Rethinking the Evolution of Consciousness

Thomas W. Polger
Department of Philosophy
Program in Brain and Mind Studies
University of Cincinnati
Cincinnati, OH 45221-0374 USA

thomas.polger@uc.edu

1. Introduction

Suppose that consciousness is a natural feature of biological organisms, and that it is a capacity or property or process that resides in a single organ. In that case there is a straightforward question about the consciousness organ, namely: How did the consciousness organ come to be formed and why is its presence maintained in those organisms that have it? Of course answering this question might be rather difficult, particularly if the consciousness organ is made of soft tissue that leaves at best indirect fossil records, or if it has been fixed in the populations for such a long time that there are few available examples of organisms that lack the consciousness organ on which to conduct comparative experiments. No doubt there are other confounding practical obstacles as well. But these are just the complications that face biologists and natural historians on a regular basis, and they do not reflect any special problems about the study of consciousness. This is just to say that if consciousness is a natural feature of biological organisms then its origins and history can be studied in the same manner as other features of the biological world. It's a hard business, but biologists are pretty good at it.

The situation that I have asked you to imagine is a caricature that lies somewhere between simplification and sheer fantasy. In all likelihood there is no consciousness organ. But then again, there is no single circulatory organ, or respiratory organ, or digestive organ.

Nevertheless, it is a respectable pursuit to inquire about the natural histories of circulation, respiration, and digestion; and to inquire about the organs and systems that enable those capacities and activities. The idea that hearts by themselves circulate blood is fine for grade school. But full understanding of the metabolic interdependence of the totality of systems that compose an organism surely reveals the idea of an isolable circulatory system as a gross simplification. This, of course, is no obstacle to studying the natural history of circulatory systems. (But for qualifications see Allen 2002.) Indeed, although complexity makes the task hard it also provides some of the most compelling evidence.

What, then, of the imaginary consciousness organ? Is this idea a useful simplification or a misleading fantasy? My own view is that the imaginary consciousness organ is more like a simplification than a fable, just like the grade school stories about hearts, lungs, and stomachs. Conscious experiences are natural features or processes that occur in biological organisms. I doubt that there is a single consciousness organ that is localized and modular. This does not mean that the goals of discovering the mechanisms and natural history of conscious experiences are hopeless. It does suggest that the task will be difficult. Later I will return to consider what such projects might look like, and what progress may have been made.

But most discussion of the origins and maintenance of consciousness is not about the relative merits of one or another natural history explanation of consciousness. Instead, the focus tends to be on various lines of reasoning that purport to show that *if* some particular explanation (or general class of explanations) of the history of consciousness were correct, then this would reveal something about the fundamental nature of consciousness. In contrast to the relatively straightforward "natural history" reasoning about consciousness, this second kind of consideration concerns theoretical connections between the etiology of consciousness and

philosophical theories of its nature. These lines of reasoning are speculative or philosophical; they focus on what some evidence might show rather than on what evidence we actually have. In this way, discussions concerning the origins of consciousness are quite different from those about the origins of hearts, lungs, and stomachs. And it is these lines of reasoning that concern me in this chapter.

There is a third line of reasoning about the etiology of consciousness that I will mention only to set it aside. These are the so-called teleological or teleofunctional theories of consciousness. Roughly speaking, these are theories according to which conscious mental states are a special kind of representational or functional state of brains or nervous systems, and according to which representational or functional states must be understood in terms of biological function. The most explicit applications of such theories to consciousness come from William Lycan (1987), Fred Dretske (1995), and Robert Van Gulick (1980). Likewise, Jerry Fodor (1968), Daniel Dennett (1991), and Owen Flanagan (1992) have hinted at such a theory for some mental states, if not conscious mental states specifically. Although the teleofunctional view of mind is perhaps most often associated with Ruth Millikan (1984, 1993), she does not seem to offer it as a theory of consciousness. I will set these theories aside because they are best thought of as representational theories of consciousness which also take an teleological or etiological approach to explaining representation. This is not to suggest that they have nothing to offer say the natural history of consciousness—see especially Dretske 1995. But my focus here is on the second kind of reasoning about consciousness.

2. Natural History, Adaptation, and Just-So Stories

Excepting the title of this chapter, I have not yet used the term 'evolution' or any of its related terms. Instead I have spoken only about the origins and natural histories of biological organisms, and their features, capacities, or organs. Now I will begin to use the terminology of evolutionary theory, the theory of the origins and natural histories of organisms and their traits.

If conscious experience is a natural trait of biological organisms then there is an evolutionary explanation for its presence in those organisms. But we must be cautious. Not every property of an organism is a trait—Stephen J. Gould famously argued that the panda's "thumb" and the shape of human chins are not traits. And not all evolutionary explanations of traits are adaptation explanations, for not all traits are formed or maintained by a process of adaptation through natural selection. Some traits could be formed or sustained by chance—mutation or drift—or by self-organization. Nevertheless, adaptation explanations are the default explanations for complex traits. (Needless to say, deciding what is a complex trait directs our attention back to my first caution, concerning which features of organisms are genuine traits.) Traits formed by natural selection are adaptations, and they are sometimes said to have evolutionary, etiological, or "proper" functions.

