Consciousness, adaptation and epiphenomenalism

Thomas Polger and Owen Flanagan
University of Cincinnati/Duke University

Consciousness and adaptation

The question of the adaptive advantage of consciousness has been introduced into the debate among philosophers of mind as support for one or another view of the nature and causal efficacy of consciousness. If we can explain how and why we came to be conscious, then that will shed light on what sort of thing, or process, or property consciousness is. If consciousness has been selected for because it is fitness enhancing then we may rest assured that it is causally efficacious, and the epiphenomenalist suspicion becomes less worrisome (Flanagan 1992). And the odds seem good that some kinds of consciousness are adaptations.

For example, surely acute pain states are adaptive. You place your hand in a fire. The fire is hot. Your hand hurts. The pain causes you to remove your hand from the fire. Pain has certain effects relative to human bodies that figure in explanations of our overall capacity to avoid serious injury. *Prima facie*, pain in humans is an adaptation for, among other things, causing us to remove our hands from fire and other sources of injury. Generalizing from cases like pain, the standard view is that consciousness evolved because it conferred its bearers an adaptive advantage.

However it is no easy task to show that consciousness is an adaptation in the strict sense, that is, that it was originally selected for because it increased the fitness of its bearers (Fox Keller and Lloyd 1992). One reason is that consciousness is at once phenomenologically homogeneous and heterogeneous.

Considered at the coarsest grain, conscious states share the property of being experienced: all and only conscious mental states seem a certain way. Indeed only conscious mental states *seem* any way at all; without consciousness there is no subjective, phenomenological point of view.[1] This phenomenological unity of experience distinguishes conscious states from non-conscious states. Exam-

ined more closely, however, conscious mental states vary widely. Experiences of red differ from experiences of green, experiences of colour differ from auditory and olfactory experiences, and so forth. This heterogeneity or variety of conscious mental states has caused some philosophers to wonder whether there is any single phenomenon, *consciousness*, after all (Churchland 1983).

We recognize consciousness under its phenomenal descriptions; conscious states share the Nagel-property: there is something that it is like to have a conscious state (Nagel 1974). But we do not know how to describe it in non-phenomenological terms. Is phenomenal consciousness a single trait, or a host of related traits? That is, are conscious mental states realized in one way in the brain, or in many ways?

1.1 Consciousness: Unity and variety

There is the consciousness in the sensory modalities; there are emotions, moods, dreams, and conscious propositional attitude states; there are various kinds of neuroses and psychoses. All of these are kinds of conscious states. This is not a final description or taxonomy of consciousness. But one must start by picking out the phenomena to be explained and these are ways we point at the phenomena.

If we knew how consciousness was realized in the brain we could give a neural specification. There is important work going on to investigate whether there is some such neural property and what it might be that the states we call 'conscious' have in common.

One possibility is that all conscious mental states are realized by a single neural property, say 40Hz oscillations (Crick and Koch 1990), or recurrency or reentry (Churchland 1995; Edelman 1989). Supposing this were so, then it might be the case that consciousness arose when human brains settled on a certain oscillation frequency or on a certain functional architecture. Settling on the brain states that give rise to consciousness might have been an adaptation or it might have been an evolutionary accident. Even if consciousness was an adaptation, it would not follow that all manifestations of consciousness, paranoia or dreams, for example, were themselves things Mother Nature aimed to be experienced because they were fitness enhancing.

The abilities to walk and run are likely adaptations. Being able to walk and run enable us to be able to waltz, and tango, and tap dance, and pole vault. But Mother Nature did not give a hoot about these bonuses. The ability to tango is not an adaptation. Suppose that, considered from both the phenomenological and neuroscientific points of view, consciousness is a general trait with a common underlying neural feature or set of features that was selected for; it would not follow that all the varieties of consciousness, all the manifestations of consciousness, are adaptations.[2]

Another possibility is that the underpinnings of consciousness are as various as the phenomenology. Perhaps, considered neurophysiologically, consciousness is a disunified phenomenon — an array of processes that are similar only in that they happen to be states that have phenomenal properties. It might then be the case that each conscious process, each kind of consciousness, independently came to be. Some kinds might be adaptations, others neutral free-riders; still others might be exaptations, traits that were not initially selected for but that were later co-opted for their adaptive advantage. There might then be no one answer to the question of the adaptive advantage of consciousness. But it might also be that despite the variety in their instantiation, conscious states were all independently selected for the same reasons; that is, that the having of phenomenal properties, however realized, always confers the same sort of advantage to its bearers. The question of the adaptive advantage of consciousness might then be like the question of the adaptive advantage of camouflage: no sensible person thinks there is a single trait, camouflage, that some creatures have. The ability to camouflage one's body can be achieved by any of many heterogeneous physiologies. Nevertheless, it seems that sensible things can be said about the adaptive advantage of camouflage in general.

The phenomena of consciousness may be realized by a single neurophysiological process that manifests itself it many ways; or the variegated phenomenology of consciousness may be realized in similarly diverse physiology. Whether the physiological realizations of consciousness turn out to be more or less heterogeneous than their phenomenological manifestations will surely be relevant to questions about the adaptive advantage of consciousness. But the relation between these questions runs in both directions: Evolution makes traits; it may be that we cannot determine the answers to questions about the homogeneity or heterogeneity of consciousness prior to answering questions about its evolutionary history. That is, whether phenomenologically distinguishable states will count as one biological trait or many can depend on whether they are the result of one selection process or many.

