



Is Consciousness a Spandrel?

ABSTRACT: *Determining the biological function of phenomenal consciousness appears necessary to explain its origin: evolution by natural selection operates on organisms' traits based on the biological functions they fulfill. But identifying the function of phenomenal consciousness has proven difficult. Some have proposed that the function of phenomenal consciousness is to facilitate mental processes such as reasoning or learning. But mental processes such as reasoning and learning seem to be possible in the absence of phenomenal consciousness. It is difficult to pinpoint in what way phenomenal consciousness enhances these processes or others like them. In this paper, we explore a possibility that has been neglected to date. Perhaps phenomenal consciousness has no function of its own because it is either a by-product of other traits or a (functionless) accident. If so, then phenomenal consciousness has an evolutionary explanation even though it fulfills no biological function.*

KEYWORDS: philosophy of mind, consciousness, philosophy of science, evolution, philosophy of neuroscience

Phenomenal consciousness seems to be the core of who we are. Without it, we wouldn't feel anything; it would be all dark inside us. Worse, without consciousness, we can't even make sense of what 'inside' would be. What could be more important than phenomenal consciousness? Surely if evolution produced conscious beings through the long process of natural selection, phenomenal consciousness must have a biological function—it must have been selected for, thanks to its adaptive value.

'Consciousness' means different things to different people. Our target is *phenomenal* consciousness, variously referred to as the qualitative aspect of experience, phenomenal properties, qualia, 'what it is like' to experience something, or the difference between sleeping dreamlessly and being conscious of something (the world, ourselves, the content of our dreams) (Nagel 1974; McGinn 1989; Chalmers 1996; Block 1995). We are not directly concerned with access consciousness, that is, the availability of information to the cognitive system for the fulfillment of its functions (Block 1995). Access consciousness is obviously important and may be closely connected to phenomenal consciousness, but there is no special problem of finding the adaptive value of access consciousness because, by definition, access consciousness contributes to the cognitive system's functions. By contrast, whether phenomenal consciousness contributes to the cognitive system's functions remains to be determined. We follow the mainstream in assuming that



deflationary views about phenomenal consciousness, to the effect that there are no such things as qualia or phenomenal properties or qualitative aspects of experience (e.g., Dennett 1991), or that phenomenal consciousness is not a proper subject of scientific investigation (Irvine 2013), miss something. We assume that phenomenal consciousness is a real feature of human mentality—worth taking seriously and studying scientifically. For simplicity, from now on when we write ‘consciousness’, we mean phenomenal consciousness.

Many authors find it obvious that consciousness has a biological function, which explains its origin through the slow processes of natural selection that operated during the evolution of our species from its ancestors. By ‘biological function’, we mean a contribution to the survival or inclusive fitness of the organism—the kind of effect on which natural selection operates (Maley and Piccinini, forthcoming). Thus, adaptationism about consciousness is the view that consciousness evolved to fulfill some biological function. How could anyone believe otherwise? For instance, psychologist Nicholas Humphrey writes:

Since [phenomenal] consciousness, as we know it, is a feature of life on earth, we can take it for granted that—like every other specialized feature of living organisms—it has evolved because it confers selective advantage. In one way or another, it must be helping the organism to survive and reproduce (Humphrey 2011: 14–5).

We will argue that taking for granted that consciousness has adaptive value, as Humphrey does, is a fallacy.

Much work has been done to identify the evolutionary benefit of consciousness in functional terms. These efforts have led to a variety of theories that depict consciousness as the means to other valuable traits such as the ability to learn or engage in higher-level reasoning. Such adaptive value, proponents argue, accounts for why evolutionary processes would lead to the emergence of consciousness.

But natural selection cannot distinguish between function f being performed by a conscious organism and f being performed by a nonconscious counterpart. Natural selection is indifferent to what happens inside the organism; only the external effects matter. For example, natural selection would be indifferent to two implementations of a foraging strategy that looked identical from the outside (i.e., resulted in the exact same external behavior), but were computed by different algorithms (i.e., were the result of different neural computations).¹ Adaptive value is determined by function, and function may not require consciousness. Thus, as several authors have pointed out, establishing that evolutionary processes produced consciousness to fulfill some biological function is a tall order (cf. Flanagan and Polger 1995; Polger and Flanagan 2002; Rosenthal 2008; Nagel 2012).

The difficulties with adaptationist accounts appear so serious that some reject standard evolutionary theory—and even materialism along the way (Nagel 2012).

¹ Of course, there may be internal differences that could matter. For example, one algorithm might require greater energy expenditure, and that would result in the more energy-efficient algorithm being selected over time. We are assuming that the different algorithms are equivalent in the respects that matter to natural selection.

This is an overreaction. We will argue that, even if adaptationist accounts fail, there is no need to abandon standard evolutionary theory, let alone materialism.

Instead of an adaptation, consciousness might be a spandrel (in the sense of Gould and Lewontin 1979)—a by-product of some other trait that has adaptive value although consciousness itself has no adaptive value of its own (or may even be dysfunctional). Another possibility is that consciousness is a functionless accident, possibly even a dysfunctional accident.² We call the view that consciousness is either a spandrel or a functionless accident the By-product or Accident View (BAV).

Consider the wiring of the optic nerve, whose fibers extend from retinal cells in the eye to the lateral geniculate nucleus (and a few other brain regions). These fibers exit the eye through a hole in the retina. This arrangement creates the blind spot. The explanation for the blind spot is not that the blind spot provides some adaptive value to the organism. The blind spot is just a by-product of the way retinal axons are wired. By contrast, the way the retinal axons are wired—going toward the inside of the eye rather than the outside—is most likely not a by-product of anything else. It was once thought to be a frozen evolutionary accident, and a somewhat dysfunctional one (it was thought to decrease acuity). The more recent consensus is that the vertebrate retina is organized in this way to allow intrinsically photo-sensitive retinal ganglion cells, a special subclass of retinal neurons, to sit closer to the light source than traditional rods and cones, in order to absorb some ambient light in the red spectrum. If the vertebrate retina were organized otherwise, rods and cones would absorb all the light and intrinsically photo-sensitive retinal ganglion cells would be unable to do their job, which is to drive both circadian rhythms and the pupillary light reflex. Even if the inward wiring of the retinal axons has a function, however, for a long time it was thought to be an unfortunate accident. That is enough to illustrate the concept. In any case, the blind spot is a functionless by-product that has no adaptive value.

