EVOLUTION AND EPIPHENOMENALISM

William S. Robinson
Iowa State University May, 2006

Abstract: This paper addresses the question whether accepting evolutionary principles is compatible with accepting epiphenomenalism, and argues for an affirmative answer. A general summary of epiphenomenalism is provided, along with certain specifications relevant to the issues of this paper. The central argument against compatability is stated and rebutted. A specially powerful version of the argument, due to William James (1890) is stated. The apparent power of this argument is explained as resulting from a problem about our understanding of pleasure and an equivocation on "explanation". Finally, an argument by Plantinga (2004), which applies to cognition rather than phenomenal qualities, is stated and rebutted.

Epiphenomenalism is regarded as anathema for a variety of reasons. Among these, one of the more widely cited is the alleged incompatibility of epiphenomenalism with evolutionary considerations. In this paper, I will examine several arguments for this incompatibility and I will argue that appeals to evolution do not provide good reasons to reject epiphenomenalism.

1. Epiphenomenalism

Epiphenomenalism is the view that

(E1) Mental events are inefficacious.[1]

Some readers may find "mental state" to be a more natural phrase than "mental event". When mental states are thought to be efficacious, however, they are thought to be causes of particular events (which may be conceived in mental, behavioral, or neural terms). Since the alleged effects are events, what is needed for their causes are events and if states are said to be efficacious, that can only be because they are "activated", or "tokened" on some particular occasion. Such activations or tokenings are events.

A fuller understanding of (E1) depends on recognizing it as a conjunction of two claims – a claim about qualitative experiences and a claim about such events as occurrently believing or desiring something. The first case is the more familiar one and can be stated this way:

(E2) Qualitative experiences are inefficacious.

I shall assume a background that is usual in discussions of (E2). Thus, I shall take it that, in a given subject, S, each qualitative experience, Q, is caused by some neural event, N. Besides causing the experience, N causes neural events and it is these neural effects of N

that cause the changes in belief or desire, or the behavioral events, that are commonly (but mistakenly, according to epiphenomenalism) regarded as effects of Q. A further piece of background is that, in general, the neural effects of N would not be able to be brought about by any other kind of neural event in S. On this assumption, the counterfactual "S would not have behaved in manner B at *t* if Q had not occurred in S" will be true – not because Q caused neural events that led to behavior B, but because Q is a co-effect of the only neural events that would have been able to cause B in S at *t*.

Assumptions about occurrent believing and desiring are less familiar and less agreed upon than assumptions about qualitative events. I shall introduce two ways of thinking about such occurrents and will follow out the arguments for each. The first is the Assertion Theory.

(AT) To have an occurrent belief that *p* is to subvocally assert that *p* in absence of dispositions to correct or withdraw one's assertion.

To have an occurrent desire that *p* is to subvocally assert that one wants it to be the case that *p* in absence of dispositions to correct or withdraw one's assertion.

The second part of (AT) is to be understood to include assertions that actually have the form "I want an *F*" (e.g., "I want a steak") and "I want to *VP*" (e.g., "I want to go to the park") instead of the more awkward, but always available, "I want it to be the case that *p*" (e.g., "I want it to be the case that I have a steak to eat", "I want it to be the case that I go to the park").

To avoid tedium, I shall henceforth use "assertion that *p*", "occurrent thought that *p*" and similar phrases with the understanding that, where the thoughts are occurrent desires, they are to be disabbreviated to "assertion that I want it to be the case that *p*", "occurrent desire that I want it to be the case that *p*", and so on. And I shall often write only about occurrent beliefs, with the understanding that parallel remarks are to be made about occurrent desires.

The other way of thinking about occurrent beliefs and desires regards them as items that are additional to subvocal assertions, but are capable of being adequately "expressed" by subvocal assertions. I shall call this the Expression of Match Theory.

(EMT) To have an occurrent thought that *p* is to have a mental event that matches a possible subvocal assertion that *p*, where "matching" means having a content that would be adequately expressed by a subvocal assertion that *p*.

(EMT) is to be understood to allow that the matching events may be either conscious or unconscious.

Overt utterances that assert claims or desires are occurrents and they plainly have effects on listeners and, in virtue of the fact that one can hear what one says, on speakers themselves. Moreover, if one overtly asserts that *p* with no insincerity and no disposition to correct or with draw one's statement, one normally believes that *p*. Nonetheless, it will

be taken as an assumption of this paper that overt utterances are not themselves occurrent beliefs or desires. This stipulation, taken by itself, leaves open three ways of thinking of overt assertions of beliefs or desires.

(OA1) Overt assertions of beliefs or desires repeat out loud previous subvocal assertions – i.e., they give voice to occurrent beliefs and desires as conceived according to (AT).

(OA2) Overt assertions of beliefs or desires are effects of occurrent beliefs and desires as conceived according to (EMT).

