

5 *Teleosemantics*

DAVID PAPINEAU

The Problem of Representation

“Teleosemantics” is a theory of representation. There are many different kinds of representations. Some representations are mental states: beliefs, perceptions, hopes, fears. Others are public, nonmental items: sentences, maps, diagrams, pictures.

What all representations have in common is “truth conditions.” Any representation will portray the world as being a certain way. It will draw a line in logical space, dividing the possibilities into those that verify it and those that do not. When I assert that “Elvis Presley once visited Paris” or think the corresponding thought, my words, or my mental state, will be true if and only if Elvis did once go to Paris and false otherwise.

(“A picture is worth a thousand words.” It is not always easy to articulate what is being claimed by a perception or by a map or other pictorial means of representation. But this does not mean that these states lack truth conditions, just that they have dense and complex ones.)

Representation can seem puzzling. How *can* one state *stand for* another? When I say or write something, my message will be conveyed by sound waves or marks on paper, and when I believe or perceive something, the vehicle of my ideas will be some arrangement of neurones inside my head. What mysterious force gives these ordinary physical arrangements the power to reach out and lay claim to further possible states of affairs – often far removed in space and time, such as when I represent that Saturn has forty-five moons or that England won the World Cup in 1966?

A natural first thought is that the physical vehicles of representation gain truth conditions in virtue of being *interpreted* as having those truth conditions. The English sentence “Elvis Presley once visited Paris” means what it does, for instance, because speakers of English

understand it a certain way. This is not a bad initial thought, and we shall return to it later, but without further elaboration, it can only get us so far. Interpreting a sentence as meaning a certain truth condition is most naturally understood in terms of its being associated with a certain type of mental state by speakers of the language – in our example, the state of thinking that Elvis once visited Paris. But this then leaves us with the question of what gives those mental states *their* truth conditions. And if the answer is that those mental states are interpreted with the help of yet further mental states, we are clearly off on a regress. We need some account of *original meaning* (“original intentionality,” as it is often called) – we want to explain the kind of meaning that states can have in their own right and not in virtue of being interpreted with the help of other meaningful states.

Many philosophers think that original intentionality is a product of consciousness. As they see it, it is specifically conscious states that have the intrinsic power to represent the world to subjects.¹ They appeal to the way that perceptions and thoughts strike us introspectively. Suppose that you are currently seeing a tree. Isn't it built into the conscious nature of your sensory state, these philosophers urge, that it represents there to be a tree in your environment? Some take a similar line with thoughts, holding that it can be intrinsic to the conscious nature of a thought that it represents, say, that the stock market has fallen.

This line of thought is seductive but fundamentally misguided. Conscious states do represent, of course, but not in virtue of their conscious properties. States with just those conscious properties could, in principle, have represented different things or nothing at all. But this is not the place to argue these points in detail; I have done so elsewhere (Papineau, 2016). In this chapter, instead of tackling consciousness-based accounts of representation head-on, I shall instead explore an alternative approach to representation as depending on nonconscious properties and relations. This approach will explain the representational powers of conscious states in term of such properties and relations, rather than their conscious nature, and moreover will also allow that nonconscious states also can be representational in the fullest sense.

¹ This is, of course, the dominant tradition in the history of philosophy. For contemporary defenses of this approach, see the Introduction and essays in Kriegel (2013).

One last preliminary point before proceeding: not all representations are *categorical*, in the sense of being offered or entertained as saying what *is* the case, as opposed to what *might be* the case. For example, conjectures, imaginations, hopes, and fears are representations all right, but they are not categorical. They have truth conditions and so can turn out to be true or false, like other representations, but they aren't embraced as categorically conveying how things are, in the way that assertions, beliefs, and perceptions are. They are merely possibilities to be considered. In what follows, we shall be concerned specifically with categorical representations. Once these have been explained, then perhaps an account of noncategorical representation can be built on that basis.

Representation as a Biological Category

The key to understanding representation is to view it as a biological phenomenon. According to the teleosemantic program, representations are states whose biological function is to guide behavior in ways appropriate to such-and-such conditions. Those conditions are then the truth-conditional contents of the representational states. The representations are true if those conditions obtain but not otherwise.²

Let me illustrate. Vervet monkeys in Kenya have three distinct alarm calls – for leopards, eagles, and snakes respectively. These calls are designed to prompt specific behaviors in the monkeys. As Seyfarth, Cheney, and Marler explain in their classic 1980 paper, the monkeys “respond to leopard alarms by running into trees, to eagle alarms by looking up, and to snake alarms by looking down.” These responses determine what the calls represent, in that the truth condition of each call is that circumstance in which the monkeys' consequent behavior would be appropriate to its survival.

It will be helpful to analyze this case in terms made familiar by Ruth Millikan (1984). Distinguish the “producer” of the call, the signaler, from the “consumer,” that is, the monkey that responds to the call. According to the teleosemantic analysis, it is the behavior of the consumers that determines the truth-conditional content of

² The first works developing this teleosemantic idea include Millikan (1984), Fodor (1984), Papineau (1984, 1987), and Dretske (1986, 1988).

the call, not the circumstances that prompt the producers. Suppose that the consumers respond to some call with behavior that is appropriate to an impending eagle, say. This then shows that the call means “eagle.” And this remains the case even if the producers also regularly produce that call in response to fast-moving clouds, low-flying airplanes, and so on. The truth condition of the signal depends on how the consumers behave in response to it, not on what causes producers to emit it.

In this example, the producer is one organism, and the consumer another. But the story will work just the same if the producer and the consumer are inside the same individual. This then gives us a model for representation by mental states as well as public signals. Suppose, as seems plausible, that the monkeys also have three kinds of cerebral states, “produced” by their visual systems and “consumed” by their motor control systems. Then the story runs just the same. These brain states will represent leopards, eagles, and snakes respectively, in virtue of the fact that they are designed to lead the monkeys to behave in ways appropriate to just those threats.

