too quick. Modern biological thinking might have discredited certain traditional notions of species essence, but it is an over-reaction to discard the idea altogether. I shall show that certain relational properties of biological species and other taxa can be viewed as super-explanatory properties of

the sort outlined in the last section, and to this extent can be viewed as constituting essences for these kinds.

Before considering biological species and broader taxa, however, it will be helpful briefly to consider those rather more superficial biological kinds that result from convergent evolution and owe nothing to common ancestry. I shall call these “functional kinds”. For example, consider the category of *aerial insectivore*. All the swallows, martins, swifts, insectivorous bats, dragon-flies and other flying insect-eaters share a range of properties, including acute sensory systems, ability to swoop, and large beaks or mouths. Convergent natural selection has operated on all these animals to give them the traits needed to catch flying insects. Accordingly the aerial insectivores display a range of shared properties and so qualify as a natural kind. More generally, other biological categories that have so resulted from convergent evolution will be similarly constituted as functional kinds.

It is true that functional kinds are relatively thin, in that their instances only share a few properties, by comparison with the rich commonalities shared by the members of the same biological taxon (Brigandt and Griffiths 2007, Godman and Papineau, 2020). For this reason biologists typically attach more weight to *homological* traits due to a common ancestry than to *analogical* traits resulting from convergent selective pressures. Even so, the members of functional kinds do display a non-trivial range of shared properties. And, in line with this, we can regard the common selective pressures to which they have been subject as the super-explanatory common cause which accounts for their all sharing this set of properties, and in this sense constitutes the “essence” of aerial insectivores. Nothing can be an aerial insectivore unless it has been shaped by such selective pressures, and anything so shaped is an aerial insectivore.

What then about species and other biological taxa (genera, families,...) that do share a common ancestry? What if anything plays the super-explanatory role for these homological biological categories? A first thought might be that the shared properties of the members of homological taxa can similarly be explained by the common selective pressures to which they have been subject. But this does not fully fit the bill. The members of any biological taxon will share a wealth of features beyond those that can be explained by the common selective pressures on them. This is because natural selection has to work with what it is given, is limited to the options offered by mutation, and can be deflected by genetic drift. This ensures that the members of biological taxa share many non-functional features. We need something beyond common selective pressures to account for these.

Ruth Millikan distinguishes “historical” kinds from “eternal” kinds like gold and “functional” kind like aerial insectivores (Millikan, 1999, 2000). Where eternal kinds owe their correlated features to some common intrinsic property, and functional kinds owe them to common selective pressures, historical kinds owe them to their all being copied from a common origin.

For example, consider all the copies of any work of literature, like *Moby Dick*, say. These copies obviously share many features, from their first word onwards, but this isn’t because they have some common internal property or because they have converged on the same form due to common selection pressures. Rather it is simply because they have all been copied from the same original.

This offers a natural account of the commonalities displayed by the members of any biological taxon. They are all descended from the same founding organisms. This is why they all share so many features, including many non-functional ones. This is a simple consequence of the fact that those features were all present in the taxon's founding population. In line with this, it is natural to view their common ancestral origin as providing the "essence" of any given biological taxon. What makes you a tiger, say, is that you are descended from the founding tiger population. Nothing without that descent is a real tiger, even if it is superficially similar, and anything with that descent is a tiger.

In a series of recent papers Michael Devitt has argued, against this historical account, that biological taxa have intrinsic "essences" in the form of shared genetic material and associated developmental mechanisms (Devitt, 2008, 2010, 2018, 2021). However, while Devitt's genomic developmental programmes can explain why a taxon's shared properties arise *within* each member, they don't satisfactorily explain why the properties are correlated *across* the taxon members. This is because a genome, at first pass, is a *conjunction* of different genes, each of which explains some phenotypic characteristic. So Devitt only can explain why the phenotypic characteristics are correlated across the taxon by assuming without further explanation that the relevant genes are themselves so correlated, and so leaves us with an unexplained coincidence at the genetic level (Godman et al. 2020, pp. 322–326).

I do grant that, in the case of sexually reproducing organisms that develop from single fertilised zygotes, there is a need to explain why all the organisms in a taxon develop similarly into adults with a set of shared features. And Devitt's genomic developmental programmes do provide an answer to this explanatory question. But this is not the explanatory question posed by natural kinds in general. After all, chemical substances do not grow from fertilised seeds. Nor do all the members of a functional kind share a single developmental programme. Rather the more general puzzle raised by natural kinds is to explain why it is not a coincidence that the same cluster of properties appears *across* all kind instances, not why the characteristic kind properties develop *within* each instance. Historical origins discharge this explanatory obligation for biological taxa. But Devitt's genomic programmes do not, because they invoke another unexplained coincidence in their explanation of shared biological phenotypes.<sup>3</sup>

## 8 Representation as a functional kind

Let me now return to representation. We can usefully view animal representation as itself a natural kind. There are many different examples of representational systems found in many different animals, and they all display a striking range of shared properties.