(OA3) Overt assertions of beliefs or desires *may* have the same contents as preceding subvocal assertions but they need not be preceded by an occurrent thought, whether occurrent thoughts are conceived of according to (AT) or according to (EMT); and, in any case, they are not caused by occurrent thoughts, on either conception.

In the interests of full disclosure, I note that I prefer (AT) and regard the alleged matching entities postulated by (EMT) as both unnecessary and problematic. I have discussed belief and desire elsewhere, however, and since I wish to make the defense of epiphenomenalism against objections from evolution as widely acceptable as possible I will consider both (AT) and (EMT) in the following discussion.[2] In contrast, I take (OA3) to be part of what epiphenomenalism, as understood in this paper, to be committed to.

With these understandings in place, we can formulate the second part of the epiphenomenalist view as follows.

(E3) Occurrent beliefs and desires are inefficacious.

The reason it will be possible to consider two views of occurrent beliefs and desires is that I am not attempting to argue for epiphenomenalism in this paper. My question here is only whether considerations derived from evolutionary theory undercut epiphenomenalism in ways that they are often supposed to do.

(AT) and (EMT) by themselves are accounts of occurrent beliefs and desires that are neutral with respect to epiphenomenalism. It is important to have a positive understanding of the view that results from combining them with (E3). Let us begin with epiphenomenalists who accept (AT). They will say that our brains have a complex organization that, together with current stimuli, produces the neural causes of the sequence of experiences that constitute a subvocal saying. They will not hold that the subvocal sayings are causes of behavior that we might loosely describe as “acting on what one has said”. Instead, they will hold that the same brain organization that produces the subvocal sayings also produces other effects, among them those that lead to behavior that “fits” with what was said or, in appropriate cases, overt utterances with the same content as the subvocal sayings.[3]

(E3) together with (EMT) yields a parallel account. Holders of this view will say that our

organized brains produce mental events whose content matches a possible assertion; but these mental events are not causes of corresponding assertions (if and when they are made, either subvocally or overtly) or of other behavior that fits the content of those events. Instead, our organized brains, together with current stimuli, produce both the occurrent beliefs and desires and the behavior that fits them.

Interactionist dualists can consistently doubt that our brains are adequate to the organizational tasks that (E3) together with (AT) or (EMT) requires. Physicalists, however, should not have such doubts, since even if they incline toward (EMT), they will hold that brains are the producers of occurrent beliefs and desires and are therefore adequate to ensuring the required organization among them.

If we confine ourselves to consideration of physicalism and epiphenomenalism, we can find some evolutionary reason to prefer the latter. This point is easiest to see if (E3) is combined with (AT), so let us take that view first. Our non-linguistic forebears possessed some ability to organize complex and sequenced actions such as those involved in hunting or social interactions. Evolutionary theory would lead us to expect that this ability would have been refined and not lost in human evolution.[4] It would, moreover, be inefficient even in a linguistic animal for the organization of behavior to have to wait upon the formation of a sentence (overt or subvocal). It is not clear how the events involved in a saying would be converted into neural impulses that would be required to carry out actions that would be appropriate in the light of what was said. These reflections make it plausible that evolution would have led to a system in which the brain that organizes speech organizes related behavior concurrently with its organization of speech, i.e., without waiting for the speech to become fully formed.

Once we see this point, it is easy to extend it to (EMT). The occurrents that EMT proposes are events that have a content that can adequately be expressed in a sentence. No sentence expresses the complexity of neural organization that would be necessary to produce an action that is appropriate in the light of the content of that sentence. So it is somewhat mysterious how an occurrent belief or desire could causally contribute to associated behavior. Again, it seems inefficient, and therefore costly from an evolutionary point of view, to wait upon the formation of a thought before the organization of associated behavior begins. The more plausible picture is that the fit between our occurrent beliefs and desires, and our behavior, is a result of a process in which our brains (under current stimulation) organize both our occurrent beliefs and desires and our behavior in parallel.

These reflections are not claimed to be conclusive, but only to show that there is some reason to think that an epiphenomenalist outlook accords well with evolutionary theory in the area of occurrent beliefs and desires. They are, as noted, directed to physicalists; they should not move anyone who thinks that, in any case, the brain is inadequate to account for our thinking and that an immaterial mind must be brought in to explain our cognitive abilities. However, we may observe that many of those who want to appeal to evolutionary theory in criticism of epiphenomenalism do not intend their criticism to lead to any such interactionist dualism.

Even my limited conclusion may be thought to claim too much, on the ground that epiphenomenalism with respect to occurrent beliefs and desires (whether conceived

according to (AT) or according to (EMT)) would imply that such occurrents are useless, and therefore could never have developed if evolutionary theory is true. But formation of subvocal speech can be regarded as an inhibited form of what is obviously useful, namely, overt speech. And occurrents conceived according to (EMT) can likewise be regarded as preparations for obviously useful communicative activities. Given the value of cooperative activity, it is not surprising, from an evolutionary point of view, that a good deal of what our brain produces is closely related to communicative actions. All that the reflections of the last few paragraphs imply is that there are reasons to doubt that evolution would have led to a condition in which these communication-related products should be steps in a causal series leading to behavior, as opposed to co-effects of a brain organization that leads, in parallel fashion, to both occurrent beliefs and desires and to behavior that is, by and large, appropriate in the light of the contents of those communicative preparations.