Many people who have only a casual familiarity with evolutionary theory think that all evolutionary explanations are adaptation explanations—that every evolved trait is an adaptation. But this is not correct. There is an important difference between evolution and selection. The pandas' "thumb" evolved, but if Gould (1980) is right it was not selected for by natural selection, so it is not an adaptation. Additionally, we have already noted that not all properties exhibited by organisms are traits at all—so it is with the shape of the human chin, and probably the ability to do calculus as well. (Though both the shape of the chin and the ability to do calculus are good

candidates for features that are made possible by the adaptation of other traits—of the developmental path and shape of our jaw bones, and the structure of the brain, respectively.) Also, some features of organisms that originally appear by chance may later prove to be useful and subsequently be favored in the process of natural selection. The length of the bone that forms the panda's "thumb" may be one such case of *exaptation* (Gould and Vrba 1982). Finally, some people identify evolution with gradualist theories of change over time, according to which descent with modification occurs slowly and continuously. My discussion of evolution will be entirely neutral about whether evolutionary change is gradual or "punctuated" or sometimes each. These disputes concern not whether adaptation is the primary mechanism that shapes organisms on our planet, but how dominant it is, just how it works, and what other biological processes also play a role in evolution. Such disagreements are, as they say, in-house.

Now we have the resources we need to restate our questions about consciousness. If consciousness is a natural biological trait, or is a system of such traits, then we should expect that there is an evolutionary explanation for its presence in those organisms that have it. If it is complex or is part of a complex system, then we should expect that there will be an adaptation explanation for it, or for some of its features, or for the organization of the complex system. I claimed that there will be evolutionary explanations for conscious experiences. Moreover, I expect that some of these will be adaptation explanations—that some sorts of consciousness, at least, were selected for by natural selection. Just what these explanations might be is a topic that we will return to later. But as I indicated above, most discussion of the evolution of consciousness concerns not how such evolutionary explanations ought to go. They concern, instead, whether consciousness is a natural phenomenon at all, if so whether it is a trait in the special sense relevant to evolutionary theory, and if not whether anything can be inferred about

its origins. These lines of reasoning concern whether there are general considerations about consciousness or about evolution that can help settle the questions of whether consciousness is a trait, or an adaptation, or a natural phenomenon at all. Such are the most prominent questions about the evolution of consciousness.

Before we examine some arguments concerning the evolution of consciousness, we need to understand what a good explanation of the adaptation of consciousness would look like. Such an explanation would ideally include (i) evidence that selection has occurred, (ii) an ecological explanation of adaptive advantage, (iii) evidence that the trait is heritable, (iv) information about the population structure, and (v) phylogenetic information about trait polarity (Brandon 1990: 165-74). Of course most actual adaptation explanations are not ideally complete, but that does not undermine the regulative ideal. In this framework we can make some general observations about the evolutionary explanations of consciousness that have been offered. Most theories of the evolution of consciousness simply take for granted that trait polarity (v) favors consciousness—that conscious creatures evolved from nonconscious creatures. Practically no theorist says anything at all about the population structures in the proposed adaptive environment of consciousness (iv). But almost every theorist assumes that consciousness or the capacity for consciousness is (or is dependent on) a biological trait (or set of traits) that can be passed from parent to offspring (iii). And almost every so-called evolutionary explanation of consciousness is in fact an ecological story about the purported adaptive advantage of consciousness (ii). Often it is argued that such a story, given the presence of consciousness in some creatures, shows that consciousness could have evolved; but practically no theorist bothers to give evidence that consciousness did in fact evolve (i).

My purpose in making these observations is not to offer a blanket critique of evolutionary theories of consciousness, but only to draw attention to their incompleteness. It is important to notice that most stories of the "evolution" of consciousness are stories about what adaptive advantage consciousness might have had in some hypothetical environment. These are ecological stories of the sort that are sometimes ridiculed as "just-so" or "how-possibly" stories. One reason that just-so stories are derided is that typically no evidence is offered that supports any claims about the adaptive environment for the evolution of consciousness—for example, no evidence is given regarding the other organisms that were competing in the environment. Lacking that information, we have no evidence that creatures with consciousness were more fit than their nonconscious peers, no evidence that consciousness conferred any advantages at all. Such omissions are what separates these works of historical fiction from genuine explanations. Converting "how possibly" stories into adaptation explanations requires filling in the other parts of the explanation to show that adaptation not only could have occurred but did in fact.

The above complaint would be devastating to any theory that mistook a "just-so" story for an explanation. But most philosophical and psychological theorists writing about consciousness are not aspiring to give ideally complete adaptation explanations, or even to approximate them. So while we should keep the ideal of complete adaptation explanations in mind, we should also look at the other uses for how-possibly stories. It seems that many theorists, rather than aiming to establish the facts of natural history, are arguing that the availability (or lack thereof) of some evolutionary or ecological story helps (or would help) to favor some theories of the nature of consciousness over others. As we shall see, there are problems with this methodology that are more serious than the mere failure to satisfy an explanatory ideal.

3. Questions About the Natural History of Consciousness

Later I will outline a few explanations of the etiology of consciousness that attempt to go beyond just-so stories. Only time and evidence will tell us whether any of those particular explanations is on the right track. What we can presently evaluate is the role that evolutionary explanations are claimed to play in broader theorizing about consciousness. In this section I will consider some of the most prominent questions that arise in evolutionary reasoning about consciousness. There are four basic questions and each comes in two versions.

