Evo Edu Outreach (2010) 3:265–274
DOI 10.1007/s12052-009-0168-6

ORIGINAL SCIENTIFIC ARTICLE

Biosemantics: An Evolutionary Theory of Thought

Crystal L'Hôte

Published online: 16 September 2009

Abstract Evolutionary theory has an unexpected application in philosophy of mind, where it is used by the so-called biosemantic program—also called the teleosemantic program—to account for the representational capacities of neural states and processes in a way that conforms to an overarching scientific naturalism. Biosemantic theories account for the representational capacities of neural states and processes by appealing in particular to their evolutionary function, as that function is determined by a process of natural selection. As a result, biosemantic theories have distinct advantages over other theories of mental representation—e.g., Fodor's causal theory. Foremost among the advantages of biosemantic theories is their ability to account for the possibility of mental misrepresentation.

Keywords Philosophy · Biosemantics · Teleosemantics · Philosophy of mind · Representation · Mind · Meaning · Function · Millikan · Naturalism

Introduction

Resistance to viewing human beings through the lens of evolutionary theory has many sources, not least of which is the fact that eugenics, poverty, and other systematic injustices have been perpetuated by the inchoate idea that the fittest survive in competition. Since both theorists and the wider public are prone to misuse evolutionary concepts and principles to such ends, it is reasonable to think either that extreme care must be exercised if evolutionary theory is used to illuminate any aspect of our humanity, or even to think that evolutionary theory should be utterly off-limits as we seek to understand ourselves. However, notice in this case that the reason for resistance is not that the explanatory tools that evolutionary theory provides are ill-fitted to the explanatory task (as saws for hammering, perhaps) but rather that the tools are dangerous. So understood, this first objection—call it the "moral objection"—does not deny the explanatory adequacy of evolutionary concepts and principles themselves; and it is consistent with the moral objection to think that these tools are exactly correct for the explanatory task. Even so, the moral objection reasonably implores us to resist using them.

However, it is a separable objection that most concerns us here. The "explanation objection" does deny the explanatory adequacy of evolutionary theory; in particular, it denies the adequacy of evolutionary explanations of human choice-making, of human thought and behavior. A couple of reasons are typically given in support of the claim that these phenomena are beyond evolutionary explanation. First, there is the claim that contemporary technologies and modern institutions have rendered selection in the human realm artificial rather than natural, or cultural rather than biological, a claim not unrelated to the idea that selectively breeding dogs, horses, and livestock nullifies evolutionary accounts of pertinent traits. So, although there may have been a time when evolutionary accounts of human choice-making, thought, and behavior—and perhaps other traits—were explanatorily adequate, that time has long since passed.

Second, there is the distinct yet compatible claim that humans possess some special attribute, whether free will or reason or other, that is essential to our very nature and is itself sufficient to place us outside the scope of evolutionary theory. If humans possess the attribute essentially, then

C. L'Hôte (✉)
Department of Philosophy, St. Michael's College,
Colchester, VT 05439, USA
e-mail: clhote@smcvt.edu

human exceptionality is nothing new and did not need to wait on the emergence of any technologies, even if technologies now amplify that exceptionality. Despite the difference between this claim and the first, both (if true) are reasons for denying the adequacy of evolutionary explanations of thought and behavior at the present time. And what often goes hand in hand with the explanation objection, in either form, is the view that the sciences more generally are inadequate to the task of predicting, explaining, or otherwise accounting for the mental life of present-day human beings.

It is in the face of both the moral objection and the explanation objection, then, that the work of Dawkins (1976), Wilson (1978), Pinker (1997), and others persists. The twin research programs that their work represents—evolutionary psychology and sociobiology—aim precisely to use the principles and concepts of evolutionary theory to account for human thought and behavior.[1] To the extent that these programs affirm the adequacy of such explanations, they build the case for what is sometimes called scientific naturalism. Though this variety of naturalism has several dimensions, it is adequate for our purposes to define scientific naturalism as the thesis that the best empirical science(s) can in principle explain everything–including the gamut of human phenomena–even if that best science will not be available until physicists finally hit upon their long-sought "theory of everything." Indeed, it is only against the background of the conflict between scientific naturalism and exceptionalism, ultimately a deep philosophic conflict about the place of human beings in the world, that the full significance of the biosemantic program—our present topic—can be appreciated.

Like both sociobiology and evolutionary psychology, the biosemantic program generates evolutionary explanations of aspects of human thought and behavior. However, the claims made by the biosemantic program are more specific than the claims made by these other programs. Tucked away within the philosophy of mind, the biosemantic program typically uses evolutionary theory to account for only our most basic and primitive thoughts, and to account for only a certain feature of these thoughts, at that: their capacity for representation.[2] (Much more will be said about the representational aspect of thoughts in what follows.) So while the biosemantic program insists that science (biology) has a crucial role to play in explaining an aspect of thought and thereby helps to build the case for scientific naturalism, it is also comparatively modest. As a result, the biosemantic program is one that both naturalists and exceptionalists might agree upon: though the biosemantic program builds the case for scientific naturalism by enlarging the scope of scientifically explicable phenomena, it also leaves much untouched. For this reason and because the biosemantic program—hereafter, simply biosemantics—is an unexpected and fascinating extension of evolutionary biology to a longstanding problem in philosophy of mind, it merits a close look.