2. The Argument from Evolution

I will begin this section by stating the key argument from evolution in a general form, that is, one that applies to both qualitative events and to believings and desirings. This argument is most plausible, however, for the case of qualitative events and so I shall divide the discussion of it. In this section and the next, I will assume that the mental events under consideration are qualitative events (or, experiences), and I will assume, in accord with traditional versions of epiphenomenalism, that qualitative events are not identical with neural events. In section 4, I will take up a very recent version of essentially the same argument that is explicitly directed upon believings and desirings.

The argument against epiphenomenalism from evolution goes back to the late nineteenth century and is familiar in recent philosophy from Popper and Eccles (1977). It can be stated this way.

(AE1) Features that have become standard in beings that have developed under processes of natural selection must have an effect on behavior.

(AE2) The ability to have mental events of the kinds we have is a standard ability in beings (namely ourselves) that have developed under processes of natural selection.

(AE3) Our ability to have mental events of the kinds we have can have an effect on behavior only if the mental events that that ability permits have effects on behavior.

So,

(AE4) Mental events have effects on behavior.

There is a reply to this argument that has been given often enough to be regarded as the standard reply.[5] It is, that (AE1) is false. The closest we can come to (AE1) is

(AE1*) Features that have become standard in beings that have developed under processes of natural selection must either have an effect on behavior or be nomologically connected to something that has an effect on behavior.

If we revise the other statements in the argument so as to ensure validity, the conclusion becomes

(AE4*) Mental events either have effects on behavior or are nomologically connected to events that have such effects.

Since epiphenomenalism agrees to (AE4*), the revised argument is no longer an objection to it.

This reply is fully correct, but it is apt not to be found satisfying. A likely reason for the dissatisfaction with the standard reply to (AE)is that there is nothing in the reply – or in any other account we have – that explains why there should be a nomological connection between neural events and qualitative events. Thus, evolutionary considerations can have no part in explaining why we have qualitative events at all. This consequence makes the case with qualitative events worse off than familiar cases of properties that are not selected for. For example, it is the insulating properties of fur, not is weight, that explains why animals in cold climates have fur. We can, however, give an explanation of the heaviness of their coats: Insulation depends on thickness, and greater thickness entails greater weight. Now, we cannot provide an analogous explanation of why we have qualitative events. So, we remain unsatisfied with the standard reply to (AE).

We can turn this dissatisfaction into an apparent objection by saying that epiphenomenalism can achieve consonance with evolutionary considerations only by adding a claim that neither it nor evolution can explain – namely, the claim that neural events cause qualitative events. But it is easy to see that parallel objections apply to physicalism and to interactionist dualism. Let us consider these views one at a time.

If physicalists claim that qualitative events cause neural events, that is only because they have already accepted the claim that qualitative events are identical with neural events. Many of those who accept this further claim agree that we do not understand how it could be true, and no one to my knowledge claims that evolutionary considerations explain how it could be true. It is therefore fair to say that physicalism can achieve consonance with evolutionary considerations only by adding a claim that neither it nor evolution can explain – namely, the claim that qualitative events are identical with neural events.

Some physicalists may reply that no explanation for such identity needs to be given because none can be given. Identities themselves can never be explained – what can be explained is only how we come to know that an item presented in a certain way is the same as an item presented in a different way. But this reply is inadequate because particular identities are intelligible only against a background that explains their possibility, and no such background is available in the present case.[6] To illustrate: There is indeed no explanation of why Twain is Clemens; but we understand a place for such an identity claim (independently of knowing its truth value) because we understand how a person's body can move in space and acquire different labels on different occasions. But

we do not understand how a qualitative property (e.g., phenomenal redness, the taste of oregano, painfulness, etc.) could be the same property as the property of being a set of neurons instantiating a certain set of firing rates.[7] Correlation does not explain it because correlation is compatible with non-identity. Sameness of causal role cannot explain it because the claim of sameness of causal role already assumes that qualitative properties have a causal role. Notice that this point is distinct from the complaint that assuming a causal role for qualitative properties would beg the question against epiphenomenalism. That is true; but the point here is that one cannot make the possibility of X intelligible if one has to assume that X is true in order to give the "explanation". Since there seems to be no other way of making the possibility of identity of qualitative events and neural events intelligible, the conclusion remains that physicalism must rely on a claim that neither it nor evolution can explain.

It cannot be argued that we must accept the intelligibility of (a) the identity claim, because that is the only way to explain a "connection" between qualitative properties and neural properties. An alternative claim is that (b) there is an explanation of a connection between the two kinds of properties, although we do not know what it is. Of course, (b) is an empty claim.[8] I am not putting it forward here as one that the reader should accept. The point is only that (b) is no more empty than (a), and therefore physicalists are not entitled to claim that their view is acceptable on the ground that theirs is the only way to imagine that the world might be intelligible. One cannot parlay ignorance into sound metaphysics.