Q1a. *If consciousness can be shown to have evolved, does that establish that it is a natural phenomenon?* This is an odd question, admittedly—for how could we know ahead of time that consciousness has evolved? But it is just another way of asking whether consciousness could evolve if it were not a natural phenomenon. Without further constraint, the answer is clearly that nonnatural consciousness could have evolved. Versions of dualism are easy to think of, and it is not hard to concoct an epiphenomenalist version according to which consciousness is a free-rider that manifests itself in certain animals. Perhaps Thomas Huxley held such a view. Perhaps David Chalmers (1996) holds this view; and, if Chalmers is right, then all nonreductive physicalists are stuck with this view. If we want to know what such a view would look like, just imagine a dualist panpsychism according to which the nonnatural properties need to be organized in a certain way in order to constitute consciousness, and then let evolution of animals happen to sometimes form that arrangement. Consciousness, on this picture, is a nonnatural feature that supervenes on the natural features.

At this point it is useful, if not overdue, to say something about the distinction between natural and nonnatural phenomena. I have been assuming that any philosophical theory of consciousness that is broadly dualist will also be one that counts consciousness as nonnatural. This is not an unusual assumption. Nevertheless, it should be noted that some theorists adopt an expanded conception of the "natural" which allows for at least certain kinds of dualistic properties to count as "natural" (e.g., Chalmers 1996, Velmans 2000). These do not count as versions of naturalism for me. But this may be a merely terminological stipulation. The crucial point, as will become clear, is not the distinction between natural and nonnatural but between causal and noncausal. I hold that these distinctions go together: natural with causal, nonnatural with noncausal. But for present purposes this can be regarded as a terminological stipulation. In this paper I am not concerned to establish the truth of any particular theory. Rather, I am concerned about the relationships between claims of naturalism or epiphenomenalism on the one hand, and claims about the evolution of consciousness on the other.

Because consciousness could evolve even if it were nonnatural or noncausal (epiphenomenal), the mere claim that it evolved will not tell us whether or not it is natural or causally potent. So the negative answer to Q1a has little to do with evolution, and much to do with creative freedom of theorizing about nonnnatural phenomena.

Q1b. If consciousness can be shown to have been selected for by natural selection, does that establish that it is a natural phenomenon? While it is easy to imagine the evolution of nonnatural or noncausal consciousness, it is harder to see how such consciousness could be selected for. That is because selection requires causal interaction, it requires that consciousness make a difference in the world by making a difference for the creature that has it. If immaterial or

otherwise nonnatural consciousness can causally interact with the world, then I suppose that it could be selected for. In that case, the fact that consciousness was selected for (that it is an adaptation) does not show that it is a natural phenomena. But if, as I suppose, only natural phenomena can causally interact, then consciousness could not have been selected for unless it is a natural phenomenon. I conclude that if consciousness is an adaptation then it is a natural phenomenon. (Although Descartes would disagree, the position that consciousness is both nonnatural and causally efficacious is not prominent among contemporary theorists.)

Q2a. If consciousness can be shown to have evolved, does that show that it is causally potent? For the same reasons mentioned in answering Q1a, consciousness need not be causally efficacious in order for it to have evolved. Consciousness can be an impotent immaterial or nonnatural free-rider, a side effect of the evolution of natural organisms. But, also following Q1a, this conclusion is too weak to be of much interest. The mere claim that consciousness evolved does not tell us much about the nature of consciousness at all. For the evolution of consciousness is compatible with any (noneliminativist) account of its nature.

Q2b. If consciousness can be shown to have been selected for by natural selection, does that establish that it is causally potent? Consciousness would have to be causally potent in order to be selected for. If consciousness makes no difference in the world, then there would be no ecological advantage for the things that have it over the things that do not that could be the basis for selection. There might, of course, be other differences between the conscious and nonconscious creatures, such that the ones that are conscious happen to have some selection advantage. But that is precisely the Q1a/Q2a scenarios, in which consciousness evolves without

having been selected for. Per Q1b, if consciousness was selected for—if it is an adapatation—then it is causally efficacious.

Given my terminological stipulation concerning naturalism and causal powers, Q2a and Q2b only repeat Q1a and Q1b, respectively. But if causal efficacy and naturalness come apart, the Q1 and Q2 questions will be distinct in the ways noted.

Q3a. If consciousness is necessary for some capacity ϕ in creature C, does that show that it is causally potent with respect to the ϕ-ing of C? Some theories of consciousness hold that it is causally impotent, that it is not capable of bringing about causal effects. Against this kind of epiphenomenalism, some have argued that consciousness must have causal effects because it is necessary for some capacity that conscious creatures actually have. This line of response appears to be valid. If some creature C can do *ϕ* and if only consciousness enables one to *ϕ*, then it seems clear that consciousness is causally responsible for the *ϕ*–ing of C. So, yes, if consciousness is necessary for some capacity ϕ in creature C (and creature C can ϕ), then consciousness is causally potent with respect to C's ϕ-ing. But is there any such ϕ?

Many theories of consciousness attempt to identify a feature or ability ϕ that cannot occur nonconsciously. Among the most popular options are flexible behavior (William James), creativity (Selmer Bringsjord), communication or mental rehearsal (Peter Carruthers), self-knowledge (Nicolas Humphreys), mentalistic language (Todd Moody) and self-awareness of a special sort (Daniel Dennett). Of course different theorists have different ideas about the nature of consciousness, and so these proposals may seem more or less radical. What concerns me, however, is the general form of the claims: that there is some ϕ that cannot be performed (by creature C) without having trait T—where T is consciousness, in the case at hand. If this is true,

then consciousness is a very unique trait, indeed. Is there any other biological trait for which an analogous claim would be true? You might think that, say, birds cannot fly without wings. Since birds do fly, and they do have wings, then it looks as though we have a valid argument that wings are causally efficacious in bird flight. And since the conclusion is true, the reasoning looks good. But is it really true that birds could not fly without wings? Birds, being as they are, cannot fly when their wings are damaged in certain ways. But with a bit of ingenuity we can imagine that the ancestors of birds could have come to fly without evolving wings—by evolving sails, or parachutes, or balloons, or rockets or some such. These alternatives are fantastic, but fantasy is all that it takes if our only task is to undermine the incredible and overly strong claim that wings are necessary for flight—that there is no way to fly without wings.