What follows is something of a biosemantics tutorial. As a result, fine distinctions between its varieties are passed over as are other complexities that may be of interest to some readers (readers who may wish to consult the works cited). Also, biosemantics is just one of a number of related philosophic programs that endeavor to use scientific concepts and principles to explain the representational aspect of thought. Alternative accounts might aim to account for the representational aspect of thought by invoking only such concepts and principles as are afforded by physics (e.g., cause and effect) without making use of the additional higher-level concepts and principles that are afforded by biology (e.g., trait, selection, fitness, and so forth). So just one approach to understanding the phenomenon of representational thought is developed here, an approach that is typically—though not necessarily—motivated by an overarching scientific naturalism and that is distinguished from other such approaches by its ingenious appeal to the concepts and principles supplied by evolutionary biology.

I have only gestured at the significance of the representational aspect of thought. The first section below says more about what this representational aspect is and explains why this aspect presents an especial philosophic problem. The biosemantic solution to this problem, as that solution is articulated by a celebrated proponent, Millikan (1984, 1993), is presented in the second section. The third highlights the distinctive strengths of biosemantic theories of representation and points to correlative weaknesses in chief alternatives: causal theories and resemblance theories. The fourth and final section identifies ongoing challenges for the program.

The Problem of Mental Representation

June wants an apple. Mark believes that the tide is receding. Isabel imagines green pastures. Though their respective mental states—desiring, believing, and imagining—are of an everyday sort, each has an aspect that seems puzzling on closer inspection. Their respective mental states represent some state or feature of the world, which is also to say that

[1] Perhaps the most controversial thesis to come out of these programs is the thesis that it is more natural, a la biology, for stepfathers to abuse their stepchildren than their genetic children on account of selection pressures that weighed on our Pleistocene ancestors.

[2] Here and throughout, the term *thought* is used in its wide sense, to refer to mental states and processes that represent. Thoughts here include perceptions and exclude sensations insofar as they are non-representational.

there is something that each of them is thinking about, whether apples or tides or pastures. Though an everyday phenomenon, the fact that our mental states have this representational aspect can seem so puzzling that some philosophers of mind see it as a problem: the problem of mental representation or, simply, the "representation problem." Of course, the representation problem is not a problem in the sense that the existence of thoughts that are about things is at all in doubt; after all, few things are more familiar than our ability to think about things. Rather, the representation problem poses a problem insofar as this distinctive feature of our mental states—i.e., the fact that they are about things—threatens to undermine scientific naturalism. In other words, it is scientific naturalists and scientifically naturalistic philosophers of mind who are most inclined to see this representational aspect of thought as a problem, and who are especially concerned to solve that problem.[3]

Seeing just why the very existence of thoughts that represent things might present a threat to scientific naturalism requires little more than comparing thoughts with paradigm physical objects: stones, molecules, and the like. Though these objects do not think, it is more to the present point that paradigm physical objects do not as such represent anything. Since our thoughts—our neural states, arguably—are about things, thoughts can readily seem categorically distinct from paradigm physical phenomena. By being about specific items in the world and by being directed toward those items, our thoughts display a striking power to "reach outside" and somehow point beyond themselves toward those things, a power that may seem strange in comparison to paradigm physical powers (Crane 1995). Again, stones, molecules, and other physical objects do not seem to have this power.

Moreover, this peculiar representational power seems to effect an equally peculiar relationship between the thoughts and what they are about, a relationship (again) that seems quite unlike the kinds of relationships that paradigm physical objects have to one another. To wit, the kinds of relations that obtain between physical objects—e.g., on top of or next to or underneath—are typically sensitive to the time or the location of the physical *relata*. For instance, one stone cannot be on top of another stone unless the two stones exist at the same time, and a cause cannot bring about its effect unless the two are proximate. By contrast, a thinker's ability to stand in the (say) thinking-about relation to some object or event seems relatively insensitive to its time or location. It requires no more effort to have a thought about the past or about the future than it does to have a thought about the present, and no more effort to have a thought about a distant friend than a neighbor. Indeed, one often hears the opposite: that it is more difficult to think about what is present and near than to think about what is far. More striking on this score, however, is the observation that we have little trouble thinking about things that do not even exist—world peace, for instance—and it is difficult to understand how a relationship to something that does not even exist could be a relationship of the same kind as exists between physical objects. For these reasons and still others, the relationship between thoughts and what those thoughts are about gives the appearance of being other-than-physical and consequently of being beyond the explanatory scope of any empirical, physical science.[4]

Indeed, many thinkers—historically, Brentano (1995/1874)—have concluded that the representational aspect or aboutness of thoughts is itself sufficient evidence against a comprehensive scientific naturalism. For the naturalist to meet Brentano's challenge, she must show that this peculiar aspect of thought is not so puzzling after all. She must show that the relation between thoughts and what they are about is ultimately a physical relationship like others, despite appearances to the contrary. Before turning to the bio-semantic response to the challenge, it is important to consider one popular response, one that denies the need for a response.