If interactionist dualists claim that qualitative events cause neural events, that is only because they have already accepted the claim that nonphysical qualitative events cause neural events. Many of those who accept this further claim agree that we do not understand how it can be true, and no one to my knowledge claims that evolutionary considerations explain how it can be true. It is therefore fair to say that interactionist dualism can achieve consonance with evolutionary considerations only by adding a claim that neither it nor evolution can explain – namely, the claim that nonphysical qualitative events cause neural events.

I conclude that, unless all of these views are held to fail on parallel grounds, it cannot be an objection to epiphenomenalism that it achieves consonance with evolutionary considerations only by relying on a claim that neither it nor evolution can explain.

This point might be more evident than it is, were it not for the possibility of confusion between Biological Evolutionary Theory (BET) and Ideological Evolutionary Theory (IET). These titles summarize the following views.

(BET) The physical constitution of complex organisms is shaped by natural selection operating on variability due to genetic recombination and mutation.

(IET) Evolutionary considerations explain everything that exists now but did not exist one billion years ago.

(BET) is a statement that is overwhelmingly supported by a century and a half of elegant science. (IET), by contrast, is merely an article of faith. Assuming that there were no

qualitative events a billion years ago, we still do not have any explanation of why they exist now – that is merely one way of expressing the familiar explanatory gap.[9] Dissatisfaction in the face of the explanatory gap is understandable, but it is no excuse for positively affirming the pious hope that evolutionary considerations will one day overcome it. No such view follows from the evidence that supports (BET), and premising (IET) in order to argue against epiphenomenalism is simply begging the question.

3. James's Argument

In a variant of the foregoing argument, William James supported efficacy of consciousness by appeal to "the well known fact that pleasures are generally associated with beneficial, pains with detrimental, experiences".[10] I will refer to this fact as the "hedonic/utility match". James says that there are "numerous" exceptions to the general rule and notes the delightfulness of drunkenness to some people as an example. But we do not find cases of people who find burning to be pleasant, or breathing to be noxious.

James's argument may be set out as follows.

(J1) The hedonic/utility match stands in need of explanation.

(J2) The hedonic/utility match has to be accounted for either by a scientific explanation or by appeal to an *a priori* parallelism.

(J3) *A priori* parallelism is no real explanation.[11]

(J4) The only scientific explanation for the hedonic/utility match is evolution.

(J5) To apply evolutionary considerations to explaining the hedonic/utility match, one must suppose that pleasures and pains are efficacious (and efficacious in virtue of their being pleasant or painful).

So,

(J6) Pleasures and pains are efficacious (in virtue of their being pleasant or painful).

Although this argument really does not add anything to the discussion already given, it gives a strong appearance of doing so, and thus it must be carefully addressed. I shall argue that the appearance that it raises a grave difficulty for epiphenomenalism arises from two sources, namely, a general difficulty in our thinking about pleasure, and a specific equivocation concerning "explanation".

To begin to understand the general difficulty about pleasure, let us consider what ensues if we adopt a non-rigidifying, functional-dispositional account, according to which, for any activity or experience, *x*, to be pleasant is for *x* to be something we generally pursue without coercion. (Analogously, for *x* to be unpleasant is for *x* to be something we generally avoid without coercion. To avoid tedium, I shall generally consider only the

pleasure case and assume parallel treatment for displeasure.) It seems unproblematic that evolution should favor organisms for which the beneficial and the usually pursued without coercion should go together and, given our dispositional hypothesis about pleasure, it follows directly that the beneficial and the pleasant should go together.

The point to notice here is that this underwriting of the hedonic/utility match flows from the dispositional account and relies on no assumption about physicalism, epiphenomenalism or interactionism. For this reason, this "explanation" of the hedonic/utility match is as open to any of these theories as it is to any other. One cannot find an objection to epiphenomenalism in the hedonic/utility match if one adopts the dispositional account.

Since this consequence is evident, we can be pretty sure that James was not assuming a dispositional account of pleasure; and, historical considerations aside, it seems likely that many will find the dispositional account unacceptable on grounds independent of any issue about epiphenomenalism. So, let us see what happens if one takes it that *x*'s being pleasant does not simply consist in the fact that one tends to pursue it without coercion. *X*'s being pleasant cannot consist simply in its occurring, since activities and experiences occur that are not pleasant. So, the pleasantness of *x* must involve *x*'s standing in some relation *R* to something besides *x*, where this additional something is not the behavior that generally accompanies *x*. Let us call this additional something, *P*.

The picture that James suggests is that actions taken in circumstances in which they produce beneficial results will usually produce *P*, and *P* will be efficacious – namely, *P* will cause changes in us that tend to increase the likelihood of similar actions in similar circumstances. The epiphenomenalist reply should be evident from the preceding section. It is, that (J5) is false, because we can get the same result from epiphenomenalist assumptions. All we have to do is suppose that *P* is an effect of some neural event, which we may call N(*P*), and that N(*P*) produces the effects on behavioral tendencies that James supposes are effects of *P*.

