James developed an evolutionary objection to epiphenomenalism that is still discussed today. Epiphenomenalists have offered responses that do not grasp its full depth. I thus offer a new reading and assessment of James’s objection. Our life-essential, phenomenal pleasures and pains have three features that suggest that they were shaped by selection, according to James: they are natively patterned, those patterns are systematically linked with antecedent brain states, and the patterns are “universal” among humans. If epiphenomenalism were true, phenomenal patterns could not have been selected (because epiphenomenalism precludes phenomenal consciousness affecting reproductive success). So epiphenomenalism is likely false.

1. Introduction. William James developed an evolutionary objection to epiphenomenalism that is still regularly discussed today.1 Perhaps because the classic passage where he lays out the objection is so pithy, epiphenomenalists have offered responses that do not, I will argue, grasp its full depth. I begin with a brief history of James’s own encounters with epiphenomenalism. Then I make use of more recent theoretical tools from evolutionary biology to unpack his worry.

*To contact the author, please write to: University Hall 310A, McMaster University, 1280 Main Street West, Hamilton, ON L8S 4K1, Canada; e-mail: kleina7@mcmaster.ca.
†I would like to thank Trevor Pearce and Elisabeth Lloyd for helpful feedback, as well as the US-UK Fulbright Commission, which financially supported this work.

The most important source for James’s early views on phenomenal consciousness generally is a *Mind* essay entitled “Are We Automata?” (James 1879). That essay’s attack on epiphenomenalism was directed at a host of authors who had advocated the latter view, including William Clifford and Herbert Spencer. But chief among James’s targets was T. H. Huxley, whose 1874 Belfast address was a high-water mark for epiphenomenalism. That position had emerged as an attractive account of consciousness, particularly for a new generation of physiologists (like Huxley) who were attempting to extend the strictly mechanistic conception of a reflex arc to cover even voluntary action (Klein 2018).

Although James seems to have been the first to use the term “epiphenomenon” in its philosophical sense (1890/1981, 133, 139, 1186; see Robinson 2015), neither he nor his opponents used the word “epiphenomenalism” in published work, typically employing the phrase “automaton theory” instead. For convenience I will use this latter phrase interchangeably with our more current “epiphenomenalism.”

Hodgson characterized such a view as asserting that “states of consciousness are not produced by previous states of consciousness, but both are produced by the action of the brain” (1865, 278). Huxley would offer a famous metaphor: like the relationship between the sound of a steam whistle and an engine that forces steam through said whistle, conscious states are always products of bodily states, but conscious states never in turn make a causal difference to bodily states, according to epiphenomenalism (1874/1894, 240). This view portrays humans as automata in the sense that all behavior is understood to be a mechanistic product of prior bodily states, with no causal intervention from consciousness. Consciousness thereby becomes “a simple passenger in the voyage of life,” as James would put it, something “allowed to remain on board, but not to touch the helm or handle the rigging” (1879, 1).

2. Against Epiphenomenalism. It is useful to distinguish two related objections James raised, one domain general and the other domain specific. The domain-general objection targets the notion that any conscious states could have been shaped by selection if epiphenomenalism were true. The domain-specific objection targets the notion that phenomenal pleasures and pains associated with “fundamental vital processes” could have been shaped by selection. The latter issue is now the more widely discussed.

2. Parts of the essay were later used in chaps. 5 and 9 of James’s opus, *The Principles of Psychology* (James 1890/1981).

3. For background on James and Spencer and on James’s engagement with evolutionary theory more generally, see Pearce (2018), which greatly improves upon the discussion of James in Wiener (1949). On James and evolution, also see Richards (1987, chap. 9).
The domain-general objection is that automaton theory depicts consciousness as a “mere supernumerary” in the sense that it could make no causal difference to the animal’s behavior. But then consciousness “would be useless” from an evolutionary standpoint, in that it could not itself have contributed to the animal’s reproductive success and so could not have evolved via natural selection (James 1879, 3; also see 1890/1981, 142). I take this part of James’s criticism to be uncontroversial. As we will see, the prevailing epiphenomenalist response today (which is to contend that consciousness evolved through evolutionary processes other than selection) does not try to resist this initial move.