Though not ultimately successful, a first and plausible reply to the representation problem is to deny that mental representation is a problem for an overarching naturalism on the grounds that representational capacities are to be found throughout the natural, physical world. For instance, it is argued, a pile of stones might signal a mountain summit, and physical patterns of ink (words) on a page can be about (many) things. Since not only thoughts but everyday physical objects are quite capable of representing things, the fact that our thoughts manage to be about things need not cause any particular puzzlement. Again, if a pile of stones can point beyond itself to particular objects or features in the world then a thought's ability to do so does not pose a threat to naturalism. In short, there is no representation problem.

This "no-response response" to the mental representation problem is unsatisfactory. Seeing why sharpens the puzzle. Consider Grice's (1989) distinction between, on the one hand, the way stones and ink represent features of the world and, on the other hand, the way our thoughts do. If a pile of stones manages to represent a mountain summit, it is arguably because we have made it do so; and if ink patterns

[3] Although, as I have suggested above, exceptionalists need not resist a scientific account of this aspect of our thoughts, it may be that we are exceptional for some other reason.

[4] Strictly speaking, it could turn out that our best empirical science is not physical. This would happen if empirical methods led us to conclude that physics is inadequate. However, it is hard to imagine that physics would not evolve with new empirical discoveries.

manage to mean something, it is arguably because we have somehow made them. Would they mean what they do—or mean anything—without us? As Wittgenstein observed, albeit to a different end, "each sign *by itself* seems dead" (Wittgenstein 1953). According to the Gricean, the semantic life stones and ink occasionally enjoy is one our thoughts and activities breathe into them, and their specific meanings are a function of which specific thoughts we have in our use of them. Notice that the same explanation cannot, however, be given for the representational powers our thoughts themselves have. The aboutness of thoughts cannot be explained in this way, by appealing to our other thoughts and intentions, without introducing a circularity that ultimately leaves the aboutness of (some) thoughts unexplained. A satisfactory explanation of the aboutness of thought cannot also take it as primitive.

Responding to the representation problem by noting that stones and ink are about things too, does nothing to lift the mystery surrounding the fundamental sort of aboutness that characterizes thought. Noting that stones and ink are about things arguably achieves the opposite: the derivative aboutness of stones and ink is revealed to be all the more puzzling insofar as it is derived from the as yet unexplained aboutness of thought. In any event, the aboutness of thought remains puzzling and seemingly different in kind than the derivative sort of aboutness we confer upon physical objects.

The mental representation problem merits a genuine response, and it merits a response whether one has naturalistic aims or not. After all, there is this fundamental question to be answered: exactly how does the aboutness of thought fit into the natural physical world? Can science explain it? What might an explanation look like? Again, biosemantics and other naturalistic theories of mental representation aim to show that the aboutness of thoughts is ultimately nothing over and above phenomena that are plainly naturalistic. They aim to show that the fundamental aboutness of thought is, as Fodor (1987) cleverly put it, "really something else"—viz., something patently naturalistic—even if it requires some effort to see that this is the case.

A Biosemantic Solution

Again, biosemantic theories of mental representation are distinguished from others by their mode of explanation: they seek to naturalize the aboutness of thought by appealing to the evolutionary, biological function of the neural states and processes that realize these thoughts. This is the strategy of David Papineau (1987) and of Millikan, who offers an especially detailed account. The aboutness of those neural states and processes that realize our thoughts is accounted for as follows: what makes a neural phenomenon represent some feature of the world is just that it is its proper evolutionary function to coordinate the relevant organism—us, in this case—with that feature of the world. We will look more closely at the notion of a proper function presently, but notice that a successful explanation of the aboutness of thought in terms of evolutionary proper functioning will constitute a naturalistic explanation only if proper functions are themselves naturalistic phenomena.

The notion of an evolutionary proper function is importantly different from the everyday notion of a function, even though it is in some ways similar to it. The proper function of a biological trait—whether structural, behavioral, etc.—is not whatever that trait in fact does, since it might very well malfunction; the proper function of a biological trait is what that biological trait should or ought to do in light of its selection history. It is the proper function of a heart to pump blood (say) even if that heart fails to pump blood or even if it does something else, even something that happens to be useful to the organism. It is the proper function of the heart to pump blood even if it is, we might be tempted to say, functioning in a way that is different than its function. In saying this, however, we are simply using the notion of a function in two distinct ways: descriptively and normatively. The biological notion of a proper function is a decisively normative one since it specifies what a trait ought to be doing and does not simply describe what it actually does. We will soon see that the normativity of the biological notion of a function makes it an especially useful conceptual tool for the task of naturalizing mental aboutness.

In light of the moral objection to evolutionary explanations of human phenomena, it is worth noting that the biological imperatives that proper functions articulate are not thereby also moral or ethical imperatives, even if they bear some significant relation to them. Although biological imperatives may overlap with moral imperatives in some cases—i.e., it may be ethically imperative to feed those for whom it is biologically imperative that they eat—these imperatives may also diverge, and perhaps do so most obviously in the case of sexual reproduction. Though human functioning leading to reproduction may be proper in a biological sense, it may be quite improper in the other. And it is for good reason that we are not in the habit of holding hearts morally accountable for pumping blood, even if it is biologically imperative that they do so. An action (or thought) resulting from the proper functioning of a biological trait is not thereby morally justified, a fact tragically missed in the cultural uptake of evolutionary theory.