Less fancifully, the purported example assumes that "winged" is a trait. But birds and insects and bats each have a specific kind of wing, as does each kind of bird. Once we notice this variation, is it particularly plausible that it would be impossible for sparrows to fly if they didn't have the exact wings that they do? After all, they could have wings of a different sort—perhaps even wings more like those of bats or insects than those of other birds. Of course if "wing" is just a stand-in for whatever produces lift to allow birds to fly, then the argument looks sound. But then we have only the empirically empty claim that birds cannot fly without some flight-enabling structure.

Question Q3a arises in the context of trying to establish some theory of the nature of consciousness. If on theory T consciousness is necessary for the capacity to ϕ and we are ϕ-ers then we are entitled to conclude that consciousness$_T$ (consciousness as explained by theory T) is what enables us to ϕ, and thereby entitled to conclude that T is the correct theory of consciousness. (This line of reasoning is usually paired with the negative argument discussed in

Q3b, below, to the effect that no other theory of consciousness can explain why consciousness is necessary.) But I do not see that we have reason to suppose that there is any ϕ that is necessary for any capacity of biological organisms in the strong sense that would be required to infer the presence of ϕ from the presence of the capacity.

Q3b. If consciousness is not necessary for some capacity ϕ in creature C, does that show that it is not causally potent with respect to the ϕ-ing of C? One reason that many theorists seem to think that consciousness must be necessary for some capacity ϕ or other is the fear that consciousness will otherwise prove to be epiphenomenal (Polger and Flanagan 2002). If there is no ϕ for which consciousness is necessary, then we don't know what consciousness does (or why nature would contrive to provide us with consciousness—see Q4b, below), and we should conclude that consciousness is epiphenomenal after all. But this line of reasoning is fallacious. Carburetors are not necessary for mixing air and fuel in combustion engines (the job can be done by fuel injectors, among other devices), but it does not follow that carburetors do not mix air and fuel in those vehicles that have them. Bird wings are not necessary for flight (rockets, helicopters, and insects can all fly), but it does not follow that bird wings are causally impotent with respect to flight. Four chambered hearts are not necessary for circulation, but it does not follow that some of the chambers of human hearts are epiphenomenal. The argument form that moves from inessentialism to epiphenomenalism is clearly invalid (Flanagan 1992, Polger and Flanagan 2002). It is hard to understand why it seems to be so attractive to so many thinkers, yet it appears over and over.

It may be useful to notice that reasoning from conscious inessentialism to epiphenomenalism is not mistaken only in the difficult case of consciousness. In general, from

the fact that *x* is not necessarily *P* it does not follow that *x* is not *P*. The argument isn't even tempting in its simple forms. Consider: Sally's car is not necessarily silver, therefore Sally's car is not silver. But for some reason this argument form has proven unusually alluring for those thinking about the evolution of consciousness. If some theory T asserts that consciousness gives us some capacity ϕ, then the opponent objects by telling a just-so story (T*) about how ϕ can be had without consciousness or without consciousness being implemented in the way that theory T supposes. The availability of the just-so story is taken to show that consciousness does not do ϕ, for a creature without consciousness$_T$—a zombie—could do ϕ. Since T says that consciousness *does* ϕ, we are urged to conclude that T is false. But the line of reasoning from "does not necessarily" to "does not" is invalid.

Notice that the emphasis in the inessentialist reasoning suggested by Q3b is on positing an alternative theory, T*, to explain ϕ. Offering an alternative explanation is quite different from showing that T is false by experimentally showing that mechanism M invoked by T can be interfered with without disrupting ϕ—experimentally dissociating M (hypothesized by T) from ϕ. The former aims to show that M is not necessary for ϕ, that it is inessential. The latter aims to show that M is insufficient for ϕ. This illustrates the difference between merely possible dissociations and actual deficit studies. Seen in this light, what is posing as an "evolutionary" argument against a theory of consciousness is revealed to be simply a skeptical argument: Because it is possible that theory T is not correct, it is concluded that T is false.

Q4a. If consciousness is necessary for some capacity ϕ in creature C, does that show that it has the evolutionary function of ϕ-ing in C? If consciousness is necessary for capacity ϕ in creature C, and C is a ϕ-er, then consciousness *is causally effective* in the ϕ-ing of C. This was the answer

to Q3a, though I expressed my doubt that there is any such ϕ. Now we are asking whether, if consciousness is necessary for ϕ in C, and C is a ϕ-er, then we can conclude that consciousness *was selected by natural selection* for (i.e., given the evolutionary function of) ϕ-ing in C. This stronger claim is too strong. But there is a related claim that is quite reasonable: Suppose that there are some features of creatures which are in fact necessary for some activities of those creatures. Again, I doubt this occurs; but let us pretend that bird wings are in some sense necessary for flight in birds. If so, then this is strong evidence that the trait in question was selected for by natural selection. However, the evidence is defeasible, and it could turn out that the trait was not selected for the capacity to ϕ, and so does not have the function of ϕ-ing. A trivial example is having mass, which is necessary for many terrestrial activities but was not selected for by natural selection. In fact, in such cases of trivial and universal features like having mass, their necessity even suggests that they were not selected for. After all, mass is had by all creatures. Science fiction aside, there were no massless creatures relative to which the massed creatures could have selective advantage. There was no opportunity for selection for "having mass." (It is doubtful that "having mass" is a biological trait at all. That is another reason for doubting that "having mass" has a biological function.)