Such complications do not spell doom for the project of providing ideally complete adaptation explanations of consciousness. But they caution against the glib assumption that the explanatory project is easy. One should not assume that all varieties of consciousness will be found to have etiological

functions, and one cannot assume from the fact that a kind of consciousness is currently adaptive that it was selected for.

1.2 Evolution by natural selection

The following fable is a possible account of the evolution of consciousness: In a finite population of interbreeding organisms, random mutation caused a portion of the population to have some sort of conscious states (i.e., for those states, there is something that it is like for the organism to be in that state.) In each case, the new phenotypic trait (speaking generally, *consciousness*) was heritable. Sadly, a nearby volcano erupted. By chance, the eruption killed all and only the non-conscious organisms. The conscious organisms, however, survived and reproduced successfully, passing on the trait — consciousness. Consciousness evolved.

Evolution occurred because there was phenotypic variation, heritability, and differential reproduction. This is an evolutionary explanation. Does it show that consciousness evolved because it was an adaptation? No. Although *evolution* of consciousness occurred in this case, it was not evolution by *natural selection* but rather by *random drift*. Only by chance did the conscious organisms out-reproduce their non-conscious counterparts; it was not *because* they were conscious that they survived. Evolution by natural selection — adaptation — requires a further element: there must be a cause other than chance for the differential reproduction that leads to evolution (Brandon 1990: 6–9). There must be something about a trait that accounts for the relative advantage of its bearer in a selective environment.

In order to give an adaptationist explanation of consciousness, we need to specify what the adaptive advantage of the feature in question was for a particular type of organism in a particular selective environment. We need to know that the trait has an etiological function. Etiological notions of function are the most common way of thinking about functions among philosophers of biology these days. The idea behind the family of views is that the functions of a thing are those effects for which it was selected.[3] The details of the etiological notion are disputed, but it is helpful to see what one way of formulating it looks like:

> It is the/a proper function of an item (*X*) of an organism (*O*) to do that which items of *X*'s type did to contribute to the inclusive fitness of *O*'s ancestors, and which caused the genotype, of which *X* is the phenotypic expression, to be selected by natural selection. (Neander 1991: 174)

Etiological functions are those that figure in explanations according to the theory of evolution by natural selection.[4] The etiological function of a trait is an effect that gave it an adaptive advantage. To claim that a trait has an etiological function is to claim that it is an adaptation (Amundson and Lauder 1994).

1.3 Dreams and other spandrels of the brain

There are many features of organisms that are not adaptations; human chins, for example (Gould and Lewontin 1978). Consciousness is a trait like any other; although it might be disappointing, it would hardly be surprising if some varieties of consciousness have no evolutionary function. Gould and Vrba call those traits of an organism that that are not themselves adaptations but are byproducts of other traits that have been selected for 'spandrels' or, if they later come to be selected for, 'exaptations' (Gould and Vrba 1982). Such traits lack etiological function.

Dreams are a plausible candidate for a type of consciousness that lacks an etiological function. Dreams are simply the byproducts of brains doing the things that brains do during sleep (Flanagan 1992, 1995, 1996, 2000). Some brain activity that occurs during sleep is an adaptation. The phenomenal mentation that occurs, although it is an effect of those processes, is an evolutionary byproduct of those brain activities for which sleep was selected. Dreaming *qua* experience makes no difference to the inclusive genetic fitness of organisms that dream. The neurochemical processes going on in their brains while they are asleep, including those that cause dreams, *do* make a difference to inclusive genetic fitness; it is just that *dreams* make no difference.

In broad strokes the hypothesis goes like this. Sleeping has an elegant neurophysiological profile, exemplified by reliable changes in brain waves and in the release ratio of aminergic versus cholinergic neurochemicals. There is good evidence that what the brain is doing during different stages of sleep is implicated in cell repair, hormone adjustment, learning, and memory consolidation.

Dreaming during NREM sleep is rationally perseverative and relatively non-bizarre. A person might think that she did not sleep because she could not stop worrying about the exam tomorrow. In fact, she did sleep. NREM sleep is like being awake in many respects and it is easily confused with being awake. NREM mentation is what gets left over from a normal brain gone to sleep. If one were awake one would first worry about the exam and then study. Since the brain does not turn off one continues to worry, but, being asleep one

doesn't get up. The perseverative dream rut doesn't affect the brain's ability to get one into a hypometabolic state in which cell repair and hormone adjustment can take place.

If one sleeps eight hours, then during two of those hours, one's eyes are bolting around under the eyelids. This is REM sleep. Neurochemically the NREM to REM shift marks (roughly) the shift from labor devoted to cell reparation to labor devoted to memory consolidation and storage. The mechanisms required to turn off certain neurons and to turn on others cause waves that incidentally activate areas throughout the brain, especially in the visual areas. Some of these activations are experienced as "thoughts" and "sensations."

Suppose the conscious brain is *independently* prone to try to make sense of thoughts it has. If so, there is no surprise that it tries — and in part succeeds — to supply a coherent story line to the noise it generates while the system as a whole is doing what it is does during sleep.

If being an adaptation is having an etiological function, then we may call the denial that consciousness is an adaptation *etiological epiphenomenalism.* Dreams, according to the story above, are etiological epiphenomena. Dreams are the spandrels of sleep.