James’s domain-specific objection targets epiphenomenalism’s prospects for making evolutionary sense of a narrower band of our phenomenally conscious experiences—that is, of a particular subset of our phenomenal pleasures and pains. James wrote that there is a “set of facts which seem explicable on the supposition that consciousness has causal efficacy” (and not explicable by epiphenomenalism; 1890/1981, 146). The facts in question concern the link we typically find between what phenomenally feels good or bad and what physically benefits or harms us. I quote this widely discussed, classic passage (CP) at length:

*It is a well-known fact that pleasures are generally associated with beneficial, pains with detrimental, experiences. All the fundamental vital processes illustrate this law. Starvation, suffocation, privation of food, drink and sleep, work when exhausted, burns, wounds, inflammation, the effects of poison, are as disagreeable as filling the hungry stomach, enjoying rest and sleep after fatigue, exercise after rest, and a sound skin and unbroken bones at all times, are pleasant. Mr. Spencer and others have suggested that these coincidences are due, not to any pre-established harmony, but to the mere action of natural selection which would certainly kill off in the long-run any breed of creatures to whom the fundamentally noxious experience seemed enjoyable. An animal that should take pleasure in a feeling of suffocation would, if that pleasure were efficacious enough to make him immerse his head in water, enjoy a longevity of four or five minutes. But if pleasures and pains have no efficacy, one does not see (without some such *a priori* rational harmony as would be scouted by the ‘scientific’ champions of the automaton-theory) why the most noxious acts, such as burning, might not give thrills of delight, and the most necessary ones, such as breathing, cause agony. The exceptions to the law are, it is true, numerous, but relate to experiences that are either not vital or not universal. . . . The only considerable attempt, in fact, that has been made to explain the distribution of our feelings is that of Mr. Grant Allen in his suggestive little work *Physiological Ästhetics*; and his reasoning is based exclusively on [the] causal efficacy of pleasures and pains. (James 1890/1981, 146–47)*
Consider a subset of our phenomenal pleasures and pains—namely, those that natively go along with “fundamental vital processes.” Examples include unpleasant feelings associated with being burned, suffocated, or physically injured and pleasurable feelings associated with eating, drinking, and resting when needed. Epiphenomenalists accept that these life-essential, phenomenal pleasures and pains (LEPPPs) are effects of the beneficial and harmful brain states (BHBSs) with which they are natively associated. But epiphenomenalists (like “Mr. [Herbert] Spencer”) cannot consistently say that they are adapted effects, or in other words proper etiological functions, of those BHBSs. In other words, if epiphenomenalism were true, there can have been no selection pressure on any BHBSs to produce their associated LEPPPs. This is because if epiphenomenalism were true, LEPPPs could have no “efficacy” and so (recall James’s uncontroversial argument) could have made no difference to reproductive success. But LEPPPs have three features that do suggest that they were shaped by selection: our LEPPPs are (a) natively patterned (they have a characteristic “distribution”), (b) those patterns are systematically linked with antecedent BHBSs (this is the “well-known fact”), and (c) the patterns are “universal” among humans. James concludes that epiphenomenalism cannot make sense of how our natively patterned LEPPPs (with their systematic connections to antecedent BHBSs) could have evolved.

There is a crucial but suppressed premise here: if a–c cannot be explained by appeal to selection, then their evolution cannot be explained at all. It is exactly this premise that today’s epiphenomenalists effectively deny.

What is now the standard response was first articulated by Jackson (1982). He suggests that phenomenal consciousness could be a spandrel (in the sense of Gould and Lewontin [1979]), in which case the evolutionary objection to epiphenomenalism would be overcome. In my terminology, Jackson is suggesting that our BHBSs could have evolved to produce our native LEPPP patterns through processes other than selection. This response has been developed more recently by Robinson (2007, 2014) and Robinson, Maley, and Piccinini (2015).

4. I take James to be picking out a subset of our phenomenal pleasures and pains by appealing to the life-essential (“fundamental” and “vital”) character of the bodily states that cause them.