The BAV account of the evolutionary origins of consciousness deserves more consideration than it has received to date. It avoids some of the problems faced by adaptationist accounts. According to BAV, complex brains are adaptively advantageous but consciousness is either their by-product or a (functionless) accident. Because consciousness itself is not selected for, its function no longer needs to be identified.

Some aspects of BAV have been independently explored by Polger and Flanagan (2002) and Rosenthal (2008). Polger and Flanagan (2002) formulate a view similar to BAV, which they call ‘etiological epiphenomenalism’. (We will discuss the relationship between BAV and epiphenomenalism in section 4.) They defend it with respect to dreams (cf. Flanagan 1995) and argue that ‘no credible adaptationist account of consciousness has been given’ (Polger and Flanagan 2002: 30).³ Nevertheless, they maintain that ‘given [consciousness’s] apparent adaptedness,

² For present purposes, the hypothesis that consciousness is an exaptation is treated similarly to the hypothesis that consciousness is an adaptation. The important divide is between the view that consciousness was selected for thanks to an adaptively valuable function it performs and the view that it was not.

³ Another example: Sufka (2000) argues that chronic pain is a maladaptive by-product of the combination of ordinary pain (which by itself is adaptive) and neural plasticity (which by itself is also adaptive).

the null hypothesis must be that consciousness was selected for in the process of evolution by natural selection' (30). Rosenthal (2008) argues that consciousness is a by-product of other traits based on his specific account of consciousness. He does not consider that consciousness could be an evolutionary accident. Neither Polger and Flanagan nor Rosenthal articulate BAV as a general thesis independent of specific accounts of consciousness to the degree that we do (sections 2–3), nor do they systematically articulate the relations between BAV and the metaphysics of mind to the degree that we do (section 4).

Before proceeding, we should emphasize that our aim here is not to raise concerns about adaptationist accounts of mental states in general, but only about consciousness. Our mental life has many aspects other than phenomenal consciousness, and many of these fit easily with adaptationist accounts. Clearly, cognitive processes have important biological functions to perform. Our question is whether consciousness does, too.

I. Adaptationist Accounts

Adaptationist accounts of consciousness maintain that consciousness confers adaptive advantages that are rooted in functional advantages. There are many different adaptationist accounts. For example, it's been suggested that consciousness is what allows us to know how the world is, to perform certain inferences, to learn in certain ways, to perform voluntary actions, or to possess a representation of ourselves (Carruthers 2000; Dretske 1997; Eccles 1992; Humphrey 2011; Johnston 2006; Kriegel 2004; Place 2000; Ramachandran and Hirstein 1997; Seth 2009; Trehub 2013; Tye 1996; van Gulick 1989).

Despite their differences, all adaptationist accounts require that consciousness has features that make it valuable in the process of species adaptation. This value explains why evolutionary processes selected for consciousness. Many adaptive value proponents offer quite sophisticated reasons for the development of consciousness—from the roots of reasoning through the benefits of prediction to the usefulness of creativity—that supposedly account for its value. For our purposes, the differences between these theories are inconsequential; it is enough to characterize them as adaptationist accounts, regardless of the specific value each may assign to consciousness.

Adaptationist accounts may be interpreted in one of two ways. On a *strong* reading, they maintain that consciousness is necessary for certain mental processes. That is, the relevant mental processes cannot take place without consciousness. If so, then it is impossible to have nonconscious functional duplicates of conscious beings. Instead, there must be functions performed by conscious beings that a nonconscious being could not perform. In order for such claims to stand, a nonconscious organism must be unable to perform the same function; otherwise consciousness would not be needed to obtain the adaptive advantage.

A major challenge for such strong adaptationism is the apparent possibility of (nonconscious) mental functioning. It is by now a commonplace that many mental

functions are performed by humans and other animals unconsciously. By the same token, there are now machines, such as computers and robots, that perform many mental functions,⁴ and such machines do not appear to be conscious. At the very least, we can explain, in functional terms alone, everything that these machines do: there is no point at which we seem to be leaving something out by not attributing consciousness to them. The intelligence of machines continues to grow, and many experts argue that in the near future we will build machines that can perform every human mental function as well as or better than human beings. The possibility of human-level machine intelligence remains somewhat speculative, although less so with each advance in artificial intelligence. But the strong adaptationist must argue that these machines are *impossible*, and it's not clear why that should be. In other words, the strong adaptationist must argue that consciousness is a functional improvement in systems with human-level mentality and that such improvement cannot be achieved in any other way. But it seems that all relevant mental capabilities could be possessed without having phenomenal consciousness. It is unclear why developing phenomenal consciousness should be necessary to improve a system's functionality.

One possible reply on behalf of adaptive value accounts is that while unconscious processing *could* accomplish everything that conscious processing does, in actual biological organisms consciousness is necessary for certain cognitive processes to occur—it is a necessary part of a sufficient condition. This is the *weak* reading of adaptationist accounts. According to the weak reading, consciousness is merely a necessary aspect of certain mental processes as they occur *in us*. Perhaps machines or organisms very different from us can perceive, plan, and carry out intentional actions without consciousness, but normal biological organisms of the kind that evolved on planet Earth cannot do so. This would be analogous to saying that machines or organisms that did not use DNA could self-replicate (presumably using some other genetic mechanism), although organisms of the kind that evolved on Earth cannot do so. Therefore, consciousness does have important biological roles that make it a phenomenon explainable as a result of selection processes.