Etiological epiphenomenalism is an empirical claim. It claims that the presence of a certain type of consciousness has no adaptation explanation — there is no effect for which that type of consciousness was selected. One can be an etiological epiphenomenalist about specific types of consciousness, e.g., dreams, or one might be an etiological epiphenomenalist about consciousness generally. To say that dreams are etiological epiphenomena is to say that there is no effect of dreams for which they have been selected. Etiological epiphenomenalism about dreams does not draw into question the existence of dreams, or the causal role of dreams in, say, the project of self-knowledge. It just says that, as a matter of historical fact, having dreams is not a trait that was selected for by natural selection. Culture may come to select for dream interpretation. But it is unlikely that sleep activity itself, say, dopamine reuptake, is enhanced by dreaming things that can be interpreted as having certain significance rather than other, or by a population becoming virtuoso dream interpreters.[5]

Defending the etiological epiphenomenalism of dreams does not commit one to any particular conclusion about other varieties of consciousness — least of all those that may be surreptitiously activated during dreaming. One advantage of adopting an approach that treats consciousness as an array of states that share the Nagel-property is that it allows space for the discovery that some sorts

are epiphenomenal, e.g., dreams, while other sorts may have etiological functions, e.g., visual perception.

2. Explaining the evolution of consciousness

Although it seems likely that some varieties of consciousness are adaptations, specifying what the adaptive advantage of a kind of consciousness *might be* is difficult. But it is child's play compared to finding the sort of evidence that would indicate that any such "how possibly" story reflects how some variety of consciousness *actually* gave an organism an adaptive advantage in a selective environment. This problem, the problem of establishing that a "how possibly" explanation is a "how actually" explanation, requires empirical data.[6] By specifying the adaptive advantage of a variety of consciousness, we give an *ecological* account of its relative adaptedness. But even if we can discover the adaptive advantage of some variety of consciousness, that is only one piece of an adaptationist explanation.

2.1 Ideal adaptation explanation

Robert Brandon (1990) formulates five elements for an ideally complete adaptation explanation:

1. evidence that selection has occurred,
2. an ecological explanation of relative adaptedness,
3. evidence that the traits in question are heritable,
4. information about population structure,
5. phylogenetic information about trait polarity.

These five elements figure in explanations in terms of evolution by natural selection.[7]

The second element of ideal adaptation explanation, the ecological explanation, is the one that most cognitive scientists and philosophers of mind focus on when discussing the evolution of some feature of mind — consciousness in the present case. Ecological explanations of relative adaptedness tell why some trait increased the fitness of its bearer in a particular selective environment. Such explanations describe the etiological function of that trait. But giving a plausible story that satisfies the demand for an ecological explanation of relative adaptedness is not by itself sufficient for giving an adaptation explanation. There are four other conditions that need to be satisfied.

Regarding the first element, if one is going to give a story about why consciousness was favored by natural selection, one needs some evidence that *selection* for consciousness has occurred. This is different from the demand for some evidence that the *evolution* of consciousness occurred. In the volcanic random drift story discussed above we have the evolution of consciousness but without selection for consciousness. Evolution by natural selection requires that the cross-generation change was due to some advantage conferred by the trait that was selected for.

Some sorts of evidence that would fit the bill would be fossil evidence, especially if such evidence involved fossils from competing groups of, say, hominids. It is sometimes thought that *Homo erectus* and *Homo sapiens* roamed the earth together. And it is widely thought that *Homo sapiens* were favored because they were more intelligent than these other hominids, allowing, for example, development of linguistic capacities. What might be the evidence that such selection occurred? Intelligence, it has been argued, is linked to encephalization, and language to specific cortical regions of larger hominid brains (Byrne 1995; Wills 1993; Nahmias 1997). According to this line of thought, the fossil evidence provides support for the idea that selection occurred because it shows increased ratio of brain size to body size in *Homo sapiens* compared to other hominids, as well as space for, e.g., Broca's and Wernicke's areas in the larger skull space of *Homo sapiens*.[8]

Similar difficulties arise for finding evidence of heritability, the third requirement for an ideally complete adaptation explanation. Since consciousness is fixed in our population, that is, we are *all* conscious whereas we are not all six feet tall, we can't observe selection for consciousness in the way we can for height. But there are variations in consciousness between persons which might give us the information we seek. Colour blindness may be a case where there is a heritable variation in the qualitative structure of conscious visual experience. Other congenital sensory deficits (blindness, deafness) also suggest the heritability of consciousness bearing traits.

The fourth element requires information about population structure. Information about the frequencies of different traits in the population is needed in order to determine whether selection is at work, rather than, e.g. random drift. In addition, some evolutionary models, for example, group selection models, refer directly to the frequency of a trait in a given population. The idea behind those models is that some traits are group traits, and are sensitive to population density.

The fifth element of an ideal adaptation explanation requires information about trait polarity; that is, what evolved from what. We need evidence that non-conscious creatures evolved into conscious creatures, not vice versa (Brandon 1990: 165–174).

It should be quite clear that few if any ideal adaptation explanations can be given for any trait, much less for consciousness. The point of having an ideal model is not to satisfy it (though that would be nice) but to have a principled standard against which proposed explanations may be evaluated. Brandon's criteria for an ideal adaptation explanation is such a gauge.

Understanding and taking seriously what it would require to give an adaptationist account of consciousness makes it very clear why it's so hard. Such explanations are difficult to give for any trait. The task of giving an adaptation explanation for consciousness inherits those difficulties intrinsic to adaptationist explanation, and complicates them with all the philosophical and scientific problems attendant to consciousness. Nevertheless, if we are going to take consciousness seriously as a natural biological phenomenon then our explanations (adaptationist, mechanical, and otherwise) of consciousness will have to be measured by the same criteria applied to other biological phenomena (Polger and Flanagan 1999).