5. I follow Lloyd and Gould’s definitions, according to which an “adaptation” is “a trait that has a direct proper (etiological) function.” They define a proper, etiological function this way: “a trait has the function of x-ing, if x-ing increased fitness in recent evolutionary history . . . (over alternative, non-x-ing, versions of the trait). . . . increased fitness by x-ing, [and] explains the prevalence of the x-ing trait” (Lloyd and Gould 2017, 51).

6. By “native” I mean inheritable—I use the former term for linguistic convenience. The “distribution” of pleasures and pains at issue in CP must be inheritable since only inheritable traits can evolve.
Jackson is correct that selection is not the only cause of evolutionary change. Common examples of nonselective evolutionary factors include phylectic inertia, genetic drift, linked genes, pleiotropy, and developmental constraints. The latter three can produce so-called by-products, or traits that have proliferated because they happen to have been genetically or developmentally linked to another trait that itself has been selected. So when James claims that epiphenomenalism makes the evolution of our native LEPPP patterns inexplicable, did he simply overlook the possibility that those patterns could have been evolutionary by-products rather than adaptations?

The answer is an emphatic no. Elsewhere in the *Principles* James repeatedly explained other mental traits as evolutionary by-products. He also identified clear and broadly acceptable criteria for distinguishing between by-products and adaptations. When one reads James’s domain-specific objection to epiphenomenalism in light of his criteria for distinguishing by-products from adaptations, a much more powerful objection emerges than has been appreciated in the literature, namely, that our BHBSs have phenomenal effects that have clear adaptive hallmarks, according to widely accepted standards in biology, and so contra epiphenomenalism, those effects are highly likely to have evolved through direct selection.

So what features of traits are adaptive hallmarks for James? His criteria for identifying likely adaptations are on especially clear display in his explanation of the evolution of the morbid fear of heights, which he takes to be a by-product.

High places cause fear of a peculiarly sickening sort, though here, again, individuals differ enormously. . . . That they [i.e., fears of high places] are a mere incidental peculiarity of the nervous system, like liability to seasickness, or love of music, with no teleological significance, seems more than probable. *The fear in question varies so much from one person to another, and its detrimental effects are so much more obvious than its uses, that it is hard to see how it could be a selected instinct. Man is anatomically one of the best fitted of animals for climbing about high places. The best psychical complement to this equipment would seem to be a ‘level head’ when there, not a dread of going there at all. . . . A certain amount of timidity obviously adapts us to the world we live in, but the fear-paroxysm is surely altogether harmful to him who is its prey.* (James 1890/1981, 1036–37, italics mine)

James offers two reasons for thinking acrophobia is likely a by-product (or what he often calls an “incidental peculiarity”). A First, “individuals differ

7. There can be no doubt that James had by-products in mind here. Throughout chaps. 24, 25, and 28 (respectively on instinct, emotion, and cognition), James identified a host of
enormously” with respect to the trait. Rather than approaching ubiquity in the human population, as he thinks one would expect if the trait were a product of selection, the fear of heights “varies so much from one person to another” that it is likely to be a by-product. Second, James suggests that the trait’s “effects” lack any clear “uses” that would outweigh its costs. If natural selection had designed our psychological response to heights for climbing utility, one would expect that response to be “level head[edness],” perhaps dampened by a twinge of “timidity,” not the near paralysis we find in cases of acrophobia. In short, James expects traits that are products of selection to be ubiquitous and to be well designed, perhaps in the sense of producing effects whose uses outweigh costs.

These criteria are actually reflected in CP, albeit obliquely. Recall that I already identified three features of LEPPPs that CP identifies as evidence of adaptation. What I called feature c is just the “universality” of a trait in a population. And James might plausibly take features a and b (LEPPPs are natively organized into patterns and systematically linked to antecedent BHBSs) as amounting to the kind of natural design that only evolves through selection.

But what is it, specifically, about features a and b that might suggest evolution by selection? And in any case, how does James’s argument fare with respect to more recent biological practice?