A likely response from those who are sympathetic to James's argument is that the epiphenomenalist reply falls short, because it does not give an evolutionary account (or any other account) of why N(*P*) causes *P*. In this context, however, this criticism amounts to an equivocation on "explanation". To see this, notice that James's argument likewise makes an assumption about a causal connection – namely, the one that holds between *P* and its alleged effects in us. This connection is not explained by evolution (or in any other way). So, if James's account qualifies as an "explanation" of the hedonic/utility match, it does so despite not explaining the laws that govern the alleged efficacy of *P*; and by parity, the epiphenomenalist account should be counted equally "explanatory". Or, if the epiphenomenalist account is regarded as non-explanatory because it does not explain the causal connection between N(*P*) and *P*, parity leads to the rejection of the Jamesian picture, because it does not explain the causal connection between *P* and its effects in us.

Physicalists may be inclined to welcome the foregoing discussion because it appears to make both interactionism and epiphenomenalism equally difficult to accept. But there are two points that stand in the way of easy acceptance of their view. One parallels a point made in section 2, and in the specific form for the present case it goes as follows. The

pleasantness of *x* must, according to physicalism and our present assumption that rejects the dispositional account of pleasure, consist in *x*'s standing in *R* to *P*. But *P* can only be a neural event – we might as well call it "N(*P*)", except that instead of saying it is the cause of pleasure, a physicalist will say that this neural event *is* the pleasantness of the *x* that stands in *R* to it. But this specific case of identity is puzzling in the same way that the general identity claim is, i.e., we do not know how it could be that *x*'s being pleasant just is its standing in a relation to a neural event.

The second point that makes physicalism difficult in this context is that evolutionary developments are contingent upon particular circumstances. It is thus imaginable, within thoroughly evolutionary principles, that creatures much like ourselves could have developed that had behavioral preferences, motivational systems, and reward systems that are somewhat similar to ours but that do not involve N(*P*). On present assumptions, they would be creatures in which beneficial actions were not generally accompanied by pleasure. So, even on physicalism, there is a sense in which evolution does not explain the hedonic/utility match; only evolutionary principles plus the particular history of our development explains that.

The pressure of this observation may well lead some theorists to say that pleasure is whatever neural state functions so as to raise the probability of behavior that produced it (in circumstances similar to those of its production). But this is to de-rigidify "pleasure"; and if that is the right way to treat "pleasure", then epiphenomenalists should adopt it, with the consequence that their "explanation" will be as good as that of physicalists.

I do not believe that it will be easy to achieve consensus on the matter of exactly how to think about pleasure, and therefore, I do not think we can resolve all the questions that consideration of James's argument raises. We shall have to be content with the thesis, which I think I have adequately defended, that epiphenomenalism is no worse off in the face of matters raised by James's argument than other views are. Further, the difficulties that remain are not due to any conflict between epiphenomenalism and evolution. They are either specific versions of matters that were discussed more generally in section 2, or they are special difficulties introduced by uncertainty about how to think about pleasure (and, analogously, displeasure).

4. Plantinga's Argument

Alvin Plantinga (2004) has advanced an argument that focuses on what he calls "semantic epiphenomenalism" (hereafter, SE). This argument is part of a larger project, namely, an argument that metaphysical naturalism plus evolution (hereafter, N&E) implies that the probability of our having reliable processes of belief formation is low or inscrutable – which, in turn, implies that a belief in N&E is self-defeating.[12] Plantinga says that his treatment of SE is an "essential aspect" of the argument for the claim that N&E imply low or inscrutable probability of our having reliable cognitive processes and the response I will give to his treatment of SE will have an evident effect on the larger project.

Plantinga's argument depends on regarding beliefs as material processes or events (602-603) that have two kinds of properties, (a) neurophysiological properties such as numbers of neurons involved, connection strengths, firing rates, rates of changes of firing rates, and so on, and (b) a content property, i.e., the property of having some proposition as the content of the belief. Plantinga is doubtful that a naturalistic theory of content can be provided, but assumes for the sake of the argument that this problem can be overcome. I shall proceed on the same assumption here and will not argue for a theory of how the content assignment problem is to be solved. I shall, however, refer to some of the ideas that have been used in theories of this kind.

Conceding a solution to the content assignment problem, however, leaves open the key question of how a belief's content can be efficacious. For it would seem that, on a naturalistic view, the neurophysiological properties must be sufficient to cause any behavior that might commonsensically be attributed to a belief. If so, there would seem to be no causal work left for the content property to do, i.e., the content of beliefs would be causally irrelevant to behavior. This conclusion is semantic epiphenomenalism. What is allegedly wrong with the view is this.