---

<sup>3</sup> Does not an appeal to a common historical origin itself assume a coincidence, as when we explain the shared characteristics of extant tigers by positing the original co-incidence of those same characteristics in the founding tigers? But particular coincidences do not call for explanation in the same way as general ones. It's only to be expected that Jim and John will sometimes wear the same colour shirt. But if their shirts are *generally* the same colour, then either one shirt colour is causing the other or vice versa, or the colours have a common cause.

- (1) For a start, they all share the basic structure in which consuming sub-systems respond to internal states R with behaviours B that are appropriate to distal circumstances C, while producing sub-systems ensure that there is an adequate correlation between the internal R and the distal C. This then allows the joint systems better to coordinate their performance of B with the presence of the external circumstances C.
- (2) Moreover, representational systems will normally involve a *range* of different Rs, each standing proxy for one of a related range of Cs and each prompting a behaviour B appropriate to that C—as when the frog’s optic tectum has a range of different Rs for moving insects at *different* directions and distances, or the dorsal visual stream has different Rs for *differently* shaped objects in *different* positions. In cases like this, the representation-consuming system will be able to serve its function across a range of related circumstances and not just in one (Shea, 2018, chapter 2).
- (3) Representational systems will also typically be able to produce any given R in response to its relevant C robustly across a range of different peripheral stimuli. For example, a monkey will be able to visually identify a leopard from different angles, in different lighting conditions, with different parts occluded, and so on. This kind of robustness thus allows representational systems to identify a given distal C across widely varying conditions of observation (Burge, 2009, 2010).
- (4) A representational producer will standardly contain further nested sub-systems whose purpose is to produce *intermediate* representations. For example, sub-systems in primate vision systems respond to luminance and chromatic discontinuities to produce edge representations, and similarly respond to stereoptical information to produce distance representations. These edge and distance representations are then themselves processed by a consumer sub-system whose function is to produce representations of three-dimensional form, where these latter representations are in turn consumed by the behaviour-directing system. In general, we can view intermediate representations of this type as serving as premises in inferences to further more holistic representations (Papineau, 2016).
- (5) Throughout the animal world, representational systems will be fine-tuned by processes of classical and instrumental conditioning. Classical conditioning will serve to augment the range of stimuli that lead to the production of some R, in line with the robustness covered by point (3), while instrumental conditioning will select those responses B to R that are best-suited to serving the consumer’s function.

So animal representation itself has all the hallmarks of a natural kind. A range of different properties cluster together in its instances. Any system that has any of the above five properties can be expected to display the others. Perhaps there are more characteristic features of representation, but these five suffice to make the point.

As with any natural kind, we can ask what grounds this clustering of properties. Why will any system that displays one typically display the others? And the answer seems obvious enough. Representation is a *functional kind*. The similarities in representational systems across animals are not due to representing animals all descending from some common ancestor. Rather representational systems have evolved separately

on many different occasions. For example, sophisticated eyes have appeared independently on some dozens of biological lineages. The same goes for mechanisms for hearing, olfaction, echolocation and other sensory systems, and indeed for amodal cognitive representations in those species that have developed them. That is why it is not a coincidence that these different representational systems are all similarly structured, in line with points (1)–(5). The similarities are due to common selective pressures. The different representational systems have all been shaped by natural selection to make animals better at tailoring their behaviour to their circumstances. That is why they all respond to a suite of different circumstances, recognised robustly via varying stimuli, use intermediate representations to arrive at behaviour-directing ones, and fine-tune their responses by learning.

As with other functional kinds, we can thus view these common selective pressures as the super-explanatory property of the kind. Their common selective provenance is the property that explains why all instances of representation share a range of other properties. Given this, it is natural to view this selective provenance as the providing the “essence” of representational systems. Something is a representational system just in case it has been shaped by natural selection to enable animals to tailor their behaviour to their circumstances.