2.2 Necessity and natural selection

There is a confusion that arises in discussions of consciousness and necessity that we want to get clear about. Suppose there is some organism that performs function f by going into physical state p. Suppose further that p is a conscious state. That is, whatever the relationship between conscious states and physical states turns out to be (identity, supervenience, etc.), p has that relationship to conscious state c. Now, we can ask several sorts of questions about p. One kind of question is: Is it logically, metaphysically, or nomically necessary that any system in state p is thereby in conscious state c? Another question is: Is it logically, metaphysically, or nomically necessary that f be accomplished consciously? These are metaphysical questions. A third question is: Is it logically, metaphysically, or nomically necessary that evolution produce p? This is a historical question. Whatever one thinks the answer to the first two metaphysical questions is, the answer to questions of the third sort is: no.

It is a consequence of taking seriously the idea that consciousness is a natural phenomenon that we must treat it like any other naturally occurring trait of a living organism. The presumption is that those traits that are currently adaptive were formed by adaptation; the burden of proof is on the objector to

show otherwise. As Brandon (1990: vii) puts it, natural selection is "the only general and scientifically legitimate theory of adaptation." Thus, given its apparent adaptedness, the null hypothesis must be that consciousness was selected for in the process of evolution by natural selection. And selection produces contingencies — traits that did not have to be. Even if there are certain configurations of matter (or matter*, depending on your favorite theory) that are necessarily conscious states (*qua* the first sort of necessity distinguished above), and even if they are necessary for some capacity of the organism (*qua* the second sort of necessity), those configurations did not have to be realized. The answer to the question, "Why did consciousness come to be?" cannot be, "Evolution selected it because consciousness is necessary for learning and plasticity?"

This is not because consciousness doesn't give us learning and plasticity — maybe it does. It is because answers like, "Because consciousness is necessary for x" are of the wrong form. "T is necessary for x" is not the form of a proper adaptation explanation of *any* trait. If T is necessary for x then one does not need to appeal to evolutionary theory to explain the presence of x in T. Adaptation explanations do not explain why organisms have mass (presumably a necessary property of organisms) although they may explain why some organisms have the particular mass they do. Even if consciousness were in some metaphysical sense necessary for a capacity of an organism, that would not explain why it came to be. To think so would be to construe evolution as unduly forward-looking, teleological in a strong way not acceptable to most contemporary philosophers.[9]

This exposes the fundamental flaw in every proposal for an evolutionary explanation of consciousness that we have seen. We maintain that no credible adaptationist account of consciousness has been given (Flanagan and Polger 1995). We're not just lamenting that there are no *complete* explanations of consciousness out there — there aren't any that even approximate full adaptationist explanation. They are all explanations in terms of one or another cognitive function that consciousness is alleged to be, in some sense, necessary for.

2.3 Inessentialism and the adaptation question

Daniel Dennett, responding to our claim that no plausible adaptationist account of consciousness has been given, argues that it is a mistake to ask what the adaptive advantage of consciousness might be:

> The question of adaptive advantage, however, is ill-posed in the first place. If consciousness is... not a single wonderful separable thing ('experiential sensitivity') but a huge complex of many different information capacities that individually arise for a wide variety of reasons, there is no reason to suppose that 'it' is something that stands in needs of its own separable status as fitness enhancing. It is not a separate organ or a separate medium or a separate talent.
>
> To see the fallacy, consider the parallel question about what the adaptive advantage of *health* is. Consider 'health inessentialism': for any activity *b*, performed in any domain *d*, even if *we* need to be healthy to engage in it (e.g., pole vaulting, swimming the English Channel, climbing Mount Everest), it could in principle be engaged in by something that wasn't healthy at all. So what is health *for*? Such a mystery! (Dennett 1995: 324–325)

Dennett's parody misses its mark; "health inessentialism" (HI) is unmotivated. Whereas no-one holds that a human being — or an organism remotely like us — could climb Mount Everest without being healthy, many philosophers believe it is possible that there are creatures that could do all the things that we do, but without being conscious. This is the thesis of *conscious inessentialism.*

Conscious inessentialism (CI) is the view that "for any intelligent activity *i*, performed in any cognitive domain *d*, even if *we* do *i* with conscious accompaniments, *i* can in principle be done without these conscious accompaniments" (Flanagan 1992: 5).[10] CI became plausible in the wake of work in artificial intelligence. A system could be very smart but lack experience altogether. Here the contrast case involves a biological system such as a human being and an inorganic computer. There are in fact computers that play world-class chess and are not conscious; therefore, it is possible to play world-class chess without being conscious. Consciousness is not essential to the ability to play chess.

HI might seem plausible in a similar way. A bulldozer might be able to get up Mount Everest, but it is not healthy. Roughly speaking, therefore, one could get up Everest without being healthy. Dennett encourages us to think that just as consciousness is not essential for chess playing, so health is not essential for mountain climbing.

If CI and HI seem equally plausible *prima facie*, important differences reveal themselves upon reflection. CI is meant to apply to both the contrastive cases where organic creatures and inorganic ones are involved *and* where only biological organisms are involved. If we focus on organisms, the claim is that we can imagine experientially blank creatures relevantly like us behaving intelligently across most every domain, i.e., they will pass the toughest Turing-test and more.[11] They will do all the things that pre-theoretically we think

require conscious intelligence, but which upon reflection seem only to require intelligence.