The weak reading of adaptationist accounts pays a price: granting that consciousness is not necessary for *any* mental process. According to the weak reading, conscious biological organisms cannot do what they do without consciousness—but this didn't have to be the way they turned out. If evolution had proceeded differently and organisms had evolved the same mental capacities without consciousness, we—or at least some other creatures mentally equivalent to us—would not be conscious, and yet we would all be doing the same things. Thus, given this response, consciousness is a contingent trait on a par with countless other contingent traits developed through evolution. Clearly, this is not the same as saying that consciousness is a by-product of another trait or a sheer evolutionary accident. For by-products and evolutionary accidents are functionless, whereas the weak reading of adaptationism about consciousness does attribute a function to

⁴ Some may object that computers and robots do not perform *mental* functions at all, because they do not have minds. We are simply referring to functions that, when performed by humans, require the exercise of mental capacities, such as playing chess, solving problems, sensing the environment, and so on.

consciousness. But given how attached many of us are to consciousness, we might have hoped that consciousness was indispensable—something that truly makes us functionally distinct from any nonconscious being.

Note that the weak reading of adaptationism about consciousness makes room for the possibility that some biological species or populations are not conscious but have evolved alternative adaptations for the biological functions fulfilled by consciousness. At the very least, we *could* have evolved in such a way as to perform the same mental functions without consciousness, and maybe some of us have! If so, we might as well explore the possibility that consciousness is superfluous simpliciter—that it adds nothing to our mental power.

Regardless of whether adaptationism about consciousness is interpreted strongly or weakly, it is difficult to establish. Many of the difficulties have been canvassed in the literature, especially by Polger and Flanagan (2002). To illustrate these difficulties, let's take just a brief look at an especially powerful argument for adaptationism about consciousness that Polger and Flanagan (2002) did not consider.

Nichols and Grantham (2000) argue that consciousness is an adaptation because it has a certain kind of complexity, and any trait with that kind of complexity is likely to be an adaptation. The complexity in question has to do with unifying several sensory streams into a unified experience in such a way that several subsystems must cooperate to bring about such unity. This argument is ingenious because it relies on an analogy between consciousness and other complex biological traits, such as the eye, in which several subsystems cooperate to bring about a capacity, without attributing any particular function to consciousness (or to its unity). Thus, Nichols and Grantham manage to defend adaptationism about consciousness without committing themselves to any particular hypothesis about the function of consciousness (or its unity), which would be hard to defend.

Nevertheless, Nichols and Grantham's argument can be resisted. A nonconscious system may well unify different sensory streams, and yet consciousness may be a by-product or evolutionary accident that occurs after the unity is unconsciously produced. Therefore, even granting Nichols and Grantham that our experience is unified in the way they assert and that the system that brings about such unity is complex in the way that Nichols and Grantham assert, it may turn out that the system that brings about the unity is unconscious while the unified experience is either its by-product or an accidental correlate. Nichols and Grantham say nothing to rule out this possibility.⁵

For example, consider breathing or eye-blinking. These behaviors are the result of complex systems, yet we are completely unaware of them most of the time. As

⁵ Nichols and Grantham discuss an objection somewhat similar to ours and respond that the objection requires showing that a nonconscious system responsible for the relevant complexity “is structurally distinct from the phenomenal consciousness system we have been discussing” (2000: 665). But our objection requires no such thing. If consciousness is a by-product of a nonconscious system or an accidental correlate, there need not be any structural distinction between the two systems because there need not be any *system* for consciousness at all. Although Nichols and Grantham mount an impressive argument, they appear not to have seriously considered that consciousness may be a by-product or evolutionary accident. Instead, they address their polemic against epiphenomenalism. We'll discuss the relation between BAV and epiphenomenalism in section 4 below.

it turns out, we can also *consciously* do these things, and at those times, there is something that it is like to breathe or to blink. Yet, it is quite easy to imagine that these behaviors are *not* ones we could be conscious of; much like digesting, they would be complex behaviors over which we have no conscious control. These examples illustrate that systems can be complex and unify multiple sensory streams and yet be either completely unconscious (like digestion) or only occasionally (and thus not necessarily) conscious, like breathing and blinking.

It is beyond the scope of this paper to do justice to arguments for adaptive value accounts and anticipate counterarguments; we won't elaborate further on the strengths and weaknesses of adaptationism. It should be clear that it faces challenges. Whether or not these challenges are surmountable will be a matter for future assessment. Meanwhile, we should at least consider other options. If there is an alternative account that fits squarely within evolutionary theory, fits well with a broad range of metaphysical views about the mind, and avoids the problems of adaptationist accounts, it should be explored.

Gould and Lewontin (1979) introduced us to the notion of phenotype by-products through the example of spandrels. Often found in Renaissance architecture, spandrels are the curved areas between arches that fill the space below a dome. They are frequently quite elaborate, and some have become canonical examples of art from the period. In fact, upon first consideration, it would be tough to dismiss their artistic purpose. However, their reason for existing is more prosaic: spandrels are a necessary by-product of the design of the arches and the dome above. While they appear to serve the function of a canvas for priceless art, that usage is an afterthought. In truth, their artistic merits are due to the actions of resourceful artists making use of available space.

In his more recent work, Gould (1991) has applied this concept of spandrels to the human brain in an effort to explain fine arts, religion, and reading. These functions may originate not through some adaptive value but instead as by-products of a large and complex human brain. As another example, many small children are fascinated by navels and wonder what they are for. In and of themselves, navels are useless. But they are a remnant—or by-product—of a very important feature (the umbilical cord) that has obvious adaptive value. In looking at these examples, we can see a process by which certain traits can come into being despite a lack of useful function. These traits are by-products of other traits that are useful. These by-products, then, may or may not have a function (e.g., reading may well have a function whereas navels do not), but their evolution is due to the function of whatever trait they are a by-product of.

Another possible explanation of a trait is that it's a (functionless) accident: it occurred via chance events such as genetic drift. Biologists often use the metaphor of 'sampling error' to explain this phenomenon. For example, eye color and blood type are believed to have no adaptive value. The proportion of individuals with one eye color versus another or one blood type versus another may increase or decrease simply by chance. Given that it is unlikely that exactly the same number of individuals with any given eye color or blood type will reproduce (or that they will reproduce at the same rate or the same number of times, particularly given the complexities of contingent environmental circumstances), it is likely that one

trait will become more prevalent than the other. In certain circumstances—again, through nothing but random processes—one trait could possibly dominate the other entirely (e.g., a large number of the striped organisms might have been grazing on the side of a volcano that erupted, killing them off). If consciousness is like this, it may be a trait that conferred no adaptive advantage to those who first developed it but that nevertheless spread through the population via genetic drift.