## 9 Answering Schulte’s objection

We are now in a position to answer Peter Schulte’s objection. Schulte assumes that, if the nominal features of some kind are “causal-dispositional”, then only some intrinsic property of the kind’s instances can qualify as its real essence. His idea is that an explanation of causal-dispositional behaviour must necessarily appeal to some current feature of the instances, not some relational historical property. It’s what the instances are made of that explains how they currently behave, he argues, not where they come from.

We can now see why this argument does not go through. We can agree that the nominal features by which we pre-theoretically identify instances of representation are all causal-dispositional, to do with the way representational systems currently operate. These will no doubt be the more salient aspects of those causal-dispositional properties shared by all representational systems and listed in the previous section. But it does not at all follow that the real essence of representation must therefore be some intrinsic feature of representational systems.

An intrinsic feature might be needed to explain why the nominal properties of some kind develop *within* any given instance. But, as we have seen, it doesn’t follow that an intrinsic property is always the right way to explain why those nominal properties *cluster across* all kind instances. Sometimes it is relational properties that play this super-explanatory role, and in such cases they constitute the essence of the kind.

It is true that the essential properties of chemical substances and other “eternal kinds” are intrinsic to the instances. But we have seen how other kinds have relational super-explanatory essences. In the case of biological taxa and other historical kinds, it is shared origins that explain why so many properties cluster together in

the instances. And with functional kinds this super-explanatory role is played by the selective pressures responsible for convergent evolution.

So I now have an answer to Schulte's challenge. Even if nominal features of representation are all causal-dispositional, this does not exclude representation having a relational historical essence. As I have argued, representation is a functional kind. Different representational systems share a range of striking similarities due to the way they have independently evolved under similar selective pressures. This shared selective provenance thus constitutes the essence of representation. And this now vindicates the teleosemantic analysis of representation, which was always the view that representations are states which have been selected to serve the function of gearing behaviour to circumstances in pursuit of some biological end.

## 10 Coda: rigidity and essences

In the 2001 paper that is the target of Schulte's argument, I did not only articulate a defence of teleosemantics as an a posteriori reduction of representation, but also sought to distance myself from various Kripkean assumptions. In particular, I aimed to stand neutral on whether terms like "belief" function as *rigid designators*. As I saw it, teleosemantics as such has no need to take a stand on the workings of English terms for representational states. Teleosemantics is a substantial theory about the kind of states that play representational roles in the actual world, I said, and so can by-pass any controversial claims about rigid designation. It is committed to representation being realized by states with selectional histories in the actual world, but that leaves it open what teleosemantics requires of other possible worlds. Thus it would be perfectly consistent with the core commitment of teleosemantics, I held, to allow that terms like "belief" behave "flaccidly" in modal contexts, functioning like descriptions with narrow scope, and so pick out, with respect to other possible worlds, whatever states play the belief role in that world. On this option, then, if there were a significant population of Swampmen in some other possible world, then they would be *believers* all right, since in their world the belief role wouldn't be realised by states with selectional histories to start with.

In a response to my paper, David Braddon-Mitchell and Frank Jackson (2002) observed that, so understood, teleosemantics seemed little different from the "analytic functionalist" view that terms like "belief" are associated a priori with a causal role description and refer to whichever cerebral states satisfy that description. This position similarly leaves it open whether we should treat such terms as referring rigidly in modal contexts to their actual satisfiers, or referring flaccidly to whatever would satisfy them under various counterfactual suppositions. Teleosemanticists might add an insistence that we view the actual realizers as incorporating a selectional provenance, but Braddon-Mitchell and Jackson's viewed this addition as debatable, and in any case as just giving us another variant of analytic functionalism.

From the perspective of the present paper, much of this debate now looks misplaced. The issue is not to identify which "realizer" plays the belief "role". Rather the question is simply the nature of representational properties. What sort of properties are *believing* and *representing*?

Setting things up in terms of analytic “roles” and their “realizers” sends us off in the wrong direction. It suggests the referent of terms like “belief” and “representation” are something like brain states, instantiations of physical properties in the brain, and that the “roles” associated with these terms are simply specifications of contingent properties by which we pick out these brain states.

However, physical brain properties as such are not representational properties. Instantiations of physical properties might be *vehicles* of representation, but the property of *representing/believing some content* depends on more than the presence of such a vehicle. This representational property has a relation to some truth condition built into it, and the issue that teleosemantics is addressing is the nature of this relational representational property. The deployment of some internal physical vehicle might be needed for representation, but even so the property of *representing/believing* itself should not be equated with the physical vehicle property, but rather with the more generic property of *deploying some vehicle related to the relevant content*.

