

James and Consciousness

Alexander Klein

The Oxford Handbook of William James

Edited by Alexander Klein

Subject: Philosophy, History of Western Philosophy (Post-Classical)

Online Publication Date: Apr 2021 DOI: 10.1093/oxfordhb/9780199395699.013.4

Abstract and Keywords

Between 1872 and 1890, William James developed an evolutionary account of phenomenal consciousness. He contended that consciousness enables the active evaluation of what is in (or might be in) one's environment. James hypothesized that this evaluative capacity was selected (in the Darwinian sense) because it regulated the behavior of vertebrates with highly articulated brains. His hypothesis was intended to explain some surprising results in physiology, particularly a series of experiments purporting to show purposive behavior in (of all things) decapitated frogs. This chapter reconstructs and evaluates James's evolutionary hypothesis, showing how it would explain those surprising experiments. His account requires interactionist dualism, so he also developed what would become an influential objection to epiphenomenalism: that the latter cannot explain the evolution of our natively patterned, phenomenal pleasures and pains.

Keywords: William James, phenomenal consciousness, evaluation, evolution of consciousness, adaptation, purposive behavior, epiphenomenalism, interactionism, behavior regulation

Introduction

William James offered an evolutionary-psychological account of phenomenal consciousness. In his view, consciousness enables the active evaluation of what is in (or might be in) one's environment. James hypothesized that this evaluative capacity was selected (in the Darwinian sense) because it "regulated" (ECR 1875, 303; PP 1890, 147) the behavior of creatures with highly articulated brains.

James's work on consciousness is substantively interesting in that he offers us a well-developed alternative to more familiar, naturalistic accounts available today. But his work is also methodologically interesting in that he did not develop, and did not intend to develop, what contemporary philosophers of mind are often seeking: a direct *explanation* of first-person conscious experience. He did not try to deduce the existence or nature of consciousness from another set of facts—he did not, that is, treat consciousness itself as an *explanandum*.¹ In fact, James regarded as perniciously metaphysical what we now call

the “hard problem” of explaining why *this* brain state produces a conscious experience with *that* kind of qualitative feel.²

Instead, what he sought to explain was a set of results in experimental physiology concerning apparently purposive behavior in, of all things, live decapitated frogs. These results were hotly debated in his day, and James thought them puzzling from an evolutionary-biological standpoint specifically. Thus, he did not offer a direct proof of his hypothesis about consciousness; instead, he (abductively) recommended the hypothesis because it would help explain an evolutionary puzzle about physiology. My aim in this chapter is to evaluate James’s account of consciousness by showing how it would explain the physiological results he found puzzling.

The most important source for James’s early, interactionist-dualist account of consciousness is a *Mind* essay entitled “Are We Automata?” (EPs 1879, 38–61).³ The article represents the culmination of at least six years’ reflection on the place of consciousness in nature, for James, starting with an unpublished 1872 essay in which he had actually defended epiphenomenalism.⁴ By the time he wrote two important book reviews in 1874 and 1875, he seems to have reversed course, expressing sympathy for interactionism.⁵ And in November of 1878, he offered an extensive defense of interactionism in a six-part lecture series at the Lowell Institute, entitled “The Brain and the Mind,”⁶ the sixth installment of which would become “Are We Automata?”⁷ My central concern will be with this latter essay, along with some physiological evidence James added to enrich his account in the 1890 *Principles of Psychology*.

In a letter to Shadworth Hodgson composed two months after he published “Are We Automata?” James wrote that this essay “was written against the swaggering dogmatism of certain medical materials [*sic*], good friends of mine, here and abroad. I wanted to show them how many empirical facts they had overlooked” (CWJ 1879, 5.44). In the second section, I will explore the vogue for epiphenomenalism against which James was reacting, along with some of the “empirical facts” mentioned in this letter, which were drawn largely from physiology. The third section reconstructs James’s evolutionary hypothesis and offers a further examination of experimental results his account was meant to explain.⁸ The fourth section briefly considers his Darwinian objection to epiphenomenalism.

A final caveat is in order. Although James would later claim that “consciousness,” as the term is *normally* used, names a “nonentity” (ERE 1904, 3) I do not believe he ever backed away from the view that consciousness in this other sense *does* exist. Still, it is enough in the present chapter to focus on James’s positive view of consciousness as he developed it in the earlier works just cited, from about 1872 to 1890. Consult Cooper (this volume) and (Klein 2020) for examinations of how such a view squares with James’s later “pure experience” metaphysics.