A less trivial but still silly example is the ability to do calculus. Whereas we may suppose that various brain structures are necessary (in some sense) for our ability to do calculus, it does not follow that those structures have the evolutionary function of permitting us to do calculus. It may be that those structures came about for other reasons, and were co-opted for doing calculus. The point here is that not every ability ϕ is one that is selectively relevant for a particular creature in a particular environment. If consciousness is necessary for some ability that did not make a fitness difference in its selective environment, then it will not have the function of ϕ-ing.

Evolution and natural selection produce contingent features in the world. We do not need evolution to explain necessary features of organisms. We need evolution precisely to explain those features that are not necessary, e.g., particular size, or the presence of eyes.

Q4b. If consciousness is not necessary for some capacity ϕ in creature C, does that show that it does not have the evolutionary function of ϕ-ing in C? The fact that a trait is necessary for some ϕ does not entail that it was selected for ϕ. But if it is not necessary for ϕ-ing, then does that show that it was not selected for that ability? Of course not. As I have emphasized above, evolution is an engine of contingency. It takes in contingencies, and spits out contingencies. Human beings have opposable thumbs, which come in quite handy. We are able to do many things with our opposable thumbs. Opposable thumbs are not necessary. They are a contingent feature, but one which evidently put some of our ancestors at a selective advantage over their peers. I don't know just exactly how to explain what opposable thumbs have the function of doing; that is, I don't know for exactly which capacity of the capacities that they enable they were selected by natural selection. But there is good reason to think that they do have some such function or functions, that opposable thumbs are adaptations (Gould 1980).

Unfortunately, like the bad reasoning explained in Q3b, the line of reasoning in Q4b has tempted many theorists to despair that an adaptation explanation for consciousness can be found if consciousness is not necessary for some capacity or other. These theorists are generally resistant to my claim that there is no capacity ϕ for which consciousness or wings are strictly necessary. Sometimes that is because they are taken in by the Q3b reasoning, and then wonder whether a causally impotent trait could be an adaptation. (They correctly conclude that it cannot.) Others succumb directly to the fallacious argument from conscious inessentialism to adaptive

irrelevance. Carruthers (2000), who is usually cautious, argues that higher-order perception (he says "experience") theories of consciousness are implausible on these grounds. Carruthers reasons that evolving higher-order perceptions requires that we already have higher-order thoughts. But once we have higher-order thoughts we do not need higher-order perceptions, they are inessential; so we ought to reject the higher-order perception theory. That is, since higher-order perceptions are not necessary, the implication is that evolution is unlikely to have provided us with them. (There is an alternative reading of this argument, on which it claims not that higher-order perception is inessential but that it is redundant. I maintain that redundancy arguments presuppose inessentialist reasoning. See Polger 2004, chapter 6.)

Now it is true that Carruthers stops at the claim that evolving unnecessary traits is unlikely, and does not go so far as to claim that it is impossible. But even the likelihood conclusion is unwarranted. From the fact that we can tell a just-so story about how a creature could do without some trait, nothing at all follows about what the trait actually does (its efficacy, per Q3b), about its history (whether it has an evolutionary function, per Q4b), or about the likelihood of its occurrence.

4. Consciousness and the Complexity Argument

The problematic lines of reasoning discussed in the previous section run into trouble for two general reasons. One is that some mistakenly try to draw conclusions about the actual state of affairs based solely on considerations about what states of affairs are or are not necessary. Another is that they try to make reasoning about the natural world into a deductive enterprise. They ask whether some facts about consciousness or evolution entail others, rather than asking what kinds of evidence we have for claims about consciousness. This is why even the positive

results are not very interesting, e.g., that if bird wings are (in some sense) necessary for flight in birds, then bird wings are causally efficacious in actual bird flight.

There is at least one line of evolutionary reasoning that avoids these pitfalls. As noted in passing above, evolution by natural selection is the most likely source of complex traits in living creatures. When we find a complex trait in a living thing we can reasonably expect that the trait was formed by natural selection. The connection is defeasible, of course. Complexity may sometimes come about and be maintained by chance alone. But as a line of reasoning about empirical contingencies, the complexity argument is a good one. Grantham and Nichols (2000) have done the most to defend the application of complexity considerations to the evolution of consciousness.

Grantham and Nichols begin with the general principle of evolutionary reasoning and apply it to the case of consciousness: "According to contemporary evolutionary biology, it is reasonable to assume that complex biological structures are adaptations—even if we do not know precisely how the organ functions or how it evolved. The complexity of phenomenal consciousness thus provides an argument that phenomenal consciousness is an adaptation" (2000: 649). The burden, then, is to argue that consciousness is complex in the appropriate way. Grantham and Nichols proceed cautiously because they are concerned to keep at bay the critic who adopts a skeptical or epiphenomenalist critique. Working under these constraints, they outline evidence for thinking that some systems implicated in conscious perception are anatomically complex. (They are unwilling to accept evidence of merely functional complexity because it is more vulnerable to epiphenomenalist concerns.) Their conclusion is that "if given an abstract characterization of the structure of phenomenal consciousness, biologists wouldn't

even entertain the hypothesis that the system is functionless" (2000: 664). Thus anatomical complexity is evidence of adaptation.