The argument of the present paper is that the property of *representing* some condition is the same property as *housing a vehicle that has been selected to gear responses to* that condition. “Belief” and other representational kind terms are simply names for this representational kind. And so, as kind terms, they continue to name this kind in modal contexts. We shouldn’t think of these kind terms as a reference-fixing descriptions for brain states, descriptions that might take wide or narrow scope in modal contexts. They are simply names for the kinds that display multiple commonalities in the actual world, and thus continue to function as such names in modal contexts. Names by their nature are rigid.

Do I not owe some defence of the assumption that kinds have *essences*, and that kind terms track those essences across other possible worlds? But this is no extra assumption, beyond the claim that representation *is* the property of having a certain selectional history.

To see this more clearly, go back to the idea that kinds are categories whose instances share multiple nominal properties. For example, all samples of water are odourless, colourless, tasteless, boil at 100 °C, and so on. Now we can ask what the property of being water *is*. One possible answer would be to equate it with some conjunction of those nominal properties. To be water is just, say, to be *odourless, colourless, tasteless, and boil at 100 °C*. But a better answer is to equate the property of being water with the property of *being composed of H<sub>2</sub>O molecules*. This equation gives us the property that pulls all the instances of water together in the actual world, so to speak. It is the property that accounts for water’s status as a significant kind. We can convey this by saying that *being composed of H<sub>2</sub>O* is the “essence” of water, if we like, but we shouldn’t think of this as pointing us to some special extra hidden feature of water that is somehow held fixed across modal space. It’s simply a matter of specifying what property *water* is, and so what will be referred to in any context by a term that names that property.<sup>4</sup>

<sup>4</sup> This is not to say that modal thoughts can play no role in *leading us* to these actual property identities. Property identity is a matter of necessary coextensiveness across modal space. And it is arguable that what drives the identification of kinds as such with their super-explanatory core properties is the idea that something *wouldn’t* be a member of the kind *if it lacked* that super-explanatory property, and *would* be a

Similarly with *representation*. The property that pulls together all the different instances of representation, with their many shared features, is the property of having been selected to gear behaviour to conditions. In line with this, we should identify the property *representing* with the property of *housing a vehicle that has been so selected*. And if we talk about this property as the “essence” of representation, we are not talking about some extra hidden feature that is tracked across possible worlds, but just saying what *representation* is.

Looking at things this way, the implications for swampman are clear. He definitely wouldn’t represent. The term “belief” names a kind constituted by a past sectional history, and none of swampman’s states would be instances of that kind. “Belief” is not an implicit description of whatever brain state satisfies some everyday descriptive role, with the implication that some non-selected brain state might satisfy that role in a possible swampman scenario. It’s just a name for the relational representational kind constituted by selectional history—a kind that no merely possible swampman would instantiate.

**Acknowledgements** I have delivered versions of this paper at King’s College London, MAPS in New York, the Edinburgh Philosophical Society, and the Federal University of Rio de Janeiro. I would like to thank all who responded on those occasions, and in particular Julien Dutant, Justin Garson, Peter Schulte, Sérgio de Souza Filho, and Nick Shea.

## Declarations

**Conflict of interest** The author confirm that this paper does not involve me in any conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abrams, M. (2005). Teleosemantics without natural selection. *Biology and Philosophy*, 20, 97–116.
- Boyd, R. (1991). Realism, anti-foundationism and the enthusiasm for natural kinds. *Philosophical Studies*, 61, 127–148.
- Braddon-Mitchell, D., & Jackson, F. (2002). A pyrrhic victory for teleonomy. *Australasian Journal of Philosophy*, 80, 372–377.
- Brigandt, M., & Griffiths, P. (2007). The importance of homology for biology and philosophy. *Biology and Philosophy*, 22, 633–641.
- Burge, T. (2009). Perceptual objectivity. *Philosophical Review*, 118, 285–324.
- Burge, T. (2010). *The origins of objectivity*. New York: Oxford University Press.
- Davidson, D. (1987). Knowing one’s own mind. *Proceedings and Addresses of the American Philosophical Association*, 60, 441–458.

Footnote 4 continued

member of the kind *if it had it*—since, after all, the super-explanatory property is what is responsible for all the other shared properties that render the kind significant. (Godman et al. 2020, Sect. 10.).