Since biosemantics does aim to account for the aboutness of thought in terms of the proper functioning of brain states, it is incumbent upon biosemantics to show that the notion of a proper function is itself suitably naturalistic.

What makes it the case that the proper function of the heart is to (say) pump blood? Demonstrating that the notion is a suitable tool for the explanatory task requires showing that a given trait's proper function is wholly given by the process of natural selection and that its specification does not presuppose thought and interpretation (as above, on pain of circularity). More specifically, proving that the notion of a proper function is suitable requires showing that a trait's proper function is nothing more than an abstraction over the evolutionary process, in the way that the average temperature of a region is nothing more than an abstraction over changing temperature. If the average temperature of a region is nothing more than an abstraction over natural phenomena, its value is independent of our calculations. If the proper function of a trait can be shown to be nothing more than an abstraction over evolutionary processes, the naturalist will be able to rely upon the notion to do naturalizing explanatory work.

Biosemantics offers this specification: the proper function of a trait is just whatever that trait did or brought about that enabled the species to survive and reproduce, i.e., whatever it did in the past that contributed to species fitness. It is just because the heart's pumping blood contributed to species fitness that pumping blood, and not the other things hearts might conceivably or randomly do, is their proper function. If, as this analysis urges, a proper function is ultimately nothing more than an abstraction over the naturally selective process, then the notion of a proper function is demonstrably suitable for showing that yet other phenomena—e.g., representational phenomena—are natural, too.

It is on the foundation of the thusly proven notion of a proper function that biosemantic theories of mental representation are constructed. Though the ultimate ambitions of theories vary, all begin modestly by theorizing only the most basic representational capacities of non-human organisms. Consider Dretske's (1994) example involving the low-level representation that occurs in certain bacteria:

> Some marine bacteria have internal magnets (called *magnetosomes*) that function like compass needles, aligning themselves (and, as a result, the bacteria) parallel to the earth's magnetic field. Since these magnetic lines incline downwards (towards geomagnetic north) in the northern hemisphere (upwards in the southern hemisphere), bacteria in the northern hemisphere, oriented by their internal magnetosomes, propel themselves toward geomagnetic north.

It turns out that heading toward geomagnetic north enables these bacteria to survive by directing them downward and hence away from the oxygen-rich surface water that is toxic for them. As Millikan (1993) sees it, the proper function of the magnetosomes is to coordinate the bacteria with (safe) oxygen-free water and for that same reason, they may be said to represent oxygen-free water. More generally, all it is for the inner state of an organism to *represent* a feature of the environment is for that state to have the proper function of coordinating the whole organism with that feature. Nothing more. And, again, whether the inner state of an organism does have the proper function of coordinating an organism with this or that particular feature of the environment (or some other) is wholly determined by the selection history of the organism-type, by the basis upon which ancestral organisms were selected over other organisms of the same type.

In this way, the selection history of a species imposes constraints on what its biological representations can subsequently mean. For instance, the orientation of a present-day marine bacterium's magnetosome could not now mean oxygen-free water unless ancestral bacteria were selected because they had magnetosomes that coordinated them with oxygen-free water. That is, in order for the orientation of the magnetosome to now represent oxygen-free water, it must be the case that other of the marine bacteria failed to survive and reproduce because they lacked magnetosomes that coordinated them with oxygen-free water—a condition that would be satisfied if, for instance, bacteria that lacked such magnetosomes veered perilously off into toxic oxygen-rich water. In addition, magnetosomes could not have been selected for coordinating the bacteria with oxygen-free water, in particular, unless oxygen-free water actually existed in the ancestral environment. In sum, it is now the proper function of magnetosomes to represent oxygen-free water only if (1) bacteria with oxygen-free water-coordinating magnetosomes were selected over bacteria without oxygen-free water-coordinating magnetosomes and, what is presupposed by this, that (2) there was actually oxygen-free water in the environment of ancestral bacteria. Only because these two conditions were met, the proper function of present-day magnetosomes is to represent the direction of oxygen-free water even if they fail to do so.

In this fashion, biosemantics provides a naturalistic analysis of the representational capacities of very basic biological structures. And since the low-level representational capacities of bacterial magnetosomes are not derivative in the way that the aboutness of non-biological stones and words is—i.e., since it is plainly not anything we think or do that makes magnetosomes represent what they do—naturalizing the representational capacities of magnetosomes is also naturalizing aboutness of the right general kind. And, in the wake of Millikan's analysis, the magnetosome's ability to represent features of the world is not so mysterious-seeming. The hope of Millikan, Papineau, and other biosemanticists is that shifting attention toward representational phenomena in the biological domain—

and away from representational phenomena in the non-biological domain (stones and ink)—will make the thesis that mental representation is naturalistic more plausible.