> If false belief caused maladaptive action, natural selection could presumably modify belief-producing structures in the direction of greater reliability . . . . But if content does not enter the causal chain that leads to behavior, then of course it will not be the case that a belief causes maladaptive behavior by virtue of its being false, and it will not be the case that a true belief causes the behavior it does by virtue of its being true. And then it is hard to see how natural selection can promote or enhance or reward true belief and penalize false belief. (605)

Since reliability requires truth, inability to select for truth implies inability to select for mechanisms of greater reliability in belief formation. On naturalistic assumptions, therefore, evolution of reliable mechanisms of belief formation would be an "enormous piece of not-to-be-expected serendipity", i.e., a development of extremely low or inscrutable probability.

So Plantinga argues. As I shall now try to show, however, the dire conclusion does not follow from the assumption of semantic epiphenomenalism. Briefly, the reason is that Plantinga has not shown that the conditions for content assignment do not imply reliability of belief formation. But if having content and being reliably formed are not independent, then Plantinga is not entitled to concede success for a naturalistic theory of content assignment while denying a tendency toward reliability. In the remainder of this section, I will explain and justify this brief response. It should be carefully noted that, while I will aim to present some of the ideas of naturalistic theories in a way that makes them appear plausible, the point of the following discussion is not at all to argue for such theories, but only to emphasize the connection between those theories and tendency toward reliability.

The essential idea can be seen by considering a toy model that appeals to covariance.[13] Thus, suppose that some neural event type, N, were to occur in some organism, O, when and only when O's sense organs are stimulated by an object with property *F*. And suppose that O is disposed to have an occurrent belief that an *F* is present only when N occurs. Suppose, finally, that one's psychosemantics holds that a certain complex, C, of

which N is a proper part, is to be assigned the belief content *that an F is present* because the parts of C other than N are related to N in a certain way. That is, in general, a complex of a certain kind that includes an event that strictly covaries with a property is a belief that a thing with that property is present, and C is a complex of that kind that includes N. This set of assumptions has the consequence that beliefs that an *F* is present will be true.

Let us extend this overly simple model one more step. Suppose that C has behavioral consequences in virtue of its neural properties. Then occurrence of C can be selected for. For example, suppose that some changes in O's synapses result in a few exceptions to the covariance of N with presence of *F*s, without reducing the tendency for the rest of C to occur when N occurs. Then O will come to have some false beliefs and some unsuccessful or unnecessary behavior can be expected to ensue. Bad consequences of such behavior may alter O's synaptic strengths in a "corrective" manner. If so, that will be a respect in which O's brain is working well from an evolutionary point of view, and O's probability of surviving to produce more offspring is enhanced. If the bad consequences of the useless behavior fail to change O's synapses in a "corrective" manner, then O's cognitive equipment is not doing its job very well, and O will be somewhat at risk.

As indicated, I am not proposing this simple model as a serious psychosemantics. The point of the model is only to illustrate what is meant by claiming that the matter of content assignment may not be independent of the question of reliability. It is plausible that this interdependence will carry over to much more highly developed and defensible naturalizations of content assignment, and thus that the accessibility of cognitive mechanisms to natural selection will be preserved.

Two complications must be mentioned. First, not all of our beliefs take the form of recognizing the presence of things of perceptible kinds. Second, everyone considers large territories of some others' belief systems to be erroneous – beliefs in phlogiston, contingent identities, and communications from the dead come to mind as examples. How can these facts be made compatible with the view that evolution selects for reliable processes of belief formation? Once again, an adequate reply would be very long and there is no possibility of giving it here. We can, however, see how reliability and content assignment can be interdependent in non-perceptual cases by noting that, very plausibly, a succession of neural events could not be assigned contents corresponding to an inference unless there were a reliable connection between event types taken to correspond to premises and event types taken to correspond to conclusions. Mistakes in logic are, of course, possible, but, very plausibly, no credible theory of content assignment could imply that O believes that [(P v Q) & ~P] → Q and also that O usually concludes ~Q from [(P v Q) & ~P].

Regarding the second point, the existence of many false beliefs held by human beings could not support the view that N&E is self defeating. It would support only the quite correct view that we should be cautious and take great care. What Plantinga needs is the much stronger idea that N&E imply that our processes of belief formation cannot be selected to be self-correcting, even when extensive evidence is available and

consideration of alternatives and review of reasoning is undertaken. But he has not shown that a credible theory of content assignment will not have the consequence that continued acquisition of evidence and reflection upon it affect belief formation in self-correcting ways and that this relationship is itself a consequence of N&E.

One may object that Plantinga is unlikely to have overlooked the possibility of dependence between naturalistic content assignment and reliability. But there is an explanation of his having overlooked it; namely, his conceding at the outset that a naturalistic theory of content may be possible. This early concession made looking into such theories unnecessary; but it is only in their internal structure that the connection between content assignment and reliability can be found.