One way we now establish that some trait is a likely adaptation (particularly when we cannot yet identify a specific etiological function) is to appeal to a set of features one expects to find in a trait that has evolved through selection. The standard list comes from Williams (1966, 10), who suggests looking at a trait’s “precision, economy, [and] efficiency.” That list is sometimes extended, with Symons (1990, 429) adding “complexity, and constancy with which effects are achieved” and Andrews, Gangestad, and Matthews (2002, 496) adding “specificity, proficiency . . . reliability of development, [and] complexity of design,” to give two examples. I will call these features “adaptive hallmarks.” James was effectively highlighting two adaptive hallmarks that are still recognized today as evidence of adaptation: complexity and precision.

mental traits for whose existence “no plausible reason can even be conceived.” They are “purely mechanical results of the way in which our nervous centres are framed,” traits that did not evolve “independently, for any utility they might possess,” but rather are “incidental to others [that have themselves] evolved for utility’s sake” (James 1890/1981, 1097; also see 1046, 1047, 1049, 1225, 1246). These traits are clearly what we would call by-products; James typically characterizes them as “incidental” effects of other selected traits. What is more, notice that he likens the fear of heights to seasick proneness and an inveterate love of music—two of his favorite examples of by-products (e.g., at 1049, 1096–97).
Let us look more closely at the first of these. Complexity of design is still widely thought to constitute evidence of evolution by selection (Symons 1990, 429; Andrews et al. 2002, 496). In fact, that complexity is a hallmark of adaptation is a point of basic biology one can find in respected textbooks on evolution (e.g., Futuyma 2005, 250).

Futuyma offers the pecten as an illustrative example. This is a highly complex structure in birds that protrudes in front of the retina and is supplied with an extensive network of blood vessels (for an illustration, see Futuyma [2005], 262). Because of its complexity, biologists have long believed that the trait is an adaptation, even before they had a good explanation of what the specific, adaptive function of this structure might actually be. Futuyma writes: “Even if we cannot immediately guess the function of a feature, we often suspect it has an adaptive function if it is complex, for complexity cannot evolve except by natural selection” (261–62). I will now argue that feature a (LEPPPs are natively organized into patterns) amounts to the kind of complexity that biologists still think can only evolve via selection.8 We can get an initial sense of this phenomenal complexity by consulting Grant Allen, whose account of “the distribution of our feelings” of pleasure and pain we saw James approvingly citing in CP.

Allen documented and classified a rich diversity of pleasures and pains. His taxonomy of pains, for example, finds a place for discomfort associated with “the amputation of a limb, the excision of an ulcer, . . . the removal of a scalp,” “wounds, cuts, pricks, and scratches,” “burning off a finger, having the feet frozen so that the joints drop off, destroying the skin and muscles with a corrosive acid,” “par[ing] or break[ing] the nails below the quick, . . . pull[ing] open a sore, . . . hav[ing] the face or lips chapped,” cases in which “portions of the body waste away in eating sores, such as abscesses, cancers, ulcers, whitlows, &c,” “corns, bunions, bedsores, and lacerations,” “intestinal pain,” “the passage of renal calculi, gall stones, or clotted catamenial discharges,” “sprains, cramps, and spasms,” rubbing “salt or pepper” on “a wound or burns,” attempts to “to tear off the nails, to flay alive, to pull out the hair, to draw a tooth,” “mustard and cayenne pepper in excess,” “very loud sounds,” “fatigue after muscular exertion; mental weariness; inanition from want of food; faintness from anaemia, loss of blood, sleeplessness, or over-exertion; weakness from fever or other depressing disease; nervous debility; and those undefinable organic feelings which result from general

8. A precise definition of what constitutes complexity has been frustratingly elusive for reasons set out in Hazen et al. (2007, 8574–75). What is more, attempts to define complexity precisely are typically undertaken in connection with the question of whether a given species can be said to be more complex than another (typically an ancestor). At the trait level, biologists seem to rely on a more intuitive, informal grasp of what constitutes complexity, as we can see in Futuyma’s discussion.
ill-health,” among others (Allen 1877, 6–9, 11–15). His taxonomy of pains gives order to a surprisingly diverse set of phenomenal states, in short, and similarly with his taxonomy of pleasures (which can be found at Allen [1877], 20–26).