So here we have a simple explanation of representation. It isn’t magic. It is just a matter of certain states having the biological function of instigating behavior that is appropriate to such-and-such conditions.

There is a sense in which this account preserves the intuitive idea that the meaning of a representational state depends on how it is *interpreted*. Representation arises whenever a consumer *interprets* some state as signifying some circumstance, in the sense that it *acts* in a way appropriate to that circumstance. The crucial point is that the idea of interpretation now in play is not the idea of the representation prompting some further *mental* state in the agent. As we saw earlier, if we understand interpretation in this mental way, the appeal to interpretation is no good for explaining original intentionality because it inevitably degenerates into regress. However, the current proposal cuts through this regress. Interpretation is now a matter of acting, not thinking further thoughts. A representation is interpreted as having a certain truth if it leads a consumer to *act in a way appropriate to that condition*, not to *form a thought with that truth condition*. By this means, we explain representation without presupposing it.

Generalizing the Story

It will be useful to schematize more fully how the teleosemantic account explains representation in terms of biological functions. Suppose that we have some consumer system that responds to representation R with some behavior B and that the biological purpose of this consumer is to achieve some end E . Then the system that produces R will have the function of producing R when condition C obtains, where C is that condition that will ensure that B causes E . If all this is in place, then R will *represent* C : the producer is biologically supposed to produce R when C because this will then enable the consumer to serve its function of achieving E .

In the vervet monkey example, we took the relevant end E simply to be survival and reproduction. But we can also view biological ends in a more fine-grained way, as aimed at more specific results than survival and reproduction, and this will then allow the teleosemantic approach to deal appropriately with more complex kinds of representation.

To understand how biological items can have biological functions that are more specific than survival and reproduction, note that biological systems can be decomposed into nested structures of interlocking components. For example, the human body is composed of the brain, the temperature-regulating system, the cardiovascular system, and so on. The cardiovascular system can itself be decomposed into the heart, the lungs, and the blood vessels. Now all these components have the eventual biological function of fostering survival and reproduction. But they are all supposed to contribute to this in special ways: the brain by managing behavior and hormonal levels, the temperature-regulating system by maintaining a constant temperature, the cardiovascular system by circulating oxygen and nutrients and removing carbon dioxide and toxins, and, in turn, the heart, lungs, and blood vessels are supposed to do the latter by pumping, oxygenating, and transporting the blood, respectively.

Given this, we can identify the functions that are *specific* to the components in an overall biological system. For example, the specific function of the heart is to pump the blood. Of course, the heart also has the further functions of circulating oxygen and nutrients and removing carbon dioxide and toxins and thereby of fostering survival and reproduction. But these further functions are not specific to the heart, as is shown by the fact that their nonfulfillment does not mean that the

heart isn't doing *its* job: if oxygen isn't circulated, this might be because the lungs aren't oxygenating the blood, not because the heart isn't pumping. In general, we can say that the specific function of some biological component is the most immediate effect it is supposed to produce at that level of decompositional analysis where it itself appears as an unanalyzed component. So, for example, pumping is the immediate effect attributed to heart once we decompose the cardiovascular system into its components.³

This means that the representational functions attributed to contentful states need not always be geared to survival and reproduction as such. If the producer-consumer system in which some representation *R* is a component itself has some specific end *E*, then the representational content of *R* will be the condition *C* that ensures resulting behavior will achieve *E*, whether or not survival or reproduction follows.

One obvious application of this idea will be to organisms that form and activate *desires* and other motivational states. These motivational states can be viewed as themselves consumer mechanisms whose specific biological purpose is to achieve certain proximal results, such as water, or sex, or social esteem – or whatever the motivation is aimed at. These motivational states are not aimed at such general results as survival and reproduction because it is not their fault, so to speak, if these further results do not follow once their specific ends have been achieved. The specific job of my desire for water, for example, is to get water into my body, and it will have fulfilled this aim even if my stomach is malfunctioning and this does not help my survival and reproduction.

Given this, representations that direct the selection of actions in pursuit of desires and other motivations will then represent circumstances relevant to achieving the specific ends of those motivations. The representational state that tells me what to drink when I am thirsty has the specific function of tracking water and fulfills this function even in cases where water will not aid my survival or reproduction.

Another important range of cases will be where the purpose of the relevant consumer is itself to produce further representations. For example, many producer mechanisms within our perceptual system have the purpose of detecting “features” (e.g. edges of physical objects)

³ This account of specific functions is borrowed from Neander (1995).

on the basis of which further consumer mechanisms will construct representations of more complex phenomena (say, whole three-dimensional objects). In this kind of case, the producer (the edge detector) will represent *edges*, say, in virtue of the fact that its outputs are treated by the consumer (the object representer) in ways appropriate specifically to the presence of edges. (And this consumer will then have the purpose of representing *objects* in virtue of the fact that *its* outputs are in turn consumed by further mechanisms in ways appropriate, given their specific purposes, to the presence of objects.)

These examples show that the general teleosemantic approach will discern different kinds of representations in different components of different organisms depending on the details of their internal cognitive architectures. What the states in an organism's component mechanisms represent will depend on such things as its structure of motivational states, the computational structure underpinning its visual perception, and so on. This is not a weakness of teleosemantics. On the contrary, it shows that it is a powerful framework that can be applied to a wide range of cognitive architectures to identify the specific representationalist purposes served by their components.

Is Truth Functional?

The teleosemantic approach to representation hinges on the idea that the *truth* of a representation coincides with *fulfilling its biological function*. This opens the teleosemantic approach to a common objection – namely, that truth and biological function can come apart for representations. In particular, so the objection goes, there are plenty of cases where representations fulfill their biological functions even though they are false. Evolution doesn't care about truth, but just about practical biological success, object the critics. If we want to understand truth, they conclude, we need to look beyond the biological realm with its exclusive focus on practical results (Plantinga 1993; Burge 2010).