Epiphenomenalism and Physiology

Though James seems to have been the first to use the term “epiphenomenon” in its philosophical sense (PP 1890, 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.⁹ 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” (Hodgson 1865, 278).¹⁰ Huxley would offer a famous metaphor: like the relationship between the sound of a steam whistle and an engine, conscious states are always products of bodily states, but conscious states never in turn make a causal difference to bodily states, according to epiphenomenalism (Huxley 1874/1894, 240). This view portrays humans as automata in the sense that all behavior (indeed all bodily motion) 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” (EPs 1879, 38).

Here is how James introduced the topic of epiphenomenalism in the 1879 piece:

The theory itself is an inevitable consequence of the extension of the notion of reflex action to the higher nerve centres. Prof. Huxley starts from a decapitated frog which performs rational-seeming acts although probably it has no consciousness, and passing up to the hemispheres of man concludes that the rationality of their performances can owe nothing to the feelings that co-exist with it. This is the inverse of Mr. Lewes’s procedure. He starts from the hemispheres, and finding their performances apparently guided by feeling concludes, when he comes to the spinal cord, that feeling though latent must still be there to make it act so rationally. Clearly such arguments as these may mutually eat each other up to all eternity. (EPs 1879, 39)

This is a curious passage—James portrayed the debate about epiphenomenalism as stemming from a controversy over decapitated frogs, a controversy that pitted Huxley against G. H. Lewes. Contending that their dispute had become stalemated, James went on to say he would adduce a set of “facts hitherto ignored in the discussion” which “wholly favors the efficacy of Consciousness” (EPs 1879, 40), a promise that resonates with the letter to Hodgson quoted earlier. But what do decapitated frogs have to do with epiphenomenalism?

James was referring to experiments on vertebrates that had been de-cerebrated in various ways, experiments that “occupied the attention of almost all physiologists who lived during the second half of the 19th century” (according to Fearing 1930/1964, 161).¹¹ Sometimes these animals were fully decapitated, and sometimes a more precise procedure called “pithing” was used (this is a technique in which experimentalists use a blunt

needle to destroy a creature's brain in whole or in part, leaving the spinal cord intact; for a description, see Huxley 1872, 54–55). Physiologists pithed fish, birds, turtles, and dogs, but the common frog was the most popular vertebrate for this purpose (Fearing 1930/1964, 166).

One aim of such experiments was to get a grip on which brain structures enabled which specific behaviors. For instance, Huxley reported that the cerebellum is necessary for jumping, since frogs that are pithed below¹² this brain structure cannot jump (Huxley 1870, 3–4).

The preoccupation with these animals was not simply due to an interest in correlating brain structures with physiological capacity, though. In an 1853 book, Eduard Pflüger had reported a particularly controversial experiment on pithed frogs. It seemed to establish that pithed frogs exhibit *purposive behavior*, even when they have been pithed at the bottom of the brain stem.¹³

The experiment involved dripping acid on the knee of a living, decapitated frog (Pflüger 1853, 16). Such a frog reliably wipes the acid away, even though it lacks a brain. By itself, this result would not have been surprising—it was already well known that many vertebrates not only survive, but exhibit *reflex action* even after being decapitated, and the simple acid-dripping case seems but one such example.¹⁴ What surprised Pflüger's readers is what happens when one amputates whatever foot the decapitated frog habitually uses for wiping away the acid. If the acid-wiping behavior were a mere reflex performed non-consciously, one might expect the amputated frog to wave its stump around helplessly in response to the irritant. But that is not what happens.

The amputated frog actually chooses another limb to try to wipe the acid away. Or if a suitable foreign surface is nearby, the frog may maneuver its body to wipe the irritated skin against the surface. If one accepts such choosing-behavior as an instance of purposiveness, and if one thinks purposiveness is a mark of consciousness, then one can infer from this experiment that the decapitated frog is somehow conscious. This is just the sort of view one finds in Lewes, Pflüger's most visible British champion (e.g., at Lewes 1877, 427–30).¹⁵

But Pflüger's experiment created a curious theoretical dilemma. If such purposive behavior is a mark of consciousness, then the brain cannot be the sole organ that gives rise to conscious experience (because the brainless frog apparently acts with purpose). This is Pflüger and Lewes's position—they argued that the *spinal cord* must also produce some measure of consciousness by itself (e.g., at Pflüger 1853, 123–26; Lewes 1859; 1859–1860, vol. II, ch. ix, sec. 3; 1873). On the other hand, if purposive behavior is *not* a mark of consciousness (as critics like Huxley had contended), then the way is open for regarding conscious experience as epiphenomenal. For, suppose one assumes brainless frogs obviously *cannot* be conscious even though they still *seem* to act with purpose. It follows that purposiveness alone cannot establish conscious control of behavior even in *intact* vertebrates—indeed, even in intact humans (Huxley 1874/1894, 222–226).