In summary, there are two separate questions that have been unduly neglected:

1. Is consciousness either a spandrel (by-product of other traits) or a (functionless) accident?
2. Is consciousness either adaptively neutral or detrimental?

In this paper, we propose to take affirmative answers to these questions seriously and explore some of their ramifications.

2. The By-product or Accident View

If adaptationist accounts of consciousness fail, the challenge is to explain the origin of consciousness despite its lack of adaptive benefit. The suggestion being explored here does just that: consciousness arose from evolutionary processes just like every other aspect of our phenotype, but it did so not due to any adaptive benefits. Rather, it is complex brains that have adaptive benefits, and consciousness is either a by-product of such brains or an evolutionary accident (BAV).

But proponents of adaptationist accounts might cry ad hocery at the suggestion that consciousness is either a spandrel or an evolutionary accident. After all, it seems quite convenient to simply declare as-yet inexplicable traits to be by-products in the absence of a readily available adaptationist account. To make BAV a viable alternative, what is needed is an account of how such by-products could come about and under what conditions it would be reasonable to expect them (cf. Nagel 2012: 62). There are nonarbitrary criteria that traits must meet in order to be candidates for classification as evolutionary by-products, and we will now argue that consciousness meets these criteria.

First, notice that by-products are either necessary or accidental. Necessary by-products are traits that necessarily (in the biological sense) co-occur with adaptive traits; accidental by-products are traits that accidentally co-occur with adaptive traits. For example, imagine a species that evolves bright white horns. Having horns (suppose) is adaptive and was selected for, but the bright white color is not adaptive and was not selected for; it is thus a by-product. Suppose this white color is a liability in the sense that, were these organisms to have nonwhite horns, they would be more likely to survive (perhaps the whiteness makes them more visible to predators). If the whiteness is an accidental (i.e., unnecessary) by-product, then we would expect the whiteness to be selected against, and eventually this species would evolve nonwhite horns (given sufficient time, a sufficiently large population, sufficient variation in horn color, etc.). But if the whiteness is a *necessary*

by-product (perhaps the horns can only be made of white material or in this species the whiteness is genetically tied to having horns at all), then any ‘liability’ would have to be outweighed by the utility of the horns. If this liability is not outweighed, then the whiteness would be maladaptive, and the horns would be selected against. Thus, if a trait is a by-product, and this trait has existed for any considerable time, it cannot be maladaptive overall.

In light of the above, if consciousness is a necessary by-product of other traits, it may even be a liability. But if it is an accidental by-product, it must not be a liability. Either way, consciousness cannot be maladaptive in the sense of leading to selection against itself, without eventually going out of existence. There is nothing about consciousness to suggest that it is, or may be, maladaptive; it seems to have existed for some time, and we are not aware of any argument that consciousness is a liability in the sense just mentioned. In fact, adaptationists defend the opposite view—that consciousness enhances fitness. While we’ve spent some time questioning adaptationist accounts here, it is at least notable that there is no evidence that consciousness is so detrimental to fitness that it should be selected against. If any such evidence is uncovered, this need not lead to the rejection of BAV so long as consciousness is a necessary by-product of other traits, and the adaptive advantages of such traits outweigh the disadvantages of consciousness.⁶

A second criterion is that traits that are candidates for classification as by-product traits must not have both obvious adaptive value and a plausible method of being produced through natural selection. If we can easily recognize why a trait has adaptive value within the ancestral environment and how an organism might obtain that trait through natural selection, BAV is not the most likely explanation. Instead, the trait’s presence must resist either an empirical account of its origin through natural selection or an account of its adaptive value or both. While it’s possible that a by-product phenotype might have one but not the other, the best candidates are traits where both questions remain unanswered. Consciousness is such a case. As we pointed out, no uncontroversial answer has yet been given for why consciousness is adaptively valuable in functional terms. The how question is similarly unanswered. There is certainly no consensus on how consciousness might have arisen through natural selection, and the answers that have thus far been given remain highly contested.

Third, the view that a trait is a by-product is more plausible if we have at least a hypothesis about what it is a by-product *of*. In the case of consciousness, this may be a certain level of organizational complexity of the brain. This organizational complexity may be a matter of having enough tightly coupled components (neurons) or of processing information in a certain way. Perhaps, as a matter of natural law, any sufficiently large number of tightly coupled processing

⁶ Varki and Brower (2013) argue that self-awareness plus theory of mind lead to anxiety about death, which in turn must be compensated for by denying the unpleasant aspects of reality, and these consequences may be maladaptive. But even if Varki and Brower are correct, self-awareness and theory of mind are not the same as phenomenal consciousness, which is the present topic. (For present purposes, self-awareness is a kind of access consciousness, not phenomenal consciousness.) In any case, even Varki and Brower argue that, at least until now, human beings have managed to compensate for the supposedly maladaptive consequences of self-awareness and theory of mind.

devices gives rise to (phenomenal) consciousness: such a system becomes a conscious system. Or perhaps, again as a matter of natural law, any sufficiently powerful information processing device gives rise to (phenomenal) consciousness. It's not that consciousness adds something to the processing power of the device—consciousness is simply a natural consequence of that kind of processing.

While the three above criteria likely do not add up to an exhaustive list, they do at least serve as a guide for what constitutes a phenotype by-product and an explanation for why such a label is not ad hoc. Because by-products and functionless accidents don't have direct adaptive value themselves, they resist adaptationist explanations, but their origins are evolutionary all the same. When thought of in this context, consciousness is a stellar candidate for being a by-product or a functionless accident; if ever there was such a thing as a phenotype by-product or functionless accident, phenomenal consciousness is as good a candidate as any.

An important corollary of BAV is that philosophical zombies may be walking among us! For present purposes, it is useful to understand 'philosophical zombie' more broadly than this term is most commonly used in the recent literature. By 'philosophical zombie', we mean any one of three things. The most discussed type of philosophical zombie is a nonconscious physical duplicate: a being that is physically identical to a conscious being—it duplicates all its physical properties including all of their physical effects—but lacks phenomenal consciousness (Kirk 2014). The possibility of this type of zombie is often believed to be incompatible with physicalism (Chalmers 1996). A second type of philosophical zombie is a nonconscious functional duplicate of a human being. A functional duplicate of a creature duplicates the behavior of that creature, but it need not duplicate all physical effects of that creature (Block 1978). For example, a sufficiently complex robot may be able to duplicate the behaviors of humans although it generates more heat and a different electromagnetic field than humans. Notice that the possibility of nonconscious functional duplicates of a human being is consistent with physicalism so long as consciousness does not contribute to human behavior but contributes to nonbehavioral physical effects.