I don't think that Grantham and Nichols' gambit of relying on structural complexity works out, but it is not one that they ought to require anyhow. They recognize that their success must be qualified:

> For those with an abiding metaphysical conviction that phenomenal consciousness can't be causally relevant, [the] complexity argument is unlikely to carry much weight. However, if we view phenomenal consciousness from the perspective of biology rather than metaphysics, we have good reason to think that phenomenal consciousness is an evolutionary adaptation and hence causally relevant. (Grantham and Nichols 2000: 664).

The troubles are twofold. First, the epiphenomenalist skeptic will not be satisfied by restricting one's concerns to only the anatomical complexity of consciousness, even if that is successful. For such a critic will be willing to be skeptic about those systems as well. Second, it is unclear how the anatomy of consciousness can be located without any appeal to evidence of functional organization. (Nor do Granatham and Nichols suppose that it can, entirely.) The mapping of the functional and phenomenal structures of experience onto anatomical structures in the nervous system is part of the argument for identifying those neural structures as the locus of consciousness (Polger and Flanagan 1999, Polger and Sufka 2006). Without that mapping we cannot be sure that we are considering the right anatomical features.

The lesson is that one should not try to fend off the skeptical epiphenomenalist and provide a positive theory of the evolution of consciousness at the same time. And if I am right, there is no need to pursue these goals simultaneously. For the main arguments appealed to by skeptical epiphenomenalists are those considered in §3 above, which reason from the fact that

consciousness is not causally or evolutionarily necessary for some or any ϕ to the conclusion that consciousness does not do ϕ. We have seen that these arguments are invalid, so we can safely set aside these kinds of epiphenomenalist worries when it comes to giving an account of the natural history of consciousness.

It is important that we are setting aside are the inessentialist-based epiphenomenal concerns, epitomized by the reasoning discussed with respect to Q3b and Q4b. If there are other reasons to consider epiphenomenalism about consciousness, then those will have to be settled. Some will think that the timing studies discussed by Libet 1993, Wegner 2002, or Gray 2004 give such reasons. (But for an alternative interpretation of these experiments, see Nahmias 20002.) If we are independently convinced that consciousness is epiphenomenal then the complexity argument will cut no ice, for epiphenomenalists will be prepared to think of consciousness as a mere by product of that complexity. Of course, as epiphenomenalists, they will also eschew any adaptationist explanation for the features of consciousness, per Q1b and Q2b, and our interest in the evolution of consciousness will be rather limited, per Q1a and Q2a. At best one would be able to say that an epiphenomenalist theory of consciousness is not incompatible with the evolution of the systems with which consciousness is associated. And, of course, epiphenomenalists will not be able to explain the complexity of conscious experience in terms of the complexity of the systems on which (they may agree) consciousness depends.

If we're not worried about epiphenomenalism, the evidence of the complexity of conscious experience is obvious and abundant. My own favorite example is the rich phenomenal and functional structure of color vision. Human color experience varies along the dimensions of hue, saturation, and brightness; these factors interact to yield a distinctive asymmetric color space that appears to be different from the perceptual spaces of other species, that is well

explained by the anatomical organization of the visual system (and by differences between our anatomy and, say, pigeon anatomy), and that is well-tuned for guiding activity within the constraints of our spectral environmental (Hardin 1988, Thompson 1995, Purves, et. al., 2003). Once we understand that evolution produces contingencies and that consciousness is part of nature, then visual consciousness evolved if eyes and brains did. The complexity argument gives us reason to think that eyes and visual systems evolved even if we had no other evidence that they did and even if we did not know what they are good for. Of course in the case of conscious visual perception we have a pretty good idea what it does for us, so complexity argument is not our only source of information.

5. Just-So Stories and Beyond

A general pitfall in evolutionary reasoning about consciousness, and about evolutionary psychology broadly, is the use of just-so stories to postulate the existence of neural or psychological mechanisms that could have, should have, or must have evolved. Since evolution is an engine of contingency, this kind of reasoning is likely to go awry. It is simply not the case that evolutionary forces should have or must have produced anything at all. So it is foolhardy to try to reason from evolutionary stories to the existence of physiological structures (Grantham and Nichols 1999).

A better methodology is to instead think about the evolutionary history of features that are known and understood. But this is hard work—this is evolutionary biology. Good theorizing requires a tremendous amount of historical and comparative study, much of which is hard to do with soft tissue systems such as the neural mechanisms that presumably mediate conscious experience. But there is, for example, some elegant comparative work on color vision across

animal species (see Thompson 1995 and Clark 1993 for discussion) that can be used in reasoning about the etiology and functions of color vision.

Once one adopts the view that consciousness is a natural process that occurs in some kind of creatures then there is no philosophical puzzle about how consciousness evolved, just the hard work of evolutionary biology. Eyes have always been central to the discussion of human evolution. No scientist now doubts that our eyes and brains are products of evolution by natural selection. None doubt that brain areas V1-V5 are implicated in visual processing, and that their structures are products of natural selection. Activation in visual cortex is also associated with conscious visual sensations. Understanding exactly how sensations are manifested by brains is a difficult problem indeed, and the object of much scientific and philosophical theorizing. The naturalist holds that whatever the evolutionary explanation of how the visual system came to be how it is, that will be the story of how visual consciousness came to be how it is. It is utterly irrelevant whether the same information gathering capabilities could be achieved by some system that has different conscious experiences or none at all, or whether our visual system could be replaced with a silicon prosthetic. In us, those capabilities are performed by conscious mechanisms.