In contrast, the concept of "health" doesn't even apply to bulldozers. It's true that we cannot imagine non-healthy organisms that are relevantly like us, i.e., with cardiovascular and muscle systems like ours, who can climb Mount Everest with weak muscles, lungs, and hearts. But HI doesn't even apply to radically different systems.

One reason HI sounds absurd is this disanalogy. HI fails not because it construes health as a "single wonderful separable thing." Rather, the very notion of *health* doesn't apply to non-organisms.[12] Furthermore, *health* is conceptually intertwined with fitness, and thus with the very notion of adaptation, in a way that consciousness is not. Asking about the adaptive advantage of health is tantamount to asking about the adaptive advantage of fitness. Such a mystery, indeed!

Dennett's health inessentialism parody does not succeed. For the reasons given, it would not succeed even if one thought that consciousness "were a single, separable, wonderful thing." But Dennett knows better than to suggest that we think that! A major theme in our work in philosophy of mind is the idea that there are multiple kinds of consciousness when specified from the phenomenological point of view and probably the neurophysiological point of view as well (Flanagan, 1985, 1992, 2000; Flanagan and Polger 1995). We treat consciousness as a superordinate, rather than as a middle-level or subordinate, category. Consciousness is to conscious vision, and conscious vision to perception of red, as vehicle is to car, and car to Mustang. Consciousness is a kind that admits of complex taxonomization.

As we argued above, that some trait has many subvariants, that its name covers a "huge complex" of traits even within a species, does not render foolish the question of its adaptive advantage. It just means that there has to be an answer for each instantiation — possibly not the same answer.

Sometimes the fact that a trait admits of variations helps explain the trait's existence, as is the case of butterfly wing patterns. There are various wing patterns within individual butterfly species; but this is not by itself sufficient reason to think that it is a mistake to seek an adaptationist explanation of butterfly wing patterns in general. Moreover, the evidence about the frequency, morphology, and similarity of variants, far from undermining an adaptationist explanation, provide crucial clues for theses about wing pattern evolution and development. Recognizing that differing wing patterns are related is crucial to understanding their evolution. To look for independent

evidence for the evolution of each individual wing pattern could be to fail to recognize an important connection. Indeed, in this case it turns out that there is a common developmental mechanism that accounts for the phenotypic variation in butterfly wing patterns (Nijhout 1990). But this discovery *followed* recognition of the adaptive advantage of not only particular wing patterns, but of variation in wing pattern. The presence of phenotypic variation is not itself an obstacle to adaptationist query. Although the variation in butterfly wing patterns turns out to result from a single underlying process, the mere fact of variation does not render senseless the question of the adaptive advantage of wing pattern unless one thinks that the question presupposes a unified answer. But that cannot be the case, since the answer (or answers) to the adaptation question may figure in deciding whether some feature is to be counted as one trait or many.

Asking if consciousness is an adaptation does not depend on knowing ahead of time whether the phenomenal varieties of consciousness will turn out to be neurophysiologically homogeneous or heterogeneous phenomena. Dennett claims that various sorts of consciousness "individually arise for a wide variety of reasons." Maybe yes, and maybe no. Our approach is neutral on this question. When we ask about the adaptive advantage of consciousness to an organism, we are asking for some of those reasons to be specified for some type of conscious state — whether the mechanisms of consciousness turn out to be one or many. We are asking what etiological functions, if any, the various kinds of phenomenal consciousness have in, say, *Homo sapiens.* It may be that consciousness was selected as a single phenomenological and brain trait — the product of recurrency or reenty (Churchland 1995; Edelman 1989). If this was so, consciousness as specified at the superordinate level would have an etiological function, but some, possibly many of its lower-level types, or varieties, might not have etiological functions.[13] Some kinds of consciousness might be etiological epiphenomena.

3. Varieties of epiphenomenalism

The etiological variety of epiphenomenalism is not the only one, and it is not the one that most often figures in the philosophy of mind. We now turn our attention to three ways in which consciousness might be thought of as epiphenomenal:

i. Etiological epiphenomenalism: consciousness depends on the physical and has physical effects, but those effects are not adaptations.
ii. Causal-role epiphenomenalism: consciousness depends on the physical and itself has physical effects, but those effects are not "mechanistic functions" they play no important causal role in the organismic system.
iii. Strict metaphysical epiphenomenalism: consciousness depends on the physical, but cannot have physical effects, period.[14]

We've seen what etiological epiphenomenalism involves. What about the other varieties of epiphenomenalism?

3.1 Causal role epiphenomenalism

Physical systems have causal powers, and causal effects. Some effects of a system, but usually not all, are functions of the system. Robert Cummins has given a formalization of this notion of function:

> *x* functions as a ϕ in *s* (or: the function of *x* in *s* is to ϕ) relative to an analytical account *A* of *s*'s capacity to ψ just in case *x* is capable of ϕ-ing in *s* and *A* appropriately and adequately accounts for s's capacity to ψ by, in part, appealing to the capacity of *x* to ϕ in *s*. (Cummins 1975: 762)

Functions are the effects of something that play a causal role in an explanation of an overall capacity of a containing system. Capacities of a containing system are explained in terms of capacities of components of the system in Cummins' "functional analysis" model of explanation. On this view, functions are those capacities that are appealed to in such an explanation. Cummins-style functions are always ascribed "against the background of a containing system" (Cummins 1975: 763). Amundson and Lauder (1994) follow Neander (1991) in calling Cummins-style functions *causal role* functions.