A third type of philosophical zombie, not yet discussed in the literature, is one that may be possible if BAV is correct. This is a system that duplicates all biologically adaptive behavior but does not duplicate biologically neutral or dysfunctional behavior that is affected by consciousness, where consciousness is either a spandrel or a (functionless) accident that nevertheless affects behavior in some biologically useless ways. If consciousness is a trait whose presence is due to genetic drift, there is no guarantee that the presence of consciousness would become fixed; individuals lacking consciousness could coexist right along with those who have it. Much like eye color (a functionless trait whose relative frequencies are due to genetic drift), the proportion of individuals with and without consciousness could be distributed unevenly.⁷

⁷ Humphrey defines a fourth type of zombie, which should not be confused with any of the types of zombie we defined. Humphrey's zombie "is physically identical to a normal human being except in one crucial respect: namely, that he or she lacks just those evolved circuits in the brain that yield the phenomenal quality of conscious experience" (2011: 70). By assuming that the circuits for consciousness are 'evolved', Humphrey builds into his

3. The Objection from Complex Partial Seizures

Is BAV plausible independently of the difficulties of finding an adaptationist account? Could consciousness be adaptively inconsequential or even detrimental? Might we be evolutionarily better off if we didn't have it?

In order to establish BAV, we should have a thorough account of how consciousness arises in the brain and whether and how it contributes to brain function. If it does not contribute to brain function but instead turns out to be a by-product of brain function or an accidental feature of the brain, then BAV becomes compelling. Unfortunately, we don't know enough about how consciousness arises in the brain, let alone whether and how it contributes to brain function, to answer these questions.

There is some tentative evidence against BAV: although much information processing happens in us unconsciously, some evidence seems to suggest that without consciousness we can't process information in certain ways. Consider cases of patients having complex partial seizures; these patients appear to behave normally in the absence of consciousness (cf. van Gulick 1989; Tye 1996). They respond appropriately to stimuli and may be able to engage in complex behaviors, such as driving or playing the piano. On closer examination, however, they only engage in highly *automated* behavior. They appear unable to make important decisions or form new plans. They are not at their most creative, to put it mildly. If this type of syndrome is a reliable symptom of the loss of the function of consciousness, then consciousness is necessary for certain kinds of information processing.

Nevertheless, an alternative explanation is that complex partial seizures affect both consciousness *and* the brain's ability to integrate information so as to make decisions and form new plans. If so, then this syndrome does not show what the function of consciousness is but only that both consciousness and certain cognitive abilities are affected simultaneously. This may be due to one of two reasons. Either this syndrome directly affects both consciousness and certain cognitive abilities, or the lack of consciousness and lack of these cognitive abilities have a common cause that is triggered by this syndrome.

We recognize that phenomenal consciousness may well contribute something adaptive to mentality. Without consciousness, perhaps integrating information from different parts of the brain is impossible. But given our still limited understanding of the role that consciousness plays in mentality (if any), it is valuable to have, in addition to adaptationist accounts, an account that accommodates the possibility that consciousness provides *no* functional advantage. Thinking through the possibility that consciousness is a spandrel or evolutionary accident is a helpful exercise, which has interesting ramifications for both our scientific understanding of consciousness and the metaphysics of mind.

definition that consciousness is an adaptive feature of minds. Therefore, Humphrey's zombies—unlike all types of zombie we define in the main text—are 'less likely to survive and reproduce' than conscious human beings (Humphrey 2011: 71).

4. Spandrels, Accidents, and the Metaphysics of Mind

We will now show that BAV is entailed by a wide range of views on the metaphysics of mind (together with plausible premises). While this makes BAV appealing and even mandatory for many metaphysicians, it also puts their views at risk. If BAV could be refuted, then several popular options on the metaphysics of mind would have to go with it.

BAV, the view that consciousness is a spandrel or a (functionless) evolutionary accident, is easily confused with epiphenomenalism about consciousness even though it is not the same view. Epiphenomenalism is the view that consciousness has no physical effects—it is causally inert. BAV does not say or entail that consciousness is causally inert. In fact, spandrels and evolutionary accidents generally have physical effects like any other traits—it's just that their physical effects have no biological function, and some of them may even be slightly detrimental. Consider the pinkish-grey color of the human cortex: this is a physical trait that can have physical effects (it reflects light in a certain way). But this pinkish-greyness is a spandrel or evolutionary accident; certainly not a trait that was selected for.

In principle, however, a spandrel or evolutionary accident may well be physically inert in the sense of having no physical effects. This makes BAV compatible with epiphenomenalism (as well as its denial).⁸ For the epiphenomenalist, however, BAV is *compulsory*. The epiphenomenalist, like everyone else, needs to find a place for consciousness in the evolution of our species. BAV provides a ready solution for the epiphenomenalist: consciousness is physically inert because rather than being an adaptation, which would require it to have adaptive physical effects, it's either a (physically inert) by-product of some other trait or a (physically inert) frozen evolutionary accident.

So epiphenomenalism gives rise to a simple argument for BAV:

Argument from Epiphenomenalism

1. Epiphenomenalism: Consciousness is causally inert
2. Causally inert properties cannot be selected for (for lack of effects on which natural selection can operate)
3. Therefore, BAV: consciousness is either a spandrel or an evolutionary accident

⁸ Polger and Flanagan call 'the denial that consciousness is an adaptation *etiological epiphenomenalism*' (2002: 26), which, as we just pointed out, is compatible with both epiphenomenalism simpliciter and its denial. Later in the same essay, Polger and Flanagan redefine 'etiological epiphenomenalism' as the view that 'consciousness depends on the physical and *has physical effects*, but those effects are not adaptations' (34, emphasis added); their second definition makes 'etiological epiphenomenalism' incompatible with epiphenomenalism simpliciter. Nevertheless, on one occasion Polger and Flanagan write 'epiphenomenal' in a context where it's unclear whether they mean etiologically epiphenomenal or epiphenomenal simpliciter (2002: 27).