Whether this explanation is accepted or not, it is clear that Plantinga has not *established* that content assignment and reliability are independent matters. But without the assumption of independence, he is not entitled to his conclusion that semantic epiphenomenalism implies that the probability of reliable belief formation is low or inscrutable. And without this conclusion, he is not entitled to the view that naturalism plus evolution is a self-defeating combination.

5. More about Believing

The response just given to Plantinga's argument relies on views that have been ably defended and are well known to contemporary philosophers of mind. Proponents of these views, however, have generally shared the common phobia against epiphenomenalism. It is thus natural to raise the question whether my response to Plantinga can consistently be given by epiphenomenalists.

A review of the discussion of occurrent beliefs and desires in section 1 will show that this question can be answered affirmatively. The theme of that discussion is that we should not think of such occurrents as causes of behavior that fits them. Instead, we should think of the occurrents and the associated behavior as common effects of an organized brain under current stimulation. This view is compatible with the claim that a naturalistic theory of content can be given. It is compatible with holding that the contents of occurrent beliefs and desires depend on facts about how a brain is organized. It is compatible with there being a regular relation between changes in sensory inputs and changes in brain organization that cause both changes in occurrent beliefs and changes in associated behavior. That is to say, it is compatible with holding that changes in beliefs can track changes in sensory inputs. Thus, whatever contributions to content assignment are made by tracking of worldly states by neural event types will be available to epiphenomenalists. Epiphenomenalism is further compatible with the view that a viable theory of content assignment could not assign a belief that $[(P \vee Q) \& \sim P] \rightarrow Q$ to an organism, O, and at the same time assign many cases of O's conjointly believing both $[(P \vee Q) \& \sim P]$ and $\sim Q$. In short, epiphenomenalism about occurrent beliefs and desires denies a certain picture of causation by those occurrents, but it does not deny efficacy of brain organization in the production of those occurrents and associated behavior. It thus offers no ground for suspicion that selectional pressures should fail to have access to the

shaping of brain organization so as to improve the fit between perceptual evidence, occurrent beliefs and desires, and behavior.

This epiphenomenalistic outlook may be more fully appreciated by considering an example in which linguistic involvement would be either absent or minimal. The knapping of stones requires one hand to hold firmly while the other strikes. The holding hand has to position the target stone correctly in relation to the blow, and there is no point to one blow if it is not part of a sequence of blows from a variety of angles. Our brains are adequate for organizing this relation among parts of our bodies at the same time and our actions over a stretch of time. It is not plausible to suggest that knappers must tell themselves what they are going to do in order to be able to do it, for it is doubtful that even we have the words to express the required arrangements at the level of detail necessary for practical success.[14] For the same reason, it is not plausible to suggest that knappers must precede their activities by mental occurrents that have a content that matches something that we could say. Even if one were to think that some degree of linguistic capability must precede the ability for successful knapping, one would thus still have to attribute considerable organizational ability to the brain that works behind the linguistic scene, so to speak. Epiphenomenalism about occurrent beliefs and desires proposes that we extend our respect for non-linguistic brain organization and credit it with adequacy for producing the successful relation between our perceptions, our needs, and our behavior, including, as a special case of behavior, our communicative activities and the subvocal preparations for those activities.

6. Conclusion

The essential line of epiphenomenalism is that certain cases that are often conceived as mental events causing behavior are better regarded as cases in which the mental events and the behavior are coeffects of neural events. The details of this line look somewhat different for the cases of qualitative events, on the one hand, and beliefs and desires on the other, and both kinds of cases have been considered here. Several arguments have been offered that purport to show that someone who accepts evolution should not be an epiphenomenalist. These have been examined and found wanting. There are, of course, other lines of criticism of epiphenomenalism. These have not been examined in this paper, the conclusion of which is only that the kind of evolutionary theory for which we have good biological reasons is not in conflict with epiphenomenalism.

There is some irony in criticizing epiphenomenalism by appeal to evolution, for the two views share a common theme: namely, that a somewhat circuitous account of our causal situation provides a better understanding than a simpler view that seems more intuitive at the beginning of our reflections.

References

Bechtel, W. and Richardson, R. (1983) “Consciousness and Complexity: Evolutionary Perspectives on the Mind-Body Problem”, *Australasian Journal of Philosophy* 61:378-395.

Broad, C. D. (1925) *The Mind and it Place in Nature* (London: Routledge & Kegan Paul).

Fodor, J. (1987) *Psychosemantics: The Problem of Meaning in the Philosophy of Mind* (Cambridge, MA: MIT Press).

James, W. (1890) *The Principles of Psychology* (H. Holt).

Levine, J. (1983) “Materialism and Qualia: The Explanatory Gap”, *Pacific Philosophical Quarterly* 64:354-361.

Lindahl, B. I. B. (1997) “Consciousness and Biological Evolution”, *Journal of Theoretical Biology* 187:613-629.

Mangan, B. (2001) “Sensation’s ghost: The non-sensory ‘fringe’ of consciousness”, *Psyche*, 10 (1).

Nisbett, R. E. and Wilson, T. D. (1977) “Telling More than We Can Know: Verbal Reports on Mental Processes”, *Psychological Review* 84:231-259.