We should count an apparent pattern in our phenomenal states as adaptively complex (i.e., as bearing the adaptive hallmark of complexity) if and only if the pattern satisfies two criteria. First, the pattern must involve a variety of phenomenal state types. Allen amply documented such variety in the case of our native LEPPPs, as I illustrated in the prior paragraph. Second, the pattern must be stable in the sense that relations between the patterned state tokens must be isomorphic not only when other tokens of the same types recur in one person but also when tokens of the relevant types occur in different persons, regardless of cultural context. For example, it cannot just be that a severe burn happens to feel worse than a hangnail here today but that my preferences might change tomorrow or might have been different had I been raised in a different culture.

We must admit that patterns among pleasures and pains that are not life essential do not count as adaptively complex precisely because those patterns do not meet this second criterion (i.e., they are not appropriately stable). For instance, music appreciation or the love of drunkenness can vary dramatically from person to person (some but not all prefer music to silence, some but not all prefer harmonious to cacophonous music, etc.).9 So the strongest version of James’s argument against epiphenomenalism must confine its scope to life-essential pleasures and pains.

And when it comes to LEPPPs, Allen documents patterns that are remarkably stable. To use two of his examples, virtually all persons would find the pain of passing a “gall stone” to be more severe than that of having the “lips chapped.” He also makes his point by asking readers to “remember the sharpness of the smart which is produced by pulling out a hair or tearing a small piece of skin off the back of the fingers, and then contrast its volume with that which is experienced on breaking a limb or undergoing a severe surgical operation” (Allen 1877, 13). Breaking a limb feels worse than pulling out a hair to virtually everyone, in any time or culture. These patterns in the “volume” of our life-essential phenomenal pains appear to be stable in the relevant sense, and the same presumably applies to life-essential pleasures. Of course, we have already noted the impressive variety of phenomenal state types involved in such patterns. So, just as the complexity of the pecten suggests evolution by selection, there is excellent reason for thinking

9. In CP, James dismisses variation in different persons’ love of drunkenness as one of several exceptions to the “well-known fact” that need not trouble us because they “are either not vital or not universal” (1879, 17–18). He acknowledges “love of music” as similarly variable (James 1890/1981, 1037).
that the complex “distribution” of our LEPPPs suggests evolution by selection as well—a point epiphenomenalists cannot accept (per James’s domain-general argument).

What is it about feature b (LEPPPs are systematically linked to antecedent BHBSs) that suggests evolution by selection? The adaptive hallmark in play, here, is what Williams calls “precision.” I will give an uncontroversial example of adaptive precision before returning to the case of LEPPPs.

Consider the Australian *Stylidium bicolor*, or triggerplant. The sexual structures are housed in a column that normally rests below the surface of the flower (for an illustration, see Futuyma [2005], 321). The petals are shaped in such a way that when a pollinator lands, its weight “triggers” the column to swing up and tag the insect with its pollen-covered tip. The precise delivery of the pollen depends on the column having exactly the right length and the petals having exactly the right shapes, surface areas, thicknesses, and so on. The only evolutionary mechanism that can plausibly explain this sort of precision is selection (Williams 1966, 10; Symons 1990, 429; Andrews et al. 2002, 496).

So, the triggerplant is shaped in such a way that it precisely delivers pollen to visiting insects. Can BHBSs be understood as producing LEPPP patterns with similar precision? Again, we do well to consult Allen, who did not just document complexity in our LEPPP patterns. He also identified a remarkable “concinnity,” as he aptly called it, between BHBSs and the LEPPPs they produce. He wrote: “the consciousness of Pain or Discomfort bears somewhat the same relation to other conscious states as the physical fact which underlies it bears to other conditions of the system” (Allen 1877, 20). Of two pains, the one that is relatively more phenomenally distasteful is also very likely to have been caused by a bodily state that is relatively more threatening to the physical organism, according to Allen (also see 17–18).