Assuming that every biological trait must be an adaptation, a spandrel, or an evolutionary accident, the conclusion follows from the premises. Thus, epiphenomenalism entails BAV, but the converse is not true. Premise 2 is unproblematic, and therefore this argument hinges on premise 1, epiphenomenalism.

Epiphenomenalism is a metaphysically strange and unappealing view. Why would there be properties that are physically inert? And how would we know about such properties? The fact that BAV accommodates but does not entail epiphenomenalism—with its metaphysically unappealing consequences—is to BAV's advantage.

BAV is also compatible with property dualism about consciousness. Property dualism is the view that the properties that constitute consciousness are nonphysical. (Property dualism is compatible with epiphenomenalism because the nonphysical properties in question may be causally inert, but property dualism is also compatible with the denial of epiphenomenalism because the nonphysical properties in question may be causally efficacious.) BAV is not committed to property dualism. In fact, typical spandrels or evolutionary accidents are as physical as any other traits.

In principle, however, a spandrel or evolutionary accident may well be nonphysical. This makes BAV not only compatible with property dualism but attractive to the property dualist, including but not limited to the property dualist who is an epiphenomenalist. Ontologically serious property dualists are typically epiphenomenalists (e.g., Chalmers 1996) due to causal exclusion arguments (cf. Kim 1998).

Causal Exclusion Argument 1: Property Dualism Entails Physical Inertness

1. Nonphysical properties are distinct from physical properties.
2. Completeness: Every physical effect has a sufficient physical cause.
3. Exclusion: If a physical property *E* has a sufficient cause *C*, then no other property *C** distinct from *C* is a cause of *E*.
4. Therefore, nonphysical properties are physically inert.

Premise 1 is a tautology. Premise 2 is highly plausible in light of two things: the generality of physics—its laws apply throughout the entirety of space-time—and physics' success in explaining phenomena (Papineau 2001). Premise 3 is a plausible metaphysical principle whose rejection leads to systematic causal overdetermination. In causal overdetermination, the same event has two or more sufficient causes. While there may be occasional events that are overdetermined (e.g., execution by firing squad, where several lethal bullets penetrate the same body at the same time), it seems unlikely that all of our behaviors are causally overdetermined by both a physical and a nonphysical cause. But even if they were, notice that there is an important sense in which the nonphysical cause remains superfluous. For, given

the presence of a sufficient physical cause, the behavior would occur even in the absence of the nonphysical cause. Thus, premise 3 is plausible, and its rejection leads to a very unappealing alternative. The conclusion, which follows from the premises, says that nonphysical properties have no physical effects.

So ontologically serious property dualism gives rise to another argument for BAV:

Argument from Property Dualism

1. Property dualism: Consciousness is a nonphysical property.
2. Physical inertness of the nonphysical: Nonphysical properties are physically inert.
3. Physically inert properties cannot be selected for (for lack of physical effects on which natural selection can operate).
4. Therefore, BAV: consciousness is either a spandrel or an evolutionary accident.

The conclusion follows from the premises, so property dualism entails BAV, but the converse is not true. Premise 3 is unproblematic. Premise 2 may be questioned but the result is ontologically problematic. Thus, this argument hinges primarily on premise 1, property dualism.

Property dualism, like epiphenomenalism, is a metaphysically strange and unappealing view for its own reasons. Nonphysical properties strike some as deeply unnatural—so unnatural as to be unintelligible and incompatible with scientific explanation or investigation. It is not clear what it means to say that something is nonphysical or how we would know about such properties. The fact that BAV accommodates but does not entail property dualism is to BAV's advantage.

BAV is also compatible with nonreductive physicalism about consciousness, which in turn is compatible with but does not entail property dualism. According to nonreductive physicalism, (i) consciousness is a higher-level physical property, and (ii) higher-level properties are distinct from, and thus not reducible to, fundamental physical properties.

Nonreductive physicalism has been criticized on the grounds that the allegedly higher-level properties that give rise to consciousness, insofar as they truly are distinct from and irreducible to fundamental physical properties, are causally redundant. This is another version of the causal exclusion argument (cf. Kim 1998).

Causal Exclusion Argument 2: Nonreductive Physicalism Entails Physical Inertness

1. Nonreductive physicalism (ii): Higher-level physical properties are distinct from fundamental physical properties.

2. Completeness: Every physical effect has a sufficient fundamental physical cause.
3. Exclusion: If a physical property *E* has a sufficient cause *C*, then no other property *C** distinct from *C* can be a cause of *E*.
4. Therefore, higher-level physical properties are physically inert.

Now the nonreductive physicalist is in a bind, for her view entails that higher-level properties, of which consciousness is allegedly one, are physically inert. BAV comes to the rescue. If the causal exclusion argument is correct, there is now a straightforward argument leading from nonreductive physicalism to BAV:

Argument from Nonreductive Physicalism

1. Nonreductive physicalism (i): Consciousness is a higher-level physical property.
2. Higher-level physical properties are physically inert (by the causal exclusion argument).
3. Physically inert properties cannot be selected for (for lack of physical effects on which natural selection can operate).
4. Therefore, BAV: consciousness is either a spandrel or an evolutionary accident.

As before, premise 3 is unproblematic. The argument hinges on nonreductive physicalism (i) (premise 1) and premise 2, which as we've seen is plausibly a consequence of nonreductive physicalism (ii). As before, BAV accommodates the needs of the nonreductive physicalist, and it does so without entailing nonreductive physicalism.

BAV is also compatible with most versions of functionalism (which, in turn, is often construed as a version of nonreductive physicalism). Functionalism is the view that mental states have a functional essence: they are functional states, individuated by their functional relations with inputs, outputs, and other functional states.