Consider the case of blindsight (Weiskrantz 1986). Philosophers and cognitive scientists have tended to focus on what you might think of as the silver lining for blindsight patients, which is that they seem to demonstrate that some perceptual information can be processed in the absence of visual sensation. After all, this is the surprising part of the phenomenon. But let us not forget that blindsight is a deficit, and that visual consciousness is lacking because there is damage to the visual system. It's true that blindsighters perform better than chance at certain tasks. But normally sighted persons perform almost perfectly in the same tasks. So at the same

time that blindsight suggests that visual sensation is (in some sense) not necessary for visual information gathering, it also provides evidence that conscious mechanisms—as a matter of contingent fact—play an important role in normal human perception. Of course the evidence is subject to further investigation. If we had actual evidence of double-dissociation between conscious visual experience and visual competence—for example, actual evidence of "super-blindsight" patients who show no performance deficit while reporting lack of visual sensation, rather than the mere philosophical possibility of such (Block 1995)—we would have experimental reason to doubt that consciousness itself is doing some work. But the mere possibility of super-blindsight at best shows that consciousness is inessential, not that it is inefficacious, per Q3b. It gives us no reason to doubt that consciousness is implicated in (rather than merely correlated with) our visual processing.

There are also some deflationary evolutionary explanations, which take what we know about existing brain systems as evidence that some manifestations of conscious experience are not adaptations. Flanagan (1995, 2000) argues that dream consciousness is not an adaptation, but a spandrel. His reasoning does not depend on the claim that brains could do what they do without consciousness, though that might be true. Instead, he argues that the best current theories of dreaming and brain activity during sleep do not invoke a role for conscious visual experience. The best candidates for the function of brain activity during sleep are memory consolidation and memory purging. But experiments show that dream experiences do not have the content that they would be predicted to have if the conscious content of dreams were to play a role. We do not dream about things that our brains are trying to remember, nor about things that we are trying to forget. Instead, the stimulation of conscious experience during sleep appears to be a side effect of those other brainy activities.

Similarly, Sufka (2000) argues that chronic pain sensation does not serve an adaptive function. Sufka assumes that the acute pain system is an adaptation. But, he argues, the neuronal changes that seem to explain chronic pain are part of the basic cellular mechanisms in neurons, not special to the pain sensory system. The cellular changes involved in chronic pain are nearly identical to those thought to be involved in the cellular basis of learning and memory. Sufka speculates that these basic cellular mechanisms are adaptations for learning and memory and that they are quite universal in neurons. Chronic pain, then, is the byproduct of two systems that are adaptations, the pain sensory system and the cellular learning mechanisms. The result is that the pain sensory system can, in effect, learn to be in pain. Something like this account may apply to some mood disorders like depression and anxiety, as well.

These evolutionary explanations of the experiences of dreaming and chronic pain are deflationary in that they deny that some kind of consciousness is an adaptation. And in each case it should be conceded that little evidence is provided to support the claim that consciousness is a spandrel of some other trait that is an adaptation. Still, these accounts are at least off on the right foot because they begin with empirical consideration of known neural mechanisms. These deflationary theories may not be borne out in the long run. But if so it will be because they do not stand up to the evidence, not because someone has an account according to which consciousness is necessary for any familiar or heretofore unnoticed capacity of human beings. Though there are many potential flaws in these accounts, they are the typically flaws of empirical theories. They are subject to experimental disconfirmation. But they avoid the pitfall of relying on claims of evolutionary necessity.

6. Conclusion

Clinical cases like blindsight lend credibility to philosophical intuitions that consciousness is not (in a sense) necessary for vision. But the empirical cases also suggest that consciousness is crucial to the ordinary operation of human cognitive and perceptual systems. The lesson, I have urged, is that it is a mistake to think about consciousness—and especially about the evolution of consciousness—in terms of necessity or lack thereof.

If birds were aware that their wings are what enables them to fly away from predators, they would be right to think that having wings was awfully important. One might even say that having wings is essential to birds being the kinds of creatures that they are. But that does not show that wings had to evolve to "solve" some evolutionary challenge in the ancestors of birds. Evolution might have pushed the bird ancestors in a different direction, making them fast runners or whatnot. This doesn't show that wings are not for flight, it just shows that wings never had to come into existence at all.

We humans are conscious creatures. We are aware that and appreciate that we are conscious creatures. We value our consciousness, for among other reasons we think that we could not be the kinds of creatures we are without being conscious. In this sense we regard consciousness as necessary for and essential to our form of life. All this is true, but it does not show that consciousness is necessary for any particular capacity that we have. Consciousness may, of course, be necessary for our way of doing things. But that will not show that consciousness had to occur unless it is also necessary that we evolve to be as we are—which surely it is not. The sense in which consciousness is necessary for us is quite a contingent sort of necessity, but that is the only kind that evolution provides.

None of these considerations undermines the claim that we are conscious beings, that consciousness plays a role in our psychology, or that consciousness has evolved. But saying more about the nature of consciousness, what it does, and where it came from will require hard empirical work, not more "just-so" stories.