I suggest that James had precisely this concinnity in mind when he cited in CP the “well-known fact that pleasures are generally associated with beneficial, pains with detrimental, experiences,” for Allen himself had put the concinnity claim in remarkably similar language (29).

Importantly, Allen pointed out that we could identify this concinnity “without entering into the ultimate question of the connexion between physical and psychical states” (1877, 19–20). He meant that one can accept concinnity without begging Huxley’s question: whether conscious states are causally efficacious. For consider a causal claim that neither interactionists nor epiphenomenalists dispute: that BHBSs reliably bring about their characteristic LEPPPs. If Allen’s concinnity claim is correct (and I think it is), then epiphenomenalists and interactionists alike should both accept that there is a kind of quantitative precision with which BHBSs bring about their associated LEPPPs. In particular, a brain state that registers relatively more (or less) harm to the organism (e.g., a brain state that registers a broken limb
as compared with a state that registers a yanked-out hair strand) will tend to produce pain states that are more (or less) phenomenally repugnant.

Epiphenomenalists and interactionists alike must also recognize a qualitative precision with which BHBSs bring about their associated LEPPPs. We experience the distinctive qualitative discomfort of a sprain if we try to walk on an injured ankle and a distinctive urge to drink water when dehydrated; we feel the distinctive urge for sexual gratification when we have been aroused and the distinctive craving for salt when sodium levels in our blood drop; we experience distinctive bodily relief when we overcome constipation and distinctive chills when we have a raging fever; and on and on.

One could perhaps believe that some pattern of neural flickering happens to produce some one phenomenal property as an accidental by-product. But it defies the standards of evolutionary biology to assert without argument that nonselective forces might have produced the spectacular precision—both quantitative and qualitative—with which our BHBSs bring about our LEPPPs.

To recap, I take James’s attack on epiphenomenalism to work like this. BHBSs produce a complex effect (i.e., stable patterns among a variety of LEPPP states) and to produce that effect with both quantitative precision and qualitative precision. This multifaceted complexity and precision strongly suggests evolution by selection. But if epiphenomenalism were true, these effects could have made no causal difference in the life of any organism, and so the effects could not possibly have been selected. Hence, epiphenomenalism is empirically implausible.10

10. Although they do not take note of James’s prior work on this issue, Nichols and Grantham argue that phenomenal consciousness is a “complexly structured system” in that it “draws information through several independent input channels” together into a “unified” conscious state; they suggest this complexity makes it likely that consciousness is an adaptation (2000, 659, 661). This is an argument from what we might call synchronic complexity, but it is subject to an objection. That we are tempted to call phenomenal states that occur over “short temporal spans” “unified” (a claim explicitly made by Nichols and Grantham [2000, 659]) suggests that such states may be better regarded as undivided wholes. Possibly, what is actually drawing together sensory information from diverse modalities is not the unified phenomenal state itself but some underlying brain structure (for this kind of response, see Robinson et al. [2015, 370]). Ironically, James himself maintains exactly this sort of physiological view of sensory integration (1890/1981, 157–59; for a discussion, see Klein, forthcoming). Thus, he obviously cannot, and does not, argue that consciousness is an adaptation on the basis of any supposed synchronic complexity that might be discoverable in our occurrent phenomenal states. Instead, James is making an argument from diachronic complexity. He identifies a complex pattern in the relative intensity of pain experiences as they occur across a lifetime of phenomenal experience, such that (e.g.) the normal phenomenal pain of breaking a limb hurts worse than breaking a nail whenever these events occur. The response to Nichols and Grantham from Robinson et al. (2015) is not easily adapted to James’s diachronic argument from complexity. Robinson et al. would have to suggest that while the brain
processes underlying these various pains are diverse, there is a sense in which there is a unity of phenomenal experience across the lifetime of the subject of these patterned phenomenal pains. Two experiences separated by a night’s dreamless sleep, for example, could not be regarded as two experiences—they could not even be treated as parts of a single, very long experience, because then the long experience is complex in the sense of being constituted by simpler parts. The problem for this response is even worse for pain experiences separated by years or decades.