Of course, their hope is also to show that the phenomenon of mental representation is naturalistic, and in a similar way. If the evolutionary concepts and principles can be used to explain aboutness, or even proto-aboutness, at the basic biological level then there is reason to think they can be used to account for the aboutness of higher-level neural phenomena. For instance, it is now open to the biosemanticist to argue that neural states or processes in us are about edges or food or danger just in virtue of the fact that directing our ancestors toward or away from these items conferred some selective advantage upon them.

A Strength of Biosemantics: Accommodating the Possibility of Misrepresentation

Representations at any level are not always veridical, accurate, correct, or true. The magnetosome may malfunction, causing the bacterium to stray into toxic water, and thoughts are often false. Indeed, the phenomenon of representation carries the possibility of misrepresentation along with it. As a result, the biosemantic theory of representation is only halfway there: any complete theory of representation must account for the aboutness of those many representations that are misrepresentations. As it happens, accounting for the aboutness of misrepresentations is the most daunting task facing naturalistic theories of representation.

Showing that error of any sort is a wholly natural phenomenon is not easy. Error is not to be found either at the sub-atomic level (muons and leptons do not make mistakes), or at the chemical level (sulfuric acid does not err). And if error does not "go that deep," then it would seem to follow that the phenomenon of representation cannot go that deep either (Fodor 1987). Again: where goes representation, there too goes the possibility of misrepresentation. But the challenge misrepresentation poses to scientific naturalism is even more daunting than this. The depth of the sub-atomic and chemical levels is not the only reason error cannot be located there. Medium-sized and enormous physical objects—stones and planets—do not make any mistakes either, even if we are guilty of mistakes involving them.

Notice, however, that error of a kind seems to occur at the biological level. The real ingenuity of the biosemantic solution ultimately rests in its appeal to a naturalistic domain that, although as natural as the domains of physics and chemistry, nonetheless includes the possibility of a kind of error. It is here that we find, as Dretske put it, "nature's way of making a mistake" (Dretske 1994): hearts and other organs malfunction, magnetosomes lead bacteria into toxic waters, and chameleons fail to change color. Things go awry. The possibility of error at the biological level gives biosemantics a distinct advantage over its chief alternatives—causal theories and resemblance theories—which do not avail themselves of biological notions. Since these alternative theories of representation do not make use of the phenomenon of biological error, they are especially hard-pressed to account for the possibility of representational errors, and hence to provide a full account of the phenomenon of mental representation.

Both resemblance and causal theories borrow on common sense views about representation-in-general. To wit, it seems as though paintings resemble what they represent, and that what a photograph represents has something to do with its causes: what makes this a photo of *Fido* is that Fido was in front of the camera at the time that the photo was taken. The resulting theories of representation can be respectively cast as follows (following Crane (1995), R stands for the representational state or process of an organism type, O and C stands for the content of the representation, i.e., what it is about):

R in O represents C if and only if R resembles C.
R in O represents C if and only if R is caused by C.

Although both resemblance theories and causal theories begin with common sense views about how representations manage to represent what they do, these theories do not and cannot end there. For instance, it is imperative that a pure resemblance theory of representation specify a relevant mode of resemblance. After all, an organism's representational state is not, as a simple theory suggests, about absolutely everything that it resembles: magnetosomes are needle-like but do not thereby represent needles. And there are several respects in which a representation does not resemble what it represents, for instance with respect to color or shape: magnetosomes do not at all look like the oxygen-free water that they nonetheless represent. Without further qualification, then, simple resemblance is neither sufficient nor necessary for representation.

For similar reasons, it is incumbent upon a pure causal theory of representation to specify a relevant mode of causation. Plainly, the representational state of an organism is not, as a simple causal theory suggests, about all of its causes: magnetosomes do not represent adjacent structures within the marine bacteria, even though causally affected by them. Nor does a representation always causally interact with what it is about, for the plain reason that what a representation represents may be absent or may not even exist in individual cases. For instance, an individual magnetosome may represent oxygen-free water even if there is none in the vicinity, as when it malfunctions. And when we make perceptual errors, as may well happen on a

dark or foggy night, the meaning or content of our representation is not what caused it, in all but the strangest cases. For instance, if we mistake a skinny cow for a normal-sized horse, the content or meaning of the representation is "horse" even though it is caused by a (skinny) cow. Without further qualification, then, it is clear that being the cause of a representation is neither sufficient nor necessary for being what the representation is about, as a simple causal theory maintains.

Again, both resemblance and causal theories of representation must specify the relevant modes of resemblance and causation, both because resemblance-relations and causal-relations are ubiquitous and because simple relations usually fail to hold if an error has occurred, i.e., if those representations are misrepresentations. Indeed, proponents of both resemblance and causal theories have made some progress in specifying those relevant modes. According to an amended causal theory, for instance, a representation is about its cause only if conditions are normal:

> R in O represents C if and only if R is caused by C in normal conditions.