7. Acknowledgements

This paper is a development of earlier ideas that were published with Owen Flanagan. I benefited from discussion of this chapter with the Cognition, Action, Perception and Performance group in the Psychology Department of the University of Cincinnati. Colin Allen, Eddy Nahmias, Shaun Nichols, Michael Riley, Susan Schneider, Robert Skipper, and Max Velmans commented on versions of the manuscript. Of course any errors are my own.

8. References

Allen, C. 2002. Real traits, real functions? In Ariew, A., Cummins, R. & Perlman, M. (eds.) *Functions: New Readings in the Philosophy of Psychology and Biology*. New York: Oxford University Press, 373-389.

Block, N. 1995. On a confusion about a function of consciousness. *Behavioral and Brain Sciences* 18 (2): 227-287.

Brandon, R. 1990. *Adaptation and Environment*. Princeton: Princeton University Press.

Bringsjord, S. and R. Noel. 2002. Why did evolution engineer consciousness? In *Consciousness Evolving*, James Fetzer (Ed). Amsterdam: John Benjamins.

Carruthers, P. 2000. *Phenomenal Consciousness: A Naturalistic theory*. New York: Cambridge University Press.

Carruthers, P. 2000. The evolution of consciousness. In P.Carruthers and A.Chamberlain (eds.), *Evolution and the Human Mind*. Cambridge: Cambridge University Press, 254-275.

Carruthers, P. 2004. Suffering without subjectivity. *Philosophical Studies* 121: 99-125.

Clark, Austen. 1993. *Sensory Qualities*. Oxford: Clarendon/Oxford University Press.

Dennett, D. 1991. *Consciousness Explained*. Boston: Little, Brown, and Co.

Dretske, F. 1995. *Naturalizing the Mind*. Cambridge, MA: MIT Press.

Flanagan, O. 1992. *Consciousness Reconsidered*. Cambridge, MA: MIT Press.

Flanagan, O. 1995. Deconstructing dreams: The spandrels of sleep". *Journal of Philosophy*, 5-27.

Flanagan, O. 2000. *Dreaming Souls*. New York: Oxford University Press.

Fodor, J. 1968. *Psychological Explanation*. New York: Random House.

Gould, S. J. 1980. *The Panda's Thumb: More Reflections in Natural History*. New York: Norton.

Gould, S.J. and E. S. Vrba. 1982. Exaptation: A missing term in the science of form. *Paleobiology* 8: 4-15.

Grantham, T. and S. Nichols. 1999. Evolutionary psychology: Ultimate explanations and Panglossian predictions. In *Where Biology Meets Psychology: Philosophical Essays,* Valerie Hardcastle (Ed). Cambridge, MA: MIT Press.

Gray, J. 2004. *Consciousness: Creeping up on the Hard Problem*. Oxford: Oxford University Press.

Hardin, C. 1988. *Color for Philosophers*. Indianapolis, IN: Hackett.

Humphreys, N. 2002. *The Mind Made Flesh: Essays from the Frontiers of Evolution and Psychology*. Oxford, UK: Oxford University Press.

Libet, B. 1993. *Neurophysiology of consciousness: Selected papers and new essays*. Boston: Birkhäuser.

Lycan, W. 1987. *Consciousness*. Cambridge, MA: MIT Press.

Millikan, R. 1984. *Language, Thought, and Other Biological Categories*. Cambridge MA: MIT Press.

Millikan, R. 1993. *White Queen Psychology and Other Essays for Alice*. Cambridge, MA: MIT Press.

Moody, T. 1994. Conversations with zombies. *Journal of Consciousness Studies*, 1, No. 2: 196-200.

Nahmias, E. 2002. When consciousness matters: A critical review of Daniel Wegner's *The Illusion of Conscious Will*. *Philosophical Psychology* 15(4): 527-541.

Nichols, S. and T. Grantham. 2000. Adaptive complexity and phenomenal consciousness. *Philosophy of Science*, 67: 648-670.

Polger, T. 2004. *Natural Minds*. Cambridge, MA: MIT Press.

Polger, T. and K. Sufka. 2006. Closing the gap on pain: Mechanism, theory, and fit. In *New Essays on the Nature of Pain and the Methodology of its Study*, M. Aydede (ed.). Cambridge, MA: MIT Press.

Polger, T. and O. Flanagan. 1999. Natural Answers to Natural Questions. In *Where Biology Meets Psychology: Philosophical Essays,* Valerie Hardcastle (Ed). Cambridge, MA: MIT Press.

Polger, T. and O. Flanagan. 2002. Consciousness, Adaptation and Epiphenomenalism. In *Consciousness Evolving*, James Fetzer (Ed). Amsterdam: John Benjamins.

Purves, D. and R. B. Lotto. 2003. *Why we see what we do*. Sunderland, MA: Sinauer.

Sufka, K. 2000. Chronic pain explained. *Brain and Mind* 1: 155–179.

Thompson, E. 1995. *Colour: A Study in Cognitive Science and the Philosophy of Perception.* Boston: Blackwell Publishers.

Van Gulick, R. 1980. Functionalism, information, and content. *Nature and System* 2: 139-162.

Velmans, M. 2000. *Understanding Consciousness*. London: Routledge.

Wegner, D. M. 2002. *The Illusion of Conscious Will*. Cambridge, MA: MIT Press.

Weiskrantz, L. 1986 *Blindsight: A Case Study and Implications*. New York: Oxford University Press.