Epiphenomenalism of the second sort denies that consciousness plays a causal role function in the explanation of human capacities, so we will call it *causal role epiphenomenalism.* Causal role epiphenomenalism is the sort that applies to the noise that an automobile engine makes: engine noise is a physical effect of a physical system, but it plays no role in a mechanistic explanation of how that system operates. It plays no explanatory role because it is not a mechanism in the functioning of the engine as an engine. Of course it may well play a role for *some* system — it might, say, function as a mechanism (or part of a mechanism) for getting an infant to fall asleep. Consider that the Harley-Davidson motorcycle company has filed to trademark the noise that its motorcycles make precisely because other companies are designing their cycles to mimic the distinctive Harley sound. It seems that engine noise can help sell a car or motorcycle. Still, although the noise may play a causal role in the motorcycle-selling system, it plays no causal role in the operation of the vehicle.[15]

3.2 Strict metaphysical epiphenomenalism

Strict metaphysical epiphenomenalism is the sort of epiphenomenalism traditionally discussed by philosophers of mind, and historically associated with property dualism (the idea that mental properties are non-physical properties of physical states.) Strict metaphysical epiphenomenalism, as the name suggests, is a metaphysical doctrine. It is a claim about a kind of state and about the properties (in particular, the lack of causal properties) of that kind of state. Strict metaphysical epiphenomenalism, as Paul Churchland puts it, is the view that mental phenomena "are entirely impotent with respect to causal effects on the physical world" (1990: 11).

When philosophers want to explain strict metaphysical epiphenomenalism, they almost invariably do so by analogy to common cases of causal role epiphenomenalism. They say things like: strict metaphysical epiphenomenalism is the view that mental states are like the sound that the whistle on a steam train makes, or like the thumping sounds that hearts make as they pump blood. Thumping plays no causal role function in hearts; and thumping of hearts plays no causal role function in explanations of human biological capacities.

Strict metaphysical epiphenomenalism need not claim that it is *logically* impossible for non-physical entities to have causal efficacy in the physical world. Rather, given the physical laws of our universe it is *nomologically* or *naturally* impossible. In contrast (i) and (ii) are claims about the actual evolutionary history of kinds of organisms, and about their actual organismic mechanisms, respectively. Strict metaphysical epiphenomenalism is a claim not about what consciousness does or does not *actually* do, but rather about what consciousness *can* and *cannot* do. It is a claim about what is naturally possible — that it is nomologically impossible that consciousness have any physical effects. That is what makes strict metaphysical epiphenomenalism the spooky sort of epiphenomenalism. It is the strange idea that there is something that is itself caused but which can have no effects at all.

Consequences of epiphenomenalism

Although we have raised worries about providing an account of the etiological function of consciousness, we have established nothing about the causal powers of consciousness or even about the possibility of an account that assigns etiological functions to some kinds of consciousness. We are have not argued for "epiphenomenalism" as it is usually discussed in the philosophy of mind. Indeed, even if consciousness is a spandrel, an accidental feature of we mutants who were lucky enough to have been missed by the lava from a volcanic eruption, nothing has been said, let alone established, about its causal role or even its current adaptedness. That some varieties of consciousness are not adaptations tells us precious little about the nature and causal efficacy of consciousness in general. The bridge of the nose is not, as Dr. Pangloss thought, an adaptation for supporting spectacles; but that fact makes it no less causally able to do so.

This point is simple but often missed. No sensible person thinks that the abilities to do calculus or quantum physics are the direct result of selection pressures, although they may well be the by-products of selection for traits needed to get around in the world. Abilities to do calculus or quantum physics are, like dreams, good candidates for etiological epiphenomena. The trouble arises because debates over the function of consciousness have often confused several notions of function, and thus several kinds of epiphenomenalism.

As research on the nature, evolution, and adaptedness of consciousness proceeds, different types of epiphenomenalism will need to be kept apart. One reason is that an entity can lack some kinds of function under a certain description but possess one or more functions under other descriptions. It can be an etiological epiphenomenon without being epiphenomenal in the causal role sense, much less in the strict metaphysical sense. Moreover, some organisms have etiological functions that they never, or no longer, perform. "The tree is dead. I keep it because I can climb it in autumn and clear the gutters. The kids love it because of the woodpeckers. When they go to college and my knees give out, I'll hire a gutter cleaner who has a ladder. Then I'll cut the tree down and use it for firewood."

In a case such as the dead tree, its branches no longer perform some or all of their etiological functions. But functions come in different kinds. Functional status is interest relative; for each effect designated as a function, there may be a variety of effects that are less interesting, considered side-effects for some purpose. The kind of function one assigns primacy to will depend on one's interests and purposes. Perhaps distributing foliage to allow increased exposure to sunlight is one function of tree branches, and one of a kind that is important to explaining why the branches evolved. Other functions, ladder-to-roof, woodpecker-attractor, etc., are secondary to the project of explaining how tree branches came to be, but are useful for other purposes. That the branches of a dead tree no longer perform their etiological function says nothing about the other sorts of functions they can serve.

Epiphenomenalism is also interest relative. 'Epiphenomenalism' is just the philosopher's epithet for lowly functional status. If the branches never had an evolutionary function, if they were etiological epiphenomena, one would not conclude that they had no causal powers at all. So too, even if consciousness in general were a spandrel, this would do nothing to defeat our beliefs about the nature and causal powers of consciousness.