There are three different kinds of cases worth discussing here. First, there are representations for which the *biological expectation is falsity* rather than truth. Second, there are representations that luckily happen to lead to biological success *despite being false*. And third, there are representations that systematically confer some positive biological benefit *in virtue of being false*.

Let us take these in turn. For an example where falsity is the biological norm, the vervet monkeys will serve well enough. Let us suppose that the monkeys are designed to err on the side of caution and will alert the troop to an eagle threat on the slightest pretext, with the result that the vast majority of “eagle” calls are occasioned by clouds, airplanes, and so on rather than eagles. (I make no claims for the ethological accuracy of this supposition.)

We can see why the monkeys might have been set up in this way. The cost of a false-positive call – one prompted by a fast-moving cloud – is far less than that of a false-negative call – ignoring a real eagle. The former mistake only means a wasted upward glance, but the latter could well mean death. It is far better, in biological terms, to bear the cost of regular false alarms than to run the risk of being caught by an undetected eagle.

But it is simply a mistake to think that cases such as these are a problem for teleosemantics. Despite the frequency of false positives, falsity is no part of the biological *function* of the “eagle” signal, nor of the mechanisms that produce them. Biological functions are always advantageous effects, results that contribute to survival and reproductive success. There are no such advantageous effects occasioned by false alarms. Pausing and looking upward because of a cloud is pure wasted effort. The advantageous effects of the “eagle” signal accrue specifically in those cases where there really is an eagle around, and the signal enables the monkeys to avoid capture. *That* is the function of the signal, and it is fulfilled specifically in cases where the signal is true, just as the teleosemantic approach would have it.

The monkey example is just a special case of the point that the biological function of some trait need not be normally or even often achieved – provided the payoff when it does occur is big enough to outweigh the cost of failed attempts. Male sperm are the standard illustration. Nearly all sperm are fated to wither and die before fertilizing an egg. But that doesn’t mean that fertilization is not the function of the sperm. We don’t want to say that perishing without achieving fertilization is the *function* of sperm just because that fate is biologically overwhelmingly probable.

Let me now turn to the second kind of example. Sometimes the behavior prompted by a false representation can lead to biological success by luck. A thirsty monkey sets off in a certain direction, prompted by the belief that there is water in the stream. As it happens,

the belief is false (the stream has dried up), but happily, the monkey comes across a pool of water halfway there. At first pass, this looks like a case of a belief serving its biological function even though it is false. After all, the belief is here consumed by the monkey's thirst-quenching mechanism, and it leads successfully to the satisfaction of that mechanism's biological end.

But this is not a problem for teleosemantics either. While the thirst-quenching mechanism has achieved its end, the belief itself has not served its own biological function. Its specific function is to coordinate behavior with the putative presence of water in the stream. It hasn't done that in this case because there was no water in the stream. It was just a matter of luck that water was found; it wasn't because the belief was working as it was supposed to.

In general, a biological trait can lead to some eventual biological success by luck in a particular case without serving the specific function for which it was selected. On some occasion the camouflage of an insect saves it from predation by delighting a child whose laughter scares away a bird. But this clearly wouldn't be a case of the camouflage serving its specific function of *hiding* the insect. Similarly, the fact that a belief can luckily engender success though false is no counterexample to the teleosemantic claim that its specific function is to coordinate its consumer's behavior with its truth condition.

Third, there are arguably some few cases where representations do genuinely serve a biological function in virtue of being false. Consider the phenomenon of "depressive realism": most psychologically healthy people have an inflated view of their own social standing, by comparison with objective measures; the only people with accurate beliefs about their status tend to be depressed. Let us suppose that these mistaken beliefs among nondepressed people have a biological purpose: the function of the widespread false beliefs is to stop people from retreating into their shells and to encourage them to be enterprising. (Again, I make no claims for the biological accuracy of this supposition.)

Now cases like these really do involve beliefs that serve a biological function *because* they are false. It is specifically when lower-status people think that they are higher status that they are encouraged to be enterprising. And this certainly seems in tension with the teleosemantic idea that we can equate the truth-conditional content of a belief with that circumstance in which it serves its specific biological function.

Here the truth condition is that you are higher status, but the function is served when you are lower status.

To deal with this issue, teleosemantics needs to recognize that some representations can serve two different functions. This is a familiar enough biological idea. For example, large earlobes can facilitate both audition and thermal regulation and be selected for both these positive effects. Similarly, in our case, the belief arguably has the functions of both (1) guiding behavior aimed at satisfying currently active motivational states and (2) boosting enterprise by fostering self-esteem.

Once we recognize these two distinct functions, we can see that there is nothing here to undermine the teleosemantic approach to representation. The function that matters to the teleosemantics of belief is the former one: guiding behavior in a way that will serve whichever motivations are currently active. To fulfill this function, your beliefs still need to be true. (Suppose that you want to make money and enter a popularity contest in the belief that your high status will win you the \$100 prize. You won't get what you want if your belief isn't true.) That some specific beliefs, such as this belief about your status, might also have some further function, such as bolstering self-esteem, that is fulfilled when the belief is false does not eliminate the former teleosemantic function any more than earlobes acquiring a thermoregulatory function eliminates their auditory function.⁴

Determinacy of Content

A standard objection to teleosemantics is that it is not able to explain the possession of fully determinate contents by representational states. Jerry Fodor (1990) has argued this point in connection with the states in frogs' brains that prompt frogs to snap their tongues in the direction of passing flying insects. Fodor challenges teleosemanticists to explain why these states should be regarded as representing flying insects rather than small, black, moving things. After all, Fodor maintains, we could as well take the frog's visual system to be biologically designed to respond to small, black, moving things as to respond to flying insects. A healthy frog will snap its tongue whenever it is presented with a small, black, moving thing, whether or not it is a flying insect.