If the functional relations that individuate mental states according to functionalism are *defined* to have an adaptive value, then functionalism is incompatible with BAV. Some versions of functionalism are formulated in this way (Lycan 1987; Millikan 1984; Sober 1990). But even if, generally speaking, the functional relations that individuate mental states have an adaptive value (which is a contentious view within functionalist circles to begin with), there is no principled reason why *all* functional relations that individuate mental states should have an adaptive value. Perhaps some of them do and some don't. If some don't, this leaves open the possibility that the functional relations that individuate phenomenally conscious states as such have no adaptive value—no biological function—and hence BAV holds.

Many versions of functionalism are formulated in terms of machines and the functions they perform. Such versions of functionalism often lead to the intuition that nonconscious machines may be functionally equivalent to conscious beings, and therefore consciousness falls outside the scope of functionalism (e.g., Block and Fodor 1972; Block 1978). For present purposes, we do not even need to go all the way to the conclusion that consciousness falls outside the scope of functionalism; all we need is the premise that consciousness makes no positive contribution to biological function. If consciousness has either no function or no biological function, we obtain yet another argument for BAV:

Argument from Functionalist Intuitions

1. Functionalist intuition: Any biological function f performed by conscious organism O^* could have been performed by a nonconscious, functionally equivalent organism O .
2. Properties that make no difference to biological function cannot be selected for.
3. Therefore, BAV: consciousness is either a spandrel or an evolutionary accident.

If consciousness has no biological function, evolutionary processes are blind to the difference between actions performed by a conscious organism and those performed by its nonconscious functional equivalent. Adaptive value depends solely on function—a nonconscious creature with useful functionality is just as evolutionarily fit as a conscious creature with the same functionality. Therefore, any adaptive value that we might attribute to consciousness could have been achieved by a nonconscious being performing the same functions. Philosophical zombies, then, have either the exact same fitness as their conscious counterparts (if consciousness makes no difference to biological function) or have slightly better fitness than their conscious counterparts (if consciousness is somewhat detrimental to biological function).

BAV is compatible with traditional reductive physicalism, also known as the type identity theory of mind. Reductive physicalism holds that mental state types are identical to or reducible to (fundamental or at least lower level) physical state types. Specifically, phenomenally conscious state types are identical to physical state types. But nothing in reductive physicalism says that the physical state types that are identical to the mental state types must all have a biological function. They may well be spandrels or evolutionary accidents.

Finally, BAV is compatible with the metaphysical view according to which the truthmakers for all truths are fundamental entities and properties, where properties are both qualitative and dispositional (Heil 2012). According to this view, the properties that make up the world, including the conscious mind, are powerful qualities—they have both a qualitative aspect, which explains the qualitateness of phenomenal consciousness, and a set of powers, which explains what minds do.

But nothing in this view says that the powers of mental states (or more precisely, the powers of the truthmakers for mentalistic truths) must have a biological function—they may well be adaptively neutral or even dysfunctional. Thus BAV is consistent with the view that fundamental properties are powerful qualities.

In summary, BAV is compatible with every metaphysical view about consciousness and is plausibly entailed by several popular metaphysical views: epiphenomenalism, property dualism, some versions of nonreductive physicalism, and some versions of functionalism. If BAV is false, any metaphysical view that entails BAV is in big trouble. Thus, BAV should be taken seriously.

4. Conclusion

Consciousness may well be an adaptation. But we've just shown that adaptation is not the only account of the evolutionary origin of consciousness. Therefore, even if adaptationist accounts of consciousness were to fail, we should not be tempted to conclude that evolutionary theory must be radically revised (Nagel 2012).

Adaptation gives an extraordinarily strong account for the origins of many phenotypic traits, and it makes sense to ask if it can explain any given trait. But the difficulties with adaptationist accounts of consciousness and the existence of many arguments for BAV make BAV a compelling possibility. If consciousness provides no adaptive benefit, then the question of why consciousness arose through evolutionary processes remains unanswered by the adaptationist.

Adaptationist accounts of consciousness are not the only alternative. Spandrels and evolutionary accidents do not fit adaptationist accounts, yet they are products of evolutionary processes. We see examples of spandrels in the nonbiological world as well as our own anatomy. These examples give us an idea of how such traits arise despite providing no adaptive advantage. But the label 'spandrel' should not be applied arbitrarily to any trait that lacks an adaptive explanation. Instead, a thorough account must be developed that expands our explanatory power. In beginning to articulate such an account, we find that consciousness is a good candidate for being a spandrel or evolutionary accident. It has no obvious adaptive value, and there is no clear account of it developing through adaptive processes. As a result, BAV deserves serious consideration as a competing account of the evolutionary origins of consciousness.

We have also seen that BAV is consistent with, but does not entail, any metaphysical view about the mind. More significantly, BAV is entailed by a number of metaphysical views—epiphenomenalism, property dualism, some versions of nonreductive physicalism, and some versions of functionalism—together with plausible premises. If BAV were refuted, those views would have to be abandoned.

Finally, why is the seeming importance of our conscious mental life not *prima facie* evidence *against* BAV? Most candidates for spandrels and evolutionary accidents are traits that are of little or no importance. The really striking features of organisms, such as the elephant's trunk, the giraffe's neck, or the tyrannosaurus's teeth, are traits that admit of adaptive explanations. Shouldn't the really striking

feature of human mentality—the presence of consciousness—also invite, if not require, an adaptive explanation? First, it may be just a mistake to equate the importance of consciousness to *us* (from the inside, as it were) with the importance of consciousness to evolutionary processes. Some things that seem important to us are also evolutionarily important (such as sex); other things that seem important to us may not be evolutionarily important (such as watching movies). If BAV holds, consciousness *couldn't* be directly important to evolutionary processes, because relative to such processes, the presence of consciousness is indistinguishable from its absence. Additionally, it may be that *feeling that consciousness is important* (or central to one's mental life) is itself a necessary side effect of being conscious. These are just speculations, but they do offer at least some promise of reconciling the diminished place of consciousness according to BAV with the exalted place that it seems to have in our mental lives.