Acknowledgments

We would like to thank audiences in Tucson, Arizona and Atlanta, Georgia for their helpful questions. Special thanks to commentator Todd Grantham. Valerie Hardcastle provided some useful suggestions. Robert Brandon and Carla Fehr helped us work out our ideas about functions and epiphenomenalism, and commented on an early draft of what has become, among other things, this essay. Francis Crick and Christof Koch pressed especially hard their intuitions about the unity and function of consciousness. We appreciate the help of our friends and critics; it is doubtful that we have satisfied all of their concerns.

Notes

1. See Graham and Horgan (this volume) on the question of grain.

2. Francis Crick has encouraged us to consider whether consciousness is realized as a unified phenomenon (40 hertz oscillations.) He regards it is obvious that information need to be bound together in order for action to be generated, and that consciousness supervenes on this binding; so both the unity of consciousness and its function are intuitively clear to Crick. We remain unconvinced; and we certainly disagree with Crick's claim that the answer is obvious. But whereas we believe that conscious states are phonemonologically diverse, our intent is to remain neutral on the issue of the unity or disunity of the subvenient base(s) of consciousness. We further maintain that the fact, if it is a fact, of the neurophysiological unity of consciousness does not entail that there is *one* answer to the adaptation question with respect to all the manifestations of consciousness.

3. The etiological account is customarily traced to Wright (1973) but has received its most widely discussed treatments from Millikan (1989) and Neander (1991). Variations of the

etiological account, broadly construed, have been endorsed by philosophers of biology (e.g., Brandon 1990; Kitcher 1993; Godfrey-Smith 1994; Sober 1985) as well as philosophers of mind (see especially Millikan 1993 and Lycan 1987, 1996).

4. Amundson and Lauder (1994) call this formulation the *selected effect* account of function to distinguish it from more broadly construed etiological accounts. For example, Wright formulates under a more general account of function that also accommodates artifacts. Millikan (1989) calls her verion "proper function." Some philosophers have argued that etiological functions can be subsumed under causal role functions (e.g., Griffiths 1993; Davies 2000). We shall use the narrow, evolutionary, notion of etiological functions.

5. We could put the main point this way: The effects of dreams are not functions relative to the nervous system, as dopamine reuptake is. Rather they are effects relative to the whole-person system. If those individuals who engage in dream interpretation are able to improve their self-understanding, and if this aids them in carrying on with their lives, and this leads to differential reproductive success (and is heritable) then it could be that there *are* selective pressures for dreaming, and for recalling and interpreting dreams. Though dreaming may not have been selected for in the past, it may come to be selected for in the future. Dreams could come to have an etiological function. Unlikely perhaps; but possible.

6. The notions of "how possibly" and "how actually" explanations are explicated in detail in Brandon (1990) §5.3.

7. These five elements are listed in the order that Brandon presents them, which is not to indicate any relative importance. He argues that the relative importance of the various elements will vary on a case by case basis. For example, Brandon (in conversation) has suggested that trait polarity is not an issue in the case of consciousness.

8. Some researchers worry that even if we can show that selection occurred for general intelligence and language, nothing has been established about consciousness. There are, one might point out, no well-established background theories about what sorts of evidence might be signs of consciousness as there are for intelligence and language. Perhaps that is an accurate assessment of the present state of the science of consciousness, but it is likely to change as attention is focussed on the evolutionary questions concerning consciousness. One researcher (Fink 1996) suggests that evidence of consciousness may be found in the complexity of pharynx, the "energy intake hub." The argument is based on the idea that the brain structures required to support consciousness are energy hungry and thus require particulary large air intake. This proposal, however speculative, illustrates that the search for evidence of consciousness need not be restricted to the relative size of fossilized skulls. Our point is not that encephalization or pharynx data is the answer, but merely that the evolutionary project in consciousness studies is not in principle obstructed by the lack of soft tissue evidence, much less the by impossibility of well fossilized phenomenal experiences.

9. Todd Grantham pressed us to distinguish this objection from a related but weaker point based on the thesis of conscious inessentialism (see §2.3 of this essay.)

10. Note that this formulation makes a claim about 'activity' not 'acts' as defined by action theory. 'Activities' in the sense relevant to conscious inessentialism do not involve conscious intentions to act. See Flanagan (1992: 129–30). The thesis of conscious inessentialism is just that — a thesis. It is an attempt to articulate the intuition that lies behind the consensus that the issues of intelligence and consciousness can, in principle, be pried apart. Flanagan (1992), in fact, denies that *Homo sapiens* could do what they do without consciousness. It is a matter of *a posteriori* necessity that certain actions we perform, e.g., lying, require conscious motivation. On this view, a non-conscious system might mislead but it cannot lie.

11. See Flanagan and Polger 1995, and other discussions of so-called zombies.

12. Some philosopher would argue that consciousness, like health, is a concept that does not apply to non-organisms. Dennett is not one of those philosophers; so we shall leave this objection aside.

13. An example of superordinate traits that could be adaptive without particular subordinate types also being adaptive would human eye colour. It seem at least plausible that colouration of the iris could be an adaptation without any particular iris pigment being selected for over others. Thus, eye colour could be an adaptation, but not blue eye colour, *per se*. We have no idea whether this is true of human eye colour, but it illustrates the possibility.