⁴ For further discussion of this last kind of case, see Papineau (1993, chap. 3).

An initial answer to Fodor is to respond that teleosemantics focuses on conditions that ensure biological success, and the alternative conditions he has in mind fail to satisfy this requirement. It is true that a healthy frog will respond to small, black, moving things even if they are not flying insects. But the conditions that teleosemantics is concerned with are not those that can be expected to cause the frog's state, but rather those that will ensure that resulting behavior causes success. The frog's state clearly has the function of helping the frog catch flying insects rather than small, black things: no selective advantage accrues to a frog that grabs some nonnutritious speck of passing dirt. And so, in line with this, defenders of teleosemantics can argue that the frog's brain state represents a flying insect in a given direction rather than small, black things. For it is precisely when its brain state is prompted by a flying insect, rather than any small black thing, that an advantageous effect will accrue (cf. Millikan 1993).

However, it is not clear that this response fully deals with Fodor's worry. I just said that the advantageous result of the frog's state is flying insects rather than small, black things. No reproductive advantage accrues when it catches a small, black thing that isn't an insect. But why stop there? The biological point of catching flying insects is to get them into the stomach. No reproductive advantage accrues if an insect is caught but isn't ingested. Again, the biological point of ingesting something into the stomach is to get nutrients into the bloodstream. No reproductive advantage accrues if an insect is ingested but yields no nutrients into the bloodstream. And so on. In the end, the ultimate point of all functional traits is survival and reproduction. No reproductive advantage ensues from any intermediate effects if they don't eventuate in survival and reproduction.

Given all this, it might not be obvious why we should interpret the frog's state as representing flying insects. Why not read it as representing stomach filler? Or as nutrient source? Or even as reproduction enhancer?

To resolve this issue, we need to go back to the idea of biological traits having specific functions of their own, in addition to those they share with other components of the whole organism. In the example used earlier, the specific function of the heart is to pump blood. It also has the functions of circulating oxygen and eventually of fostering survival and reproduction – but these functions it shares with other

organs, such as the lungs, and so they are not specific to the heart as such.

We saw earlier how this concept of specific functions matters when we apply teleosemantics to organisms that have motivational states. It allows us to think of states like desires as themselves biological components with specific functions, namely, the production of the specific effects that will satisfy them. And then we can equate the truth conditions of representational states that inform the pursuit of desires with conditions that will ensure that those desires are satisfied.

So, if we could credit frogs with motivational states, then this would resolve our issue. If the frog's behavior is motivated by the desire for a flying insect as such, then the state that prompts tongue snapping in a certain direction would signify that there is a flying insect in that direction. Whereas if the frog's is motivated by a desire for a nutrient source, say, then the state would represent a nutrient source in that direction.

However, there seem to be no good grounds for attributing such motivational states to frogs. Modern physiologic research suggests that frogs lack any integrated decision-making system in which belief-like states serve desires. Rather, each of its behavioral systems is guided by its own proprietary information, which is unavailable to its other systems of behavioral control. One channel of sensory information guides its prey-catching behavior, another guides its obstacle-avoiding behavior, and yet another its ability to jump away from looming threats. Lesions of the frog's optical system can dissociate these different abilities (Milner and Goodale 1995, sect. 1.2.2).

Still, this lack of an integrated decision-making system does not mean that the idea of specific functions has no grip on the frogs at all. There is nothing to stop us from applying this idea directly to the prey-catching system as such. As in other cases, this system has a sequence of functions. It is designed to catch flying insects, thereby to have them swallowed, thereby to allow digestion to place nutrients in the bloodstream, and thereby . . . to lead to reproductive success. But only the first of these is arguably the function peculiar to the prey-catching systems, considered as the visuomotor system that governs head turning and tongue snapping. It is not necessarily the fault of this system if a flying insect is caught but does not end up in the stomach (because the swallowing mechanism is not working) or if it is placed in the

stomach but not digested (because the stomach is malfunctioning), or so on.

If we accept, on these grounds, that the prey-catching system has the peculiar function of catching flying insects, rather than any later effects, then we can view the sensory signals that prompt behavior in this system as indicating the circumstance under which that behavior will achieve the system's peculiar end – that is, as indicating the presence of a flying insect in such-and-such a direction.

There might seem a further issue here. I have been assuming that the relevant signal is part of the prey-catching system. And this by no means seems mandatory. After all, why not regard the frog's sensory signal as part of the prey-stomaching system or as part of the prey-digesting system or so on? The effects occasioned by the sensory signal don't normally stop with the flying insect being caught – when everything is working as it should, the flying insect will also promptly be swallowed, and digested, and . . .

This then threatens to render the content of the sensory signal indeterminate once more. Each of the systems at issue – the prey-catching system, the prey-stomaching system, the prey-digesting system, and so on – will have a specific function of its own. But this won't give the sensory signal a determinate content if it is not determinate which of these larger systems is informed by the signal. It will leave it open that the signal tells the prey-catching system about insects, so to speak, but also the prey-stomaching system about stomach fillers, the digestive system about nutrient sources, and so on.

However, this point has in effect already been dealt with. The signal in question is properly seen as a component of the prey-catching system, not of the prey-stomaching or the prey-digesting system. When we first analyze the larger prey-stomaching system, say, into its component prey-catching and prey-swallowing systems, there is as yet no need to bring in the signal as such. The larger prey-stomaching system fulfills its specific function as long as its component prey-catching and prey-swallowing systems fulfill theirs, however they manage to do that. It is only when we analyze the prey-catching system itself that the signal comes into view, so to speak. The prey-catching system fulfills its specific function when its components fulfill theirs – which requires, *inter alia*, that the signal tracks the condition in which the resulting snapping behavior will secure a flying insect. The signal is thus

specifically a component in the prey-catching system and its truth condition – the presence of a flying insect in a certain direction.⁵

Outputs over Inputs

According to teleosemantics, the truth condition of a representation depends on the output of the representation, on what behavior it prompts, and not on the input to it, on what circumstances cause it. The vervet monkey's state means "eagle" because it prompts the monkeys to behave in ways appropriate to eagles, even if most of the things that cause the state are not eagles.