Plantinga, A. (2004) “Evolution, Epiphenomenalism, Reductionism”, *Philosophy and Phenomenological Research* 68:602-619.

Popper, K. & Eccles, J. (1977) *The Self and Its Brain* (New York: Springer-Verlag).

Povinelli, D. J. (2000) *Folk Physics for Apes: The Chimpanzee’s Theory of How the World Works* (Oxford: Oxford University Press).

Robinson, W. S. (1982) “Causation, Sensations and Knowledge”, *Mind* 91:524-540.

Robinson, W. S. (1986) "Ascription, Intentionality and Understanding", *The Monist* 69:584-97.

Robinson, W. S. (1988) *Brains and People: An Essay in Mentality and its Causal Conditions* (Philadelphia: Temple University Press).

Robinson, W. S. (1990) "States and Beliefs", *Mind* 99:33-51.

Robinson, W. S. (1995) "Mild Realism, Causation, and Folk Psychology", *Philosophical Psychology* 8:167-187.

Robinson, W. S. (1999) “Epiphenomenalism” in *Stanford Encyclopedia of Philosophy*. Available at http://plato.stanford.edu/entries/epiphenomenalism.

Robinson, W. S. (2004a) *Understanding Phenomenal Consciousness* (Cambridge: Cambridge U. P.)

Robinson, W. S. (2004b) "A Few Thoughts Too Many?", in Gennaro, R., ed., *Higher Order Theories of Consciousness* (Amsterdam and Philadelphia: John Benjamins Publishing Co.), pp. 295-313.

Robinson, W. S. (2005) "Thoughts Without Distinctive, Non-Imagistic Phenomenology", *Philosophy and Phenomenological Research*, 70:534-561.

Robinson, W. S. (2006) "Knowing Epiphenomena", *Journal of Consciousness Studies*, 13:85-100.

Robinson, W. S. (submitted) "What is it Like to Like?"

Ryle, G. (1954) *Dilemmas* (Cambridge: Cambridge U. P.).

Van Rooijen, J. (1987) "Interactionism and Evolution: A Critique of Popper", *British Journal for the Philosophy of Science* 38:87-92.

Wynn, T. (2002) "Archaeology and Cognitive Evolution", *Behavioral and Brain Sciences* 25:389-438.

Notes

1. This section presents a summary account of epiphenomenalism, designed to provide only the general background, and certain specifications, that are required to make the arguments in this paper fully intelligible. For further details regarding epiphenomenalism, and for other criticisms of it, and replies, see Robinson (1982, 1999, 2004a, and 2006).

2 For previous discussions of beliefs and desires see Robinson (1986, 1988, 1990, 1995, 2004b, 2005).

3. Nisbett and Wilson's (1977) results can be taken to show that the fit is not perfect. I regard it as an empirical matter just how much discrepancy, and what kind of discrepancy, there is. I shall, however, assume that in general there is a fairly good fit. The reason is that the larger the failure of fit, the less well organized we can take the brain to be.

4. Povinelli (2000, p. 65): "[B]ecause the ancestral systems were not discarded as we evolved theory of mind abilities, humans may have been left in the philosophically awkward position of having multiple psychological causes for the same behaviors – only some of which penetrate into the highest levels of our conscious experience. Indeed, we suspect that most of the ancient psychological

mechanisms which drive our moment-to-moment behaviors do not intrude into our reflective conscious experience, and therefore we are frequently left to misdiagnose the psychological causes of our behaviors."

5. See, e.g., Broad (1925), Bechtel and Richardson (1983), Van Rooijen (1987), Lindahl (1997).

6. See D. Hutto (2000, pp. 99-100) for an excellent discussion of the need for making the possibility of identities intelligible.

7. I use sets of firing rates as the relevant neural properties because they are easily understood. It may be that other properties are actually the relevant ones. For example, what is important might be ratios of such rates, or rates of change of such rates or their ratios, or intervals between arrival of packets of neurotransmitters at a neural surface, or ratios or rates of change of such intervals, and so on. Since the arguments in this paper do not depend on exactly which neural properties turn out to be the relevant ones, "sets of firing rates" should be read as an abbreviation for a disjunction of the properties just listed, together with further disjuncts that are definable in terms of, or similar to, the properties on that list.

8. But see Author (2004a) for a speculative sketch of how science could lead to discovery of such an explanation.

9. See Levine (1983).

10. James (1890, v. 1, p. 143).

11. Strictly, this premise is redundant, once (J4) has been stated. I include it only because James explicitly mentions parallelism before quickly dismissing it.

12. The structure of this criticism is parallel to the familiar criticism that epiphenomenalism is self-stultifying. I do not discuss this aspect of epiphenomenalism here. I have responded to it in Robinson (1982, 1999, 2004a, and 2006).

13. My debt to Fodor here will be obvious to readers of Fodor (1987).

14. In any case, knapping precedes language. See Wynn (2002).