Accordingly, the direction of the magnetosome represents oxygen-free water only if the bacterium is in its normal marine environments, and not if a bar magnet is waved above it or if it is transported to the opposite hemisphere (where polarity is reversed). An amended causal theory does not presume, then, that the cause and the content of a representation will coincide unless the environment in which the representing takes place is normal or—as often put—ideal. In this way, the amended causal theory proposes to distinguish between relevant (ideal) and irrelevant (non-ideal) causes. And the amendment seems to get the right result: the magnetosome does not represent oxygen-rich water in an environment in which the local polarity is reversed, even if it directs bacterium toward the oxygen-rich water. Rather, the representational content is determined by what the individual magnetosome would do in the environment that is normal or ideal for it (even if it never gets there).

Though it seems to get the right results, the amended causal theory is hard-pressed to give a non-circular specification of the normal or ideal conditions for representation. Consider: is a dark night an ideal condition for perception? It depends; dark nights may not be ideal for seeing horses but they are ideal for seeing stars. If the specification of ideal conditions ultimately depends on what the perception is about (its content), as it seems, then the specification also presupposes the very phenomenon it is invoked to explain. At the same time, if the specification of ideal conditions does not take the content of a representation into account—i.e., if it is insensitive to what the representation is about—then it seems doomed to fail. If daytime is ideal then veridical star-perceptions will thereby be ruled out; if the nighttime is ideal then all representations caused by skinny cows at night will thereby represent skinny cows, and ordinary perceptual mistakes will be ruled out. The results are unacceptable, yet naturalism prohibits the amended theory from presupposing aboutness in its specification of ideal conditions; and to do so would be to presuppose the very aboutness that the theory sets out to explain. In the end, then, even an amended causal theory falters in the face of misrepresentation.

Amended resemblance theories are similarly susceptible to the special problem misrepresentation poses, but it is not necessary to see how all of that goes in order to appreciate the difficulty of the challenge. Naturalistic theories struggle to explain the aboutness of misrepresentations in particular, because the usual coordination and correlation between an organism and its environment, or between a perceiver and her world, is often disrupted when error and misrepresentation occur. Many of the simple relationships that obtain between a veridical representation and what it represents fail when misrepresentation occurs and unusual relationships take their place. Nonetheless, a successful naturalistic theory of representation must identify a relationship that does not change, a single naturalistic relationship that holds between a representation and what it represents whether the representation is veridical or non-veridical, whether it is a representation or a misrepresentation.

Biosemantic theories are remarkably well poised to accommodate misrepresentation. In virtue of their distinctive appeal to evolution, they seem to do this without inadvertently presupposing content or aboutness. Again, Millikan's particular version of the theory is that the bacteria's magnetosome represents oxygen-free water if and only if it is the function of the magnetosome to coordinate the bacteria with oxygen-free water. Leading the bacteria toward toxic, oxygen-rich water would not constitute successful coordination. Consequently, it would constitute misrepresentation. Misrepresentation is simply a species of malfunction.

A recap: for it to be the case that this individual magnetosome has the proper function of coordinating this individual bacterium with oxygen-free water is just for it to be the case that being so coordinated with oxygen-free water contributed to the biological fitness of ancestor bacteria, i.e., that at some point in the evolutionary history of the bacteria, those that had magnetosomes that played this role were (naturally) selected over those that did not—a process that explains the persistence of the trait in descendent bacteria. According to Millikan and others, all of these episodes, events, and sequences in the organism's past are what make this present fact true: the orientation of this individual magnetosome now represents the direction of oxygen-free water. Likewise, all of these episodes, events, and sequences in the organism's past are also what make it true that the magnetosome misrepresents the

location of oxygen-free water if it orients the bacteria toward toxic surface water.

Through the biosemantic program and, in particular, by enabling the program to link the phenomenon of misrepresentation to the phenomenon of biological malfunction, evolutionary theory has made a surprising and significant contribution to the project of showing how the phenomenon of representation might fit into the natural world. In doing so, it has brought us a step closer to understanding how the phenomenon of mental representation might fit there, too. Perhaps neural structures and processes, too, have been naturally selected for representing particular features of the world. Biosemantics makes it conceivable.

Ongoing Challenges

However, the program is not without its challenges, three of which are mentioned here. A fuller treatment of them can be found in the works cited.

"The past is never dead," Faulkner (1975) once wrote, "In fact, it is not even past." Although the biosemanticist would not assent to the bare assertion that the past is the present—nor is Faulkner likely to have done so—she agrees that (selection) history determines what present representational structures and processes are about, just as it determines the present function of non-representational structures and processes (hearts). A consequence of this theory—one which offends the sensibilities of Davidson (1987) and others—is that if a bacterium were to come into existence spontaneously, its magnetosome would have no function and would not represent anything at all. Similarly, if a humanlike being were to pop into existence, and with a brain and neural firings physically identical to any of ours, it is a consequence of biosemantics that the being's brain and neural firings would have no function and not be about anything at all. No history, no function; no function, no representation. Now, whether this ultimately constitutes an objection to biosemantics depends on whether that result is acceptable, and this is not clear to many.