Teleosemantics contrasts in this respect with causal theories of representational content, that is, theories that aim to explain truth conditions in terms of their characteristic causes. The obvious problem facing such theories is to distinguish causes that constitute truth conditions from other things that cause the representation. This problem is often termed "disjunctivism": what makes "eagle" the truth condition of the monkey's state rather than the disjunctive condition "eagle-or-low-flying cloud-or-airplane-or-anything else that causes the state"?

This problem is dealt with from the start by teleosemantics precisely because it understands truth conditions in terms of outputs rather than inputs. It doesn't start by looking at the causes of representational states and then seek somehow to narrow these down to the truth-conditional causes.⁶ Rather, it simply asks what will ensure that the behavior resulting from the state will be successful.⁷

⁵ In an earlier paper (Papineau 2003) I argued that the frog's state was indeed indeterminate, on the grounds that it could equally well be considered a component in all the nested prey-catching, prey-stomaching, and prey-digesting systems. It was only after writing that paper that I came to appreciate how Neander's notion of specific functions (1995) resolves this issue in favor of the first option. (Perhaps it is worth mentioning here that in her 1995 paper, Neander herself ends up arguing that the frog's state represents "small, black thing." I would say that she is driven to this conclusion by her mistaken general assumption that traits in healthy animals cannot malfunction simply because the environment is unhelpful, which then implies that a healthy frog cannot be misrepresenting simply because a black speck of mud shoots by.)

⁶ Perhaps the best-known causal theory is Fodor's asymmetrical dependence theory (Fodor 1987).

⁷ Sometimes teleosemantics is understood as equating truth conditions with circumstances that are biologically supposed to cause the relevant state in "epistemically ideal conditions" and then criticized because it has no noncircular way of defining "epistemically ideal." But this criticism presupposes that teleosemantics is

Not all commentators view this output orientation of teleosemantics as an advantage. If the monkey's representation is triggered as readily by clouds as by eagles, would it not be better to include clouds in its truth condition?

This reaction is bolstered by the following well-known thought experiment due to Paul Pietroski (1992). The kimu are simple creatures whose only enemies are the snorf. The snorf hunt the kimu at dawn. Then one day a biological mutation endows one of the kimu with an ability to register the presence of red things and an inclination then to approach them. This is an advantage to its possessors because it leads them to climb hills at dawn, the better to observe the red sunrise, with the result that they avoid the snorf, which are ill suited to climbing hills. As a result, the disposition spreads through the kimu population.

Now consider the state a kimu gets into when it is stimulated by something red. It is natural to credit this state with the content "red." But an output-based teleosemantics sees things differently. In general, nothing good happens to kimu when they approach red things. Most red-approaching behavior is a waste of time. It is only when it takes them away from the snorf that it yields a biological advantage. So an output-based teleosemantics will construe the state in question as representing "snorf-free" or "predator-free" or something like that. Pietroski argues that this is highly counterintuitive. After all, by hypothesis, the kimu's senses are tracking redness, not snorf.

But this argument is by no means conclusive. Defenders of teleosemantics can object that Pietroski's intuitions are reading more into the story than is warranted. As Pietroski initially tells it, the kimu evolve some state that is triggered by redness and that functions to keep them away from the snorf. But his subsequent discussion invites us to suppose that the kimu have some general-purpose visual *system* whose outputs might inform an open-ended range of behaviors directed at various possible ends (such as avoiding blood, or finding apples, or indeed wanting to see red things). However, this supposition adds significant extra structure to Pietroski's initial story and so makes room for teleosemanticists to argue that an organism with that extra structure would indeed be representing redness rather than snorf-freeness: if the kimu's visual system did inform a *range* of different

in the business of distinguishing good causes from bad ones, when in truth it doesn't care about causes but only conditions for success.

behaviors directed at different ends, then the content of its visual states would be conditions that ensured in the achievement of all those ends, and one such state might well come out as representing redness. By contrast, if we stick to a minimal understanding of the snorf, in line with Pietroski's initial story, as having only a special-purpose visual sensitivity that brings no advantage except snorf avoidance, then it's not so clear that there is anything wrong with reading their states as representing "snorf-freeness": after all, if these states never do anything except trigger simple avoidance behavior, it seems natural enough to read them as representing the danger they are designed to avoid.

Doing Without History

Teleosemantics is not the only theory of representation that explains content in terms of outputs rather than inputs. "Success semantics," the origins of which can be found in Ramsey (1927), focuses specifically on belief-desire systems and in that context agrees with teleosemantics in equating the truth conditions of beliefs with circumstances in which resulting actions will satisfy desires. And, more generally, various species of convention-based signaling theory agree with the structure of producer-consumer teleosemantics in equating the truth conditions of a signal with those circumstances in which the behavior performed by the recipient of the signal will satisfy the recipient's ends (Lewis 1969; Skyrms 1996, 2010).

Where these theories differ from teleosemantics is in not viewing these structures as necessarily involving *biological functions*. For teleosemantics, the satisfaction conditions of desires, and more generally the ends of consumers of representations, are equated with the effects that these systems are *biologically* supposed to produce. And correspondingly, the producers of representations are taken to have the *biological* function of producing representations in circumstances where resulting behavior will fulfill the functions of their consumers.

Success semantics and signaling theory avoid these biological commitments. They see no reason to bring biology into the understanding of representation. In their view, we can understand what it is for a desire to be aimed at some outcome or, more generally, what it is for consumers to have ends, independently of any appeal to biological function: these are perfectly good everyday notions, and it is not clear

that they demand any further analysis. And even if they do, they can arguably be understood in terms of other nonbiological everyday notions, such as contributing to psychological or bodily equilibrium.