ZACK ROBINSON
UNIVERSITY OF MISSOURI – ST. LOUIS

COREY J. MALEY
UNIVERSITY OF KANSAS

GUALTIERO PICCININI
UNIVERSITY OF MISSOURI – ST. LOUIS
piccininig@umsl.edu

References

- Block, Ned. (1978) 'Troubles with Functionalism'. In C. W. Savage (ed.), *Perception and Cognition: Issues in the Foundations of Psychology* (Minneapolis: University of Minnesota Press), 261–325.
- Block, Ned. (1995) 'On a Confusion about a Function of Consciousness'. *Behavioral and Brain Sciences*, 18, 227–47.
- Block, Ned, and Jerry A. Fodor. (1972) 'What Psychological States Are Not'. *Philosophical Review*, 81, 159–81.
- Carruthers, Peter. (2000) *Phenomenal Consciousness: A Naturalistic Theory*. Cambridge, UK: Cambridge University Press.
- Chalmers, David J. (1995) 'Facing Up to the Problem of Consciousness'. *Journal of Consciousness Studies*, 2, 200–19.
- Chalmers, David J. (1996) *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.
- Dennett, Daniel C. (1991) *Consciousness Explained*. Boston: Little, Brown.
- Dretske, Fred. (1997) 'What Good Is Consciousness?'. *Canadian Journal of Philosophy*, 27, 1–15.
- Eccles, John. (1992) 'Evolution of Consciousness'. *Proceedings of the National Academy of Sciences*, 89, 7320–24.
- Flanagan, Owen J. (1995) 'Deconstructing Dreams: The Spandrels of Sleep'. *Journal of Philosophy*, 92, 5–27.
- Flanagan, Owen J., and Thomas W. Polger. (1995) 'Zombies and the Function of Consciousness'. *Journal of Consciousness Studies*, 2, 313–21.

- Gould, Stephen J. (1991) 'Exaptation: A Crucial Tool for an Evolutionary Psychology'. *Journal of Social Issues*, 47, 43–65.
- Gould, Stephen J., and Richard C. Lewontin. (1979) 'The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Program'. *Proceedings of the Royal Society of London B*, 205, 581–98.
- Heil, John. (2012) *The Universe as We Find It*. Oxford: Oxford University Press.
- Humphrey, Nicholas. (2011) *Soul Dust: The Magic of Consciousness*. Princeton, NJ: Princeton University Press.
- Irvine, Elizabeth. (2013) *Consciousness as a Scientific Concept: A Philosophy of Science Perspective*. Dordrecht: Springer.
- Johnston, Mark. (2006) 'Better than Mere Knowledge? The Function of Sensory Awareness'. In Tamar S. Gendler and John Hawthorne (eds.), *Perceptual Experience* (Oxford: Oxford University Press), 260–90.
- Kim, Jaegwon. (1998) *Mind in a Physical World: An Essay on the Mind-body Problem and Mental Causation*. Cambridge, MA: MIT Press.
- Kirk, Robert. (2014) 'Zombies'. In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Fall 2014. Available at: <http://plato.stanford.edu/archives/fall2014/entries/zombies/>.
- Kriegel, Uriah. (2004) 'Consciousness and Self-Consciousness'. *The Monist*, 87, 185–209.
- Levine, Joseph. (1983) 'Materialism and Qualia: The Explanatory Gap'. *Pacific Philosophical Quarterly*, 64, 354–61.
- Lycan, William G. (1987) *Consciousness*. Cambridge, MA: MIT Press.
- Maley, Corey, and Gualtiero Piccinini. (Forthcoming) 'A Unified Mechanistic Account of Teleological Functions for Psychology and Neuroscience'. In David Kaplan (ed.), *Integrating Psychology and Neuroscience: Prospects and Problems* (Oxford: Oxford University Press).
- McGinn, Colin. (1989) 'Can We Solve the Mind-body Problem?' *Mind*, 98, 349–66.
- Millikan, Ruth. (1984) *Language, Thought, and Other Categories: New Foundations for Realism*. Cambridge, MA: MIT Press.
- Nagel, Thomas. (1974) 'What Is It Like to Be a Bat?' *Philosophical Review*, 83, 435–50.
- Nagel, Thomas. (2012) *Mind and Cosmos: Why the Materialist New-Darwinian Conception of Nature is Almost Certainly False*. Oxford: Oxford University Press.
- Nichols, Sean, and Todd Grantham. (2000) 'Adaptive Complexity and Phenomenal Consciousness'. *Philosophy of Science*, 67, 648–70.
- Papineau, David. (2001) 'The Rise of Physicalism'. In Carl Gillett and Barry Loewer (eds.), *Physicalism and its Discontents* (Cambridge, UK: Cambridge University Press), 3–36.
- Place, Ullin T. (2000) 'The Causal Potency of Qualia: Its Nature and its Source'. *Brain and Mind*, 1, 183–92.
- Polger, Thomas W., and Owen J. Flanagan. (2002) 'Consciousness, Adaptation and Epiphenomenalism'. In James H. Fetzer (ed.), *Consciousness Evolving* (Amsterdam: John Benjamins), 21–42.
- Ramachandran, V. S., and William Hirstein. (1997) 'Three Laws of Qualia: What Neurology Tells Us about the Biological Functions of Consciousness, Qualia, and the Self'. *Journal of Consciousness Studies*, 4, 429–58.
- Rosenthal, David. (2008) 'Consciousness and Its Function'. *Neuropsychologia*, 46, 829–40.
- Seth, Anil. (2009) 'Functions of Consciousness'. In William P. Banks (ed.), *Encyclopedia of Consciousness* (Boston: Elsevier), 279–93.
- Sober, Elliott. (1990) 'Putting the Function Back into Functionalism'. In William Lycan (ed.), *Mind and Cognition* (Malden, MA: Blackwell), 63–70.
- Sufka, Kenneth. (2000) 'Chronic Pain Explained'. *Brain and Mind*, 1, 155–79.
- Trehub, Arnold. (2013) 'Where Am I? Redux'. *Journal of Consciousness Studies*, 20, 207–25.
- Tye, Michael. (1996) 'The Function of Consciousness'. *Noûs*, 30, 287–306.
- van Gulick, Robert. (1989) 'What Difference Does Consciousness Make?' *Philosophical Topics*, 17, 211–30.
- Varki, Ajit, and Danny Brower. (2013) *Denial: Self-Deception, False Beliefs, and the Origins of the Human Mind*. New York: Twelve.