However, it is clearly an objection to biosemantics to argue that its appeal to history rules out the possibility of thoughts about telephones or computers, not to mention thoughts about things that do not yet exist. Plainly, we do think about such things, even though being coordinated with telephones and computers (say) did not confer any selective advantage on our Pleistocene ancestors. Millikan's reply to the objection is resourceful; she argues that it is the proper function of some set of neural firings to coordinate us with these modern items for the same reason it is the proper function of some device in a chameleon to coordinate it with a color even if its ancestors never encountered that specific color before. She argues (Millikan 1994a):

> The brain structures we have recently been using in developing space technology and elementary particle physics have been operating in accordance with the very same general principles as when prehistoric man used them for more primitive ventures. They are no more performing new and different functions or operating in accordance with new and different principles nowadays than are the eyes when what they see is television screens and space shuttles.

The same may be true of structures and processes which enable us to entertain thoughts about televisions and space shuttles. According to Millikan, what occurs in such cases is merely a novel application of an old rule or principle which, when followed by particular neural structures and mechanisms in the ancestral past, led to their selection. In the chameleon case, that rule is: turn the color of the surrounding surface, whatever that is; and since the device observed the rule, the chameleon was coordinated with its environment, and the device persisted in descendant chameleons. Of course, if Millikan is to avoid the charge of being ad hoc in her reply to this objection, she must also make sure that the rules instantiated by our biology limit our representational capacities. (Rules which allow everything are not rules.)

A final objection is that biosemantics is constructed upon a foundation that proves unstable on close inspection: the notion of a proper function. Arguably, the notion of a proper function is neither clear nor determinate. Take the magnetosome case. Millikan's position is unequivocal: the magnetosome represents oxygen-free water since being coordinated with oxygen-free water contributed to species fitness. However, Millikan's ruling is not decisive. In fact, there is serious disagreement over how the function of the magnetosome should be specified. Imagine again that an unsuspecting bacterium from the Northern Hemisphere is transported to the Southern Hemisphere or (again) deposited into a Petri dish over which magnets are passed. The magnetosome will probably not lead the bacterium to oxygen-free water in these situations and may even lead the bacterium straight to its destruction. According to Millikan, the magnetosome would thereby malfunction. But surely, some theorists contend, we should hesitate before concluding that the magnetosome has malfunctioned here. Surely it would be unfair to expect the magnetosome to point the bacterium in the direction of oxygen-free water in these strange situations.[5] If this is correct, then the proper function of the magnetosome is simply to coordinate the bacterium with magnetic north, wherever that should lead, and not to coordinate the bacterium with oxygen-free water.

[5] Although, again, moral notions of responsibility do not apply at this level, they do effectively highlight the difficulty of specifying the proper function of a biological representation and, hence, what it is really about.

On the other hand, and as tempting as it may be to narrow the function of the magnetosome down to just coordinating the bacterium with magnetic north, the reason why the magnetosome has persisted clearly includes its capacity for coordinating the bacteria with more distal features of the environment. Neither the bacteria nor the magnetosome would have survived if the direction of magnetic north had also been the direction of toxic water, for instance, so it is not just pointing the bacteria toward magnetic north that explains the persistence of the bacteria and its magnetosomes, a fact of which Millikan is keenly aware.

Narrowing the proper function too drastically may let the magnetosome off the hook in every case; that is, it may be near-impossible for all but the most deformed magnetosome not to point the bacteria toward magnetic north. More generally, if the proper functions of representing traits are too narrowly specified, their specification will not allow for the possibility of misrepresentation, which would undercut the most theoretically attractive feature of the biosemantic account. In short, it is incumbent upon a successful biosemantics to specify proper functions in a principled way and in a way that is neither too broad nor too narrow, neither too demanding nor too lax. The specification must be broad enough to allow for the possibility of misrepresentation and yet narrow enough to ensure a degree of representational success sufficient for explaining the persistence of the representing trait. Executing this balancing act is just one item on the biosemantic agenda, one taken up by Neander (1995, 1991) and Sober (1993, 1984), among others.

These three challenges face any biosemantic theory of representation; which additional challenges arise depends on the specific ambitions of individual theories. Using the evolutionary past to account for the aboutness of complex human thoughts, as Millikan and Papineau ultimately aim to do, is the most ambitious goal to which biosemantics ever aspires, while a most modest biosemantics denies that there is any fruitful continuity between low-level biological representation (about which it has much to say) and human thoughts (about which it will say nothing). Though an ambitious biosemantics faces additional challenges, the committed naturalist may find these worth confronting: since the biosemantic solution to the problem of mental representation is arguably the most promising naturalistic solution available, thoughts that elude biosemantic explanation are thoughts that may well elude naturalistic explanation altogether. Clearly, a theorist who is willing to let the thesis of human exceptionalism stand will not be similarly motivated.

But it would be an easy mistake to conclude that a most modest biosemantics is more correct simply for its modesty: to deny that there is any fruitful continuity between low-level biological representations and human thought is arguably as implausible as the naïve adaptationism that often characterizes the programs of sociobiology and evolutionary psychology. After all, a theory may be just as incorrect by being too modest in its claims as by being too bold. Millikan (1994) insists, "To suspect that the brain has *not* been preserved for thinking with or that the eye has *not* been preserved for seeing with—to suspect this, moreover, in the absence of any alternative hypotheses about causes of the stability of these structures—would be totally irresponsible." Admitting that brains have been selected for thinking is admitting that thinking is their proper function and allowing that, on occasion, it may be their proper function to think useful and perhaps even correct thoughts. If so, there is good reason to believe that the representational aspect of human thought can in some part be explained by evolution, if not just yet.