True, understanding representation in teleosemantic terms will automatically carry with it an explanation of *why* the world contains the representational systems it does. Teleosemanticists work with the standard etiological understanding of biological function: a trait *T* has the function *F* if it was designed by natural selection to produce *F*, so to speak – or, less metaphorically, if *T* is now present because ancestral versions of *T* were selected because they produced effect *F* (Wright 1973; Millikan 1989; Neander 1991). On this etiological account of function, to ascribe a function *F* to a trait *T* will therewith explain the presence of *T* in terms of its selectional past. And in the case where we are dealing with representational systems, as with teleosemantics, an ascription of biological functions will carry with it an evolutionary explanation of those systems.

But those who favor nonbiological alternatives to teleosemantics can retort that it is one thing to explain the existence of the representational systems and another to invoke such systems in giving explanations of action. We don't need an evolutionary (or any other) explanation of why representation exists, just because we invoke representation in explaining further things.

As a preliminary to addressing this challenge, it will be useful to clarify exactly how representational notions do help us to predict and explain further things. A first thought might be that they help us understand how internal cerebral states such as perceptions, beliefs, and motivations interact in generating bodily movements. However, this kind of “narrow” psychological explanation makes no real use of representational notions that relate internal cerebral states to features of the environment. After all, if our focus of explanatory interest were solely in predicting and explaining bodily movements, we wouldn't need to think of cerebral states as related to things outside the head at all – we could just think of them as internal components in a structure of causal pushes and pulls (Papineau 1993, chap. 3).

The real significance of representational notions is that they allow us to predict and explain *success*, that is, the achievement of distal results. My belief that there are lobsters in that bay leads me to place my traps there – and then, if that belief is true, it further leads to my catching lobsters. It is the explanation of this eventual result for which

representation is crucial – hinging, as it does, on the way my belief is supposed to track the whereabouts of lobsters.

So the crucial pattern is this: behavior *B* in pursuit of end *E* is informed by representation *R* with truth condition *C*, and when *C* obtains (when *R* is true), not only is *B* performed, but in addition, *E* is achieved. In short, truth explains success. By ascribing representational contents, we are thus able to discern systematic patterns governing the achievement of distal ends.

Still, nothing in this, the opponents of teleosemantics can continue to object, depends on biological functionality. Teleosemanticists may pick out the relevant ends *E* as ones that fulfill biological purposes. But why does that matter? The representation-invoking explanatory pattern requires only that there be some *C* that systematically ensures that the behavior *B* prompted by *R* yields some *E*. And if this is so for the *E*'s and *C*'s identified by the teleosemanticists, it will remain so whether or not those *E*'s fulfill biological functions. What matters are the present-day patterns relating the interlocking parts of representational systems, not the evolutionary history of those systems.

It is worth distinguishing two different ways of running this anti-evolutionary line. So far I have been assuming that the output-based alternatives will reject the teleosemantic appeal to biological functions altogether. But an alternative is to view representation as a matter of biological functions all right, but to understand function in some nonetiological way. There are various nonetiological approaches to biological functions available, united by the general thought that the biological function of a trait involves a contribution to *current* or *future* survival, reproduction, or other beneficial effect rather than to the *past* effects that *causally explain* the presence of the trait (Cummins 1975; Boorse 1976; Bigelow and Pargetter 1987). These nonetiological “forward-looking” accounts of function mean that there is room for output-based theories that agree with teleosemantics that representation is a matter of guiding behavior in a biologically functional way but which identify the relevant biological functions nonetiotologically.⁸

⁸ Some writers extend the term “teleosemantics” to theories that in this way explain representation in terms of nonetiological “biological functions” (Abrams 2005; Nanay 2014). There is nothing wrong with this, but for present purposes, it will be convenient to continue to restrict “teleosemantics” to etiology-based theories.

One issue here is the right way to understand claims about biological functions. This is a much-debated matter. One relatively uncontroversial (if insufficiently remarked) point is that only etiological functions are suitable for *explaining* the presence of the traits that have them and so for taking such claims as *the function of hearts is to pump blood* at explanatorily successful, as intuitively they certainly are. Still, this does not mean that claims about biological functions might not also sometimes be properly understood in a nonetiological forward-looking way. Such forward-looking functional claims would not *explain* the presence of the items with functions, but they might for all that provide a useful way of categorizing some of their biological effects.

Fortunately, in the present context, we can bypass this issue of the proper understanding of claims about biological functions. Perhaps there are important philosophical issues that hang on which notion of *biological function* is most current in scientific and other contexts (though I rather doubt it). But in any case, and however that debate comes out, there is nothing to stop someone from *defining* a notion of biological function in some nonetiological way as involving contributions to current or future beneficial effects, or some such, and then maintaining that *representation* is best understood in terms of a contribution to biological ends so-defined – perhaps backing up this claim by arguing that this understanding best captures the present-day patterns we appeal to when we invoke representation to predict and explain success in achieving distal results.

The issue of biological functions thus turns out to be something of a red herring in the context of output-based theories of representation. Teleosemanticists want to explain representation in terms of etiological functions, effects that mattered for the selectional history of representational systems. Alternative output-based approaches deny that selectional histories matter, as opposed to identification of ends the achievement of which can systematically be tracked representationally. Whether these alternative approaches think of the relevant ends as biological functions is a subsidiary issue. Either way, we have the same challenge to teleosemantics: how can it matter to present-day predictive and explanatory patterns what roles the relevant ends played in the past?