- Devitt, M. (2008). Resurrecting biological essentialism. *Philosophy of Science*, 75, 344–382.
- Devitt, M. (2010). Species have (partly) intrinsic essences. *Philosophy of Science*, 77, 648–661.
- Devitt, M. (2018). Individual essentialism in biology. *Biology and Philosophy*, 33, 39.
- Devitt, M. (2021). Defending intrinsic biological essentialism. *Philosophy of Science*, 88, 67–82.
- Dretske, F. (1986). Misrepresentation. In R. Bogdan (Ed.), *Belief: Form, content and function*. New York: Oxford University Press.
- Garson, J. (2019). *What biological functions are and why they matter*. Cambridge: Cambridge University Press.
- Garson, J., & Papineau, D. (2019). Teleosemantics, selection and novel contents. *Biology and Philosophy*, 34, 36.
- Godman, M., Mallozzi, A., & Papineau, D. (2020). Essential properties are super-explanatory: Taming metaphysical modality. *Journal of the American Philosophical Association*, 6, 316–34.
- Godman, M., & Papineau, D. (2020). Species have historical not intrinsic essences. In A. Bianchi (Ed.), *Language and reality from a naturalistic perspective: Themes from Michael Devitt* (pp. 355–70). Cham: Springer.
- Goodale, M., & Milner, A. (1992). Separate visual pathways for perception and action. *Trends in Neuroscience*, 15, 20–25.
- Häggqvist, S., & Wikforss, A. (2018). Natural kinds and natural kind terms: Myth and reality. *British Journal for the Philosophy of Science*, 69, 911–933.
- Hull, D. (1965). The effect of essentialism on taxonomy. *The British Journal for the Philosophy of Science*, 15, 314–326.
- Jablonka, E., & Lamb, M. (1999). *Epigenetic inheritance and evolution*. Oxford: Oxford University Press.
- Kim, D. (2021). Explanation and modality: On why the Swampman is still worrisome to teleosemanticists. *Synthese*, 199, 2817–39.
- Lettvin, J., Maturana, H., McCulloch, W., & Pitts, W. (1959). What the frog's eye tells the frog's brain. *Proceedings of the IRE*, 47, 1940–1951.
- Macdonald, G., & Papineau, D. (2006). Introduction. In G. Macdonald & D. Papineau (Eds.), *Teleosemantics*. Oxford: Oxford University Press.
- Mameli, M. (2004). Nongenetic selection and nongenetic inheritance. *British Journal for the Philosophy of Science*, 55, 35–71.
- Millikan, R. (1984). *Language*. Bradford Books.
- Millikan, R. (1996). On swampkinds. *Mind and Language*, 11, 103–117.
- Millikan, R. (1999). Historical kinds and the special sciences. *Philosophical Studies*, 95, 45–65.
- Millikan, R. (2000). *On clear and confused ideas*. Cambridge: Cambridge University Press.
- Nanay, B. (2014). Teleosemantics without etiology. *Philosophy of Science*, 81, 798–810.
- Neander, K., & Schulte, P. (2021). Teleological theories of mental content. *The Stanford encyclopedia of philosophy*. Spring, 2021 ed.
- Neander, K. (1995). Malfunctioning and misrepresenting. *Philosophical Studies*, 79, 109–141.
- Neander, K. (1996). Swampman meets swampcow. *Mind and Language*, 11, 118–129.
- Papineau, D. (1984). Representation and explanation. *Philosophy of Science*, 51, 550–572.
- Papineau, D. (2001). The status of teleosemantics, or how to stop worrying about swampman. *Australasian Journal of Philosophy*, 79, 279–289.
- Papineau, D. (2016). Teleosemantics. In D. Smith (Ed.), *How biology shapes philosophy* (pp. 95–120). Cambridge: Cambridge University Press.
- Putnam, H. (1973). Meaning and reference. *Journal of Philosophy*, 70, 699–711.
- Ramsey, F. (1927). Facts and propositions. *Aristotelian Society Supplementary*, 7, 153–170.
- Schulte, P. (2020). Why mental content is not like water: Reconsidering the reductive claims of teleosemantics. *Synthese*, 197, 2271–2290.
- Seyfarth, R., Cheney, D., & Marler, P. (1980). Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. *Science*, 210, 801–803.
- Shea, N. (2018). *Representation in cognitive science*. Oxford: Oxford University Press.
- Sober, E. (1980). Evolution, population thinking, and essentialism. *Philosophy of Science*, 47, 350–383.
- Whyte, J. (1990). Success semantics. *Analysis*, 50, 149–157.
- Wilson, R., Barker, M., & Brigandt, I. (2007). When traditional essentialism fails: biological natural kinds. *Philosophical Topics*, 35, 189–215.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.