Yet even Millikan (1994), whose program is most ambitious of all, denies that "bacteria and paramecia, or even birds and bees, have inner representations in the same sense that we do." Though the bacterium's magnetosome enables it to represent, the bacterium does not thereby perceive or think. There are significant differences between low-level biological representations and human thoughts. Still, Millikan argues that the sense in which we have (mental) representations is explicable in terms of lower-level representations plus other explicable features—e.g., non-self-representing elements, storage, etc. Millikan argues that the additional features that distinguish thoughts and other mental representations from low-level biological representations are also perfectly amenable to evolutionary explanation. However, it would be sufficient for the purposes of an overall naturalism if these supplementary features were amenable to any naturalistic explanation, whether evolutionary, chemical, physical, or other. But Millikan's chief claim here is, again, that mental representations are continuous with basic biological representations. If we can explain how magnetosomes represent and misrepresent, then we will be that much closer to explaining how our thoughts represent and misrepresent, and that much closer to understanding how the aboutness of June's desire for an apple, Mark's belief about the tide, and Isabel's imaginings is possible in a natural world. And we will perhaps be that much closer to knowing if and even just when these mental states occur.

Conclusion

Philosophers have long labored to understand the place of humans and minds in the natural world. It is a surprise to many that the theory of evolution, an ostensibly biological theory, is responsible for much of the progress that has been made on this perennial philosophic problem. In the hands of the biosemanticist, the theory of evolution offers us a way to understand what it is that makes one part of the

world about another part of the world. It begins to explain how neural firings might come to represent something, and to represent one thing rather than some other. Perhaps, the program suggests, for neural firings to represent something is simply for them to have been naturally selected for coordinating an organism—perhaps us—with that something. In this way, biosemantics begins to show how the norms of representation and mind could ultimately derive from natural, biological facts.

As a way to understand how low-level biological representations represent, and what they represent, the biosemantic strategy is promising. The challenges that face it do not seem insurmountable, and Neander (1995, 1991), in particular, has made great strides in addressing and managing the worry that proper functions are indeterminate. Even as a method for understanding how simple and primitive perceptual representations in non-human organisms, and even humans, manage to represent, biosemantics offers compelling solutions. Even if biosemantics cannot explain our ability to represent such novelties as televisions and space shuttles, it is plausible to think that more primitive representational abilities have been, as it were, seared into our brains: the ability to represent precipices, food, predators, and the like. Its prospects for explaining higher level and higher order thought are less certain, but few programs aim this high.

In the end, the success of biosemantics as a theory of mental representation may depend on the extent to which the theory can be confirmed by empirical research. Though biosemantics outlines an ingenious strategy for the naturalization of the representational capacities of non-human and human organisms, its success as an account of mental representation may ultimately require showing that neural states and processes were indeed selected for thinking, as Millikan reasonably suggests, and that thinking—and thinking particular thoughts—is not simply a by-product of some more crucial function. In this respect, the biosemantic program waits on the collaboration of empirical researchers.

Acknowledgments I am grateful for the comments of Adam Goldstein, Timothy Mackin, and an anonymous reviewer.

References

Brentano F. Psychology from an empirical standpoint. 2nd English edition. London: Routledge; 1995/1874.
Crane T. The mechanical mind. New York: Penguin; 1995.
Davidson D. Knowing one's own mind. Proceedings and Addresses of the American Philosophical Association. 1987;60:441–8.
Dawkins R. The selfish gene. New York: Springer; 1976.
Dretske F. Misrepresentation. In: Stich S, Warfield T, editors. Mental representation: a reader. Cambridge: Blackwell; 1994. pp. 164–167.
Faulkner W. Requiem for a nun. New York: Vintage; 1975. Act I Scene III.
Fodor J. Psychosemantics: the problem of meaning in the philosophy of mind. Cambridge: MIT; 1987. p. 97.
Grice P. Studies in the ways of words. Cambridge: Harvard University Press; 1989.
Millikan R. Language, thought, and other biological categories. Cambridge: MIT; 1984.
Millikan R. White queen psychology and other essays for Alice. Cambridge: Bradford Books, MIT; 1993.
Millikan R. Biosemantics. In: Stich S, Warfield T, editors. Mental representation: a reader. Cambridge: Blackwell; 1994. pp. 253–255
Neander K. The teleological notion of function. Australasian Journal of Philosophy. 1991;69:454–68.
Neander K. Misrepresentation and malfunction. Philosophical Studies. 1995;79(2):109–41.
Papineau D. Reality and representation. Oxford: Blackwell; 1987.
Pinker S. How the mind works. New York: W. W. Norton and Company; 1997.
Sober E. The nature of selection: evolutionary theory in philosophical focus. Cambridge: Bradford/MIT; 1984.
Sober E. Philosophy of biology. Boulder: Westview; 1993.
Wilson EO. On human nature. Cambridge: Harvard University Press; 1978.
Wittgenstein L. Philosophical investigations. Trans. G. Anscombe. New York: MacMillan; Sec. 432.; 1953.