Swampman

This challenge is highlighted by the well-known “swampman” thought experiment (Davidson 1987). Suppose that lightning strikes a steamy swamp in the tropical jungle and by miraculous coincidence a perfect molecule-for-molecule replica of a human being assembles itself from organic materials in the swamp. By hypothesis, this “swampman” will lack a history of natural selection and so, according to teleosemantics, will be incapable of representing anything. Yet, intuitively, it seems that swampman will be capable of at least some forms of mental representation. After all, it will be physically just like a normal human, so it will be able to visually register its surroundings and make appropriate behavioral responses. Given this, there would seem every reason to credit its states with truth conditions and use this to track when it will succeed in achieving its ends. So it looks as if teleosemantics has gone wrong somewhere if it denies that swampman has any representational capacities.

The standard teleosemantic response to this difficulty is to bite the bullet and conclude that swampman will indeed be incapable of representation. Maybe everyday intuition argues that swampman can represent, allow the bullet-biters, but a good theoretical account should be allowed to overturn a few everyday intuitions. Just as our modern concept of fish excludes whales, despite any naive intuitions to the contrary, so should a developed concept of representation exclude swampman. According to this line of thought, then, we should replace our naive concept of representation by the theoretically more powerful selection-based notion, even at the cost of overturning intuitions about swampman (cf. Millikan 1996; Neander 1996; Papineau 1996).

However, there is room for an alternative and more nuanced defense of teleosemantics against swampman worries. Rather than seeking to replace the everyday *concept* of representation with one that excludes swampman as a representer, teleosemanticists can leave that concept as it is and instead appeal to the status of teleosemantics as an *a posteriori reduction* of representational facts. On this natural way of understanding of teleosemantics, it is not offered as an analysis of our everyday concept of representation (after all, it was always implausible that this everyday concept should refer to selectional histories) but rather as a theoretical reduction, which appeals to scientific theory to uncover

the important underlying features that bind different instances of representation together, just as scientific theory uncovers the underlying nature of water and other familiar chemical substances.

From this perspective, it is no argument against teleosemantics that a representationally competent swampman is consistent with our everyday concept of representation. You might as well argue against modern chemistry on the grounds that XYZ-composed water is consistent with our everyday concept of water. The fact that a swampman with representations can be *imagined* does nothing to undermine the central teleosemantic claim that in the *actual* world representational facts are *constituted* by selectional facts.

Of course, if a swampman were to exist in the actual world, then things would be different. Such a being certainly would display an important explanatory pattern covering the representationally guided achievement of its ends, in line with the everyday notion of representation. However, if a swampman were to exist, then teleosemantics would simply be false. Selectional histories wouldn't be an *a posteriori* theoretically important part of what representing agents have in common because it wouldn't be something they had in common at all. Still, none of this is relevant to how things are in the actual world, where swampman does not exist, and all representers do turn out *a posteriori* to share a selectional past. Given this, teleosemanticists can insist that an imaginary swampman is no more relevant to teleosemantics than imaginary molecular makeups are to chemistry (Papineau 2001).

Still, does this teleosemantic answer to the swampman challenge really scratch the itch occasioned by nonetiological output-based theories of representation? Maybe the general run of representers in the actual world do have selectional histories. But why think of this as the "important underlying feature" that *constitutes* representation rather than as an incidental past circumstance that happens explain why the actual world contains representational systems? After all, isn't the important thing that representational systems display present-day patterns of behavioral success, not what caused them to be like that? As long as they do display such patterns, they can fruitfully be viewed as representational, independently of any selectional history they may have. We need only think of swampman once more to see the point.

In response to this continued challenge, teleosemanticists can first make the initial point that it is no accident that all representational systems found in the actual world have a selectional history. Representational systems, even the simplest, are complex structures involving information-gathering producers and flexibly behaving consumers that are well suited to the needs of the organisms that possess them. Along with all other instances of apparent design in the natural world, the existence of such structures demands explanation. It beggars belief that such helpful complex mechanisms could in reality ever arise by chance, swampman-style. The only serious possibility is that they are results of past selection processes that preserved and refined structures that were biologically helpful to their possessors.

Teleosemanticists then can follow this with the point that anybody who wants to *know* about representational systems will inevitably also be interested in their selectional histories. This is a heuristic rather than a principled point. There is no absolute barrier to an investigator fully understanding some representational system directly, figuring out the operations of its various interlocking parts entirely from first-hand observation. But this is not how it works in practice. As with all cases of “reverse engineering” – figuring out the inner workings of some mechanism from observing its operations – it is practically essential to consider how the system was designed, to think about the way in which its parts are *supposed* to work together. Without considering the designed purposes of the parts, it wouldn’t be feasible to distinguish important effects from incidental features. And these points apply to systems designed by natural selection just as much as they do to humanly designed systems. Throughout biology, investigators appeal to possible evolutionary histories to help them understand the current workings of biological systems.

In the case of representational systems, this point applies particularly to the causal transitions between representational states made within the system – the “syntactic” moves whereby some internal states lead to others and eventually to the selection of behavior. It is normally assumed without thinking that these syntactic moves will tend to respect the semantic values of the relevant states – that is, that they will be made in a way that conduces to true states generally leading to further true states and thence to the selection of behavior appropriate to goals. Without this assumption, it would be impossible to draw the information-processing flowcharts that are crucial to many theories in

cognitive science and neuroscience. But this “syntax respects semantics” assumption is itself a design assumption. It rests on the idea that representational systems have been set up so as to ensure that the selection of behavior will be appropriately geared to circumstances.

The opponents of teleosemantics might wish to object that none of this really addresses their central point, which is that the current workings of representational systems are metaphysically independent of their causal histories, and that it is only the former that matter for predicting and explaining current behavioral success. However, by this stage, this point is wearing thin. If all actual representational systems have a design history – and anybody who is interested in understanding a representational system must also be interested in its design history – then why continue to insist that real representation only involves present-day patterns?

Consider watches. It is natural to think of watches etiologically. Watches are portable items that make the time of day visible and, what is more, have been designed for this purpose – not by natural selection, of course, but by the conscious watchmakers who constructed their complex workings specifically to ensure that they display the correct time.