But as the debate was then playing out, it turned on whether or not purposiveness was an acceptable mark of consciousness. And that was a question that had to be answered *before* one could interpret any of these experiments, it seemed, and thus was not a question the experiments themselves could settle (Klein 2018). This is the apparent stalemate to which James was referring in the “decapitated frog” passage reproduced above.

James’s intervention in the debate worked like this. He rejected epiphenomenalism because he thought it highly implausible that consciousness could have evolved through any known mechanism if epiphenomenalism were true (as we shall see later, in the fourth section). But he also rejected the Lewes/Pflüger contention that de-cerebrated vertebrates were fully conscious. James’s crucial insight was that although de-cerebrated vertebrates were in fact capable of purposive behavior, they differed from their cerebrated peers in a subtle way that suggested a lack of fully fledged consciousness: they were apparently unable to *evaluate*. I now turned to James’s positive account.

James’s Evolutionary Account

There is no one passage that lays out James’s own explanation of the pithing results. But toward the end of “Are We Automata?” we get what comes close to a summary of his argument:

We have found that the unaided action of the cerebral hemispheres would probably be random and capricious; that the nerve-process likely to lead to the animal’s interests would not necessarily predominate at a given moment. On the other hand, we have found that an impartial consciousness is a nonentity, and that of the many items that ever occupy our mental stage Feeling always selects one as most congruous with the interests it has taken its stand upon. Collating these two results, an inference is unavoidable. The “items” on the mental stage are the subjective aspects of as many nerve-processes, and in emphasising the representations congruous with conscious interest and discouraging all others, may not Attention actually reinforce and inhibit the nerve-processes to which the representations severally correspond?

This of course is but a hypothetical statement of the verdict of direct personal feeling. ... (EPs 1879, 52)

James suggests we should be surprised that vertebrates with healthy cerebral hemispheres in fact behave in ways “likely to lead to the animal’s interests.” To explain this capacity for interested action, he makes what he characterizes as a “hypothetical statement”¹⁶ or an abductive “inference”: that consciousness (which always “selects” what accords with the animal’s “interests”) may steer the animal to behave in profitable ways by “reinforc[ing]” and “inhibit[ing]” mechanical brain processes. Two pages later he makes clear that he means to be offering an *adaptive* explanation, explicitly suggesting that consciousness likely had a positive survival value (EPs 1879, 54).

This passage does not give us James's entire account of consciousness, but it does suggest at least the general structure of his reasoning. If we are willing to fill in some gaps by looking outside of "Are We Automata?"—in particular, to his treatment of relevant physiological material in the *Principles*—then we find something like the following line in James:

1. Intact, healthy vertebrates are capable of "prudent" behaviors.
2. (1) is a surprising fact that demands an explanation, since *de-cerebrated* vertebrates are incapable of prudence.
3. Phenomenological Claim: consciousness typically involves engaging in a (nonphysical) process of evaluation.
4. Quasi-Mechanistic Hypothesis: phenomenal consciousness produces prudent behaviors *by* enabling evaluation.
5. Adaptive Hypothesis: phenomenal consciousness is an adaptation for producing prudent behavior.

As I unpack these steps in turn, I will offer textual bases for attributing each to James.

A word on the modal force of James's argument is immediately called for. James did not offer (and did not intend to offer) reasons for thinking his brand of interactionism is *necessarily* true. If my reading is roughly accurate, James's argument was fundamentally empirical—he was proposing a hypothesis for accounting for some surprising physiological observations, not offering conceptual reasons for thinking interactionism *must* be right. And as we will see in the penultimate section of this paper, his attack on epiphenomenalism also was not purely conceptual—he did not take himself to have shown automaton theory to be incoherent, but rather empirically implausible.