14. There appears to be at least one instance of an interesting variation on type (iii) in the literature (Chalmers 1996):

> iii*. *Limited-exception metaphysical epiphenomenalism:* consciousness is an effect of the physical, but it cannot have physical effects *except* that it causes us to report having conscious experience.

We call this the "limited exception" version of metaphysical epiphenomenalism; it stipulates an exception in order to explain the obvious fact that we often report being or having been conscious. We are going to ignore (iii*), or at least assume that it must be assimilated into (ii) or (iii), because we cannot see that (iii*) is a sustainable position. The conservation laws do not admit of limited exceptions.

15. Notice that the claim is merely that engine noise does not in fact play such a role, not that it could not. There might be some sort of engine or vehicle in which the engine noise does play a causal role. An example from household appliances and the ever tricky business of microwaving popcorn illustrates the possibility: Some microwave ovens have sensors that detect the sound of popcorn popping and, based on the decreasing frequency of pops, automatically stop the process to prevent burning the popcorn. The sound of popping corn plays a causal role function for those ovens, though not in the popping of individual kernels of corn.

References

Amundson, R. and Lauder, G. 1994. Function without purpose: The uses of causal role function in evolutionary biology. *Biology and philosophy* 9: 443–469.

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

Byrne, R. 1995. *The thinking ape: Evolutionary origins of intelligence.* New York: Oxford.

Chalmers, D. 1996. *The conscious mind.* New York: Oxford.

Churchland, P. S. 1983. Consciousness: The transmutation of a concept. *Pacific philosophical quarterly, 64*: 80–95.

Churchland, P. M. 1988. *Matter and consciousness*, revised edition. Cambridge, MA: MIT Press.

Churchland, P. M. 1995. *The engine of reason, the seat of the soul: A philosophical journey into the brain.* Cambridge, MA: MIT Press.

Crick, F. and Koch, C. 1990. Towards a neurobiological theory of consciousness. *Seminars in the neurosciences*, 2: 263–275.

Cummins, R. 1975. Functional analysis. *The Journal of Philosophy LXXII, 20*: 741–765.

Davies, P. 2000. The nature of natural norms: Why selected functions are systemic capacity functions. No. 34, 1: 85–107.

Dennett, D. 1995. The unimagined preposterousness of zombies. *Journal of consciousness studies, 2, 4*: 322–26.

Edelman, G. 1989. *The Remembered present: A biological theory of consciousness.* New York: Basic Books.

Fink, B. 1996. The evolution of conscious behavior: An energy analysis. Presentation at Towards a science of consciousness II, Tucson, Arizona, April 1996.

Fox Keller, E. and Lloyd, E. (eds.) 1992. *Keywords in evolutionary biology.* Cambridge, MA: Harvard University Press.

Flanagan, O. 1985. Consciousness, naturalism, and Nagel. *Journal of mind and behavior, 6, 3*: 373–390.

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

Flanagan, O. 1995. Deconstructing dreams: The spandrels of sleep. *Journal of philosophy: 5–27.* Reprinted with revisions in Flanagan 1996: 32–52.

Flanagan, O. 1996. *Self expressions: Mind, morals, and the meaning of life.* New York: Oxford University Press.

Flanagan, O. 2000. *Dreaming souls: Sleep, dreams and the evolution of the conscious mind.* New York: Oxford University Press.

Flanagan, O. and Polger, T. 1995. Zombies and the function of consciousness. *Journal of consciousness studies, 2, 4*: 313–321.

Godfrey-Smith, P. 1994. A Modern history theory of functions. *Noûs 28,3*: 344–362.

Gould, S. J. and Lewontin, R. 1978. The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist program. *Proceedings of the Royal Society, London 205*: 581–598.

Gould, S. J. and Vrba, E. 1982. Exaptation: A missing term in the science of form. *Paleobiology, 3*: 115–151.

Griffiths, P. 1993. Functional analysis and proper functions. *The British journal of philosophy of science* 44: 409–422.

Kitcher, P. 1993. Function and design. In P. A. French, T. E. Uehling, Jr., H. K. Wettstein (Eds.), *Midwest studies in philosophy XVIII.* Notre Dame, Ind.: University of Notre Dame Press, 379–397.

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

Lycan, W. 1996. *Consciousness and experience.* Cambridge, MA: MIT Press.

Millikan, R. 1989. In defense of proper functions. *Philosophy of science 56*: 288–302. Reprinted in Millikan 1993: 13–29.

Millikan, R. 1993. *White queen psychology and other essays for Alice.* Cambridge, MA: MIT Press.

Nagel, T. 1974. What is it like to be a bat? *Philosophical review, LXXXIII*: 435–450.

Nahmias, E. 1997. Why our brains got so big: Reciprocal altruism deception, and theory of mind. Presented at the Southern Society for Philosophy and Psychology, Atlanta, Georgia, April 1997.

Neander, K. 1991. Functions as selected effects: The conceptual analyst's defense. *Philosophy of Science 58*: 168–184.

Nijhout, H. 1990. A Comprehensive model for colour pattern formation in butterflies. *Proceedings of the Royal Society of London B 239*: 81–113.

Polger, T. and Flanagan, O. 1999. Natural answers to natural questions. In V. Hardcastle (Ed.), *Where biology meets psychology* (Cambridge, MA: MIT Press).

Sober, E. 1985. Panglossian functionalism and the philosophy of mind. *Synthese 64*: 165–193.

Wills, C. 1993. *The runaway brain.* New York: Basic Books.

Wright, L. 1973. Functions. *Philosophical review 82*: 139–168.