However, one can imagine an antietiologist about watches. “What matters is only that watches correlate well with the time. It is these present-day patterns that we use to tell the time. True, all normal watches have probably been produced by conscious designers for this purpose. But that is a different matter. It might explain the existence of watches but is incidental to their predictive and explanatory significance.”

Well, one could insist in this way that a design etiology is not a requirement for being a watch. But what would be the point? After all, every actual portable item that displays the time has been designed for that purpose, and moreover, if you want to understand the workings of a watch, you have no option but to try to figure out what the designer intended the components to do. Given this, nothing would seem to be gained by dropping the etiological requirement for being a watch apart from unnaturalness and an unnecessary purification of categories.

The analogy should be clear. We could, in principle, opt to think of representation as a nonetiological kind. But nothing would be gained. The same range of instances would still come out as representational,

and we would simply have cleansed representation of something that is practically essential to understanding its workings.

Varieties of Selection

Let me conclude by addressing an issue that might have been worrying some readers for a while. As the last two sections will have made clear, teleosemantics is committed to regarding all original representation as deriving its functionality from past histories of natural selection.⁹ But is this at all plausible? The most familiar kind of natural selection is the intergenerational selection of genes. However, it is highly unlikely that all original representations can be explained in terms of such *genetic* selection. For a start, most human mental representations are products of ontogeny rather than phylogeny. No genes have been selected specifically to foster my desire for a new iPhone or my belief that the Mets will win the World Series.

Fortunately for the teleosemantic project, the possession of etiological functions by biological traits does not always depend on the selection of genes that give rise to those traits. I shall distinguish three ways in we can have etiological functionality in nongenetic traits. The first, emphasized by Ruth Millikan, appeals to a many layered account of functions. The second involves nongenetic selection in learning. The third depends on the intergenerational inheritance of nongenetic items. Together these three processes greatly expand the range of items that can possess etiological-selectional functions.

Let us take multilayered functions first. Millikan notes that some biological items have a “relational” function, which is a function to do something when bearing a certain relation to something else. The chameleon’s camouflage system has the relational function of matching the chameleon’s skin color to its environmental background, whatever that may be. Given a specific background to adapt to, this mechanism then generates traits with *derived* functions. When the chameleon is crouching on a brown branch, its brown color has the derived function of matching it to the branch. The camouflage

⁹ I say “original” because, of course, many *derived* representational systems – codes, computer languages, and artificial languages generally – have been constructed to serve their representational purposes not by natural selection but by conscious designers. But *original* representation needs natural selection to serve as a “blind” designer.

mechanism might never have produced that shade of brown before, but even so the skin color will have this derived function, courtesy of the fact that the overall camouflage mechanism has been selected to produce whatever color will match the background.

Such multilayered functions are relevant to the many representational systems that are *compositional*, in the sense that they construct complex representations out of simpler components. Such compositional systems often generate entirely novel representations, items that have a meaning even though they have no historical precedents.

Consider the dance whereby bees can “tell” other bees where to go to find nectar, with the direction of the dance indicating the direction of the nectar and the duration indicating its distance. Any particular dance will be adapted to the current location of nectar and so will have a derived function of guiding behavior in a way appropriate to that location. However, the dance that indicates this specific distance and direction might never have occurred before in bee history. If so, it would owe its functionality not to any beneficial effects of previous versions of that specific dance but rather to the beneficial effects of the overall representational *system* for indicating directions and distances.¹⁰

This general model can be applied to many representational systems, including human cognition. I doubt that anybody has ever thought that “old tables make good toothpicks” until I wrote that sentence down. But this won’t stop this mental representation deriving its biological function, and hence its truth condition, from the functionality of the overall system that generates complex thoughts out of simpler concepts.

Still, what about the *elements* in this compositional system, like the simple concepts “table” and “toothpick”? Won’t their functionality have to derive from their past contribution to the selection of genes, even if not in the specific combination “old tables make good toothpicks”? But it is scarcely credible that concepts like these can have any kind of gene-based function.

Here teleosemantics can appeal to the other two ideas mentioned earlier. The first was selection-based “learning.” This doesn’t involve the differential reproduction of organisms over generations but the

¹⁰ Both these chameleon and honeybee examples are discussed in further detail in Millikan (1984).

differential “reproduction” of cognitive or behavioral items during the development of a given individual. Such ontogenetic selection takes place, for example, when cognitive responses are molded by experience during learning. In such cases, we can think of the items selected as having the function of producing those effects in virtue of which they were favored by the learning mechanism. It is arguable that concepts such as “table” and “toothpick” might gain representational functions via this route.

An alternative form of nongenetic selection relevant to teleosemantics is nongenetic “intergenerational selection.” Many traits are passed from parents to children by nongenetic channels outside the sexual “bottleneck”: these traits include the possession of parasites, the products of imprinting mechanisms, and the many cognitive and behavioral traits acquired from parents via social learning. A number of biological theorists have argued that such non-genetically inherited traits can be naturally selected through the normal Darwinian process of differential reproduction of organisms (e.g. Jablonka and Lamb 1999; Mamei 2004). Non-genetically inherited traits that become prevalent in this way will have functions, namely, the effects that favored their possessors. Again, it seems possible that functions of this kind could help to explain the contents of mental representations. After all, it seems a natural enough thought that certain non-genetically inherited ways of thinking are an advantage to their possessors because they make them sensitive to certain features of their environment.

Conclusion

Teleosemantics offers a powerful framework that promises to explain the nature of representation in nonrepresentational terms. At first sight, it might seem unlikely that a simple appeal to biological functionality can account for the significance of truth, or accommodate the determinacy of representational content, or deal with aspects of representation that have no genetic basis. However, once we appreciate the full extent of the resources to which teleosemantics can appeal, we see that it has the flexibility to deal with these and other objections.