Also, we will see that James's interactionism only does explanatory work in steps 4 and 5, where consciousness is part of the *explanans*, not the explanandum. As I have already suggested, James is not explaining consciousness itself. He is suggesting that *if* a particular form of interactionist dualism were true, it *would* explain the surprising physiological results identified in step 1.

Step 1: Intact, Healthy Vertebrates Are Capable of "Prudent" Behaviors

Let us begin with this term "prudence." I draw the term from a passage in the *Principles* where James was fleshing out the notion of stable, profitable behavior. He wrote that "no animal without" cerebral hemispheres "can deliberate, pause, postpone, nicely weigh one motive against another, or compare. *Prudence*, in a word, is for such a creature an impossible virtue" (PP 1890, 33, my italics).

As James used the concept, genuinely *prudent* behavior has two components. First, it involves making choices of a robustly purposive cast (something Pflüger's frogs can do). And second, it involves making those choices by considering information other than what is directly presented by sensory stimuli at the time of choosing (something Pflüger's frogs

cannot do). The debate over epiphenomenalism had largely relied on the first component as a mark of consciousness; one of James's key (albeit rarely noticed) innovations was the addition of the second. I will take each component in turn.

First, consider the faux, purposeless brand of "choice" a magnet might be said to make when it attracts iron but not brass filings (an example considered at James 1879, 8.n). The magnet's "choice" is produced by what James called a "*vis a tergo*," or a force from behind. The magnet happens to achieve some end, but only accidentally—only because of causal factors that determine its course of actions without anything like consideration of where those actions might lead.

In contrast, suppose one submerges a frog in water, and then traps it under a glass bell when it tries to surface. The frog "will not ... perpetually press his nose against its unyielding roof, but will restlessly explore the neighborhood until by re-descending again he has discovered a path round its brim to the goal of his desires" (PP 1890, 20). The *frog's* end is achieved by what James called a "*vis a fronte*"—a force operating, so to speak, from the front. Unlike in the magnet case, there is an "ideal purpose presiding over the [frog's] activity from its outset and soliciting or drawing it into being" (PP 1890, 21). Thus James defined *vis a fronte* choosing as "[t]he pursuance of future ends and the choice of means for their attainment" (PP 1890, 21). One necessary component of prudent behavior is the exercising of *vis a fronte* choice—that is, the agent must be capable of trying different means to achieve an end she has in view.

But one more condition must be met as well. For James explicitly noted that de-cerebrated frogs often exhibit such *vis a fronte* choosing, as Pflüger had clearly established in his acid-drip experiment. Indeed, James noted that Goltz had shown something even more startling, that the task of escaping the inverted glass bell when emerging for air can also be performed by a de-cerebrated frog (PP 1890, 22).¹⁷ So here is where we find James's distinctive intervention into the Huxley-Lewes debate. James contended that the capacity for *vis a fronte* choosing is necessary, *but not sufficient*, for the kind of prudence one only finds in vertebrates with intact cerebral hemispheres. Such prudence *also* involves a capacity to make choices that are not prompted by present sensory stimuli, he held, resting this second claim on a collection of pithing results I will now examine.¹⁸

As physiologists had become more adept at pithing, their awareness of which specific brain structures were responsible for which behaviors blossomed. Figure 1 gives a schematic diagram of a frog's brain from the era, along with a description of what behaviors are lost when each structure is knocked out through pithing.



Figure 1. Diagram of a frog brain, with a summary of the effects of removing various structures (from Brunton 1898, 227).

What is most important for James's account of consciousness is what happens when the cut is made just below the cerebral hemispheres. In such cases, "an unpractised observer" might fail to notice anything unusual in the animal. A de-cerebrated frog can react to *present* stimuli largely in the way an intact frog reacts. It can even exhibit *vis a fronte* choice, e.g., managing to navigate by sight around obstacles, as Goltz (1869, 65) had shown. But what a de-cerebrated frog typically does *not* do, crucially, is *initiate* any genuine behavior that is not incited by some sensory stimulus.

Thus Goltz's de-cerebrated frog moves reflexively toward a light source, even hopping around a book placed in its path. But in this case the light source acts as a stimulus. *Without* any such sensory incitement, the frog will simply sit, noiseless and motionless. James saw the de-cerebrated frog's quiescence as evidence of an inability to consider anything but what is directly presented in sensation.

James offered an example to help crystalize his point about this second requirement for prudence:

If I step aside on seeing a rattlesnake, from considering how dangerous an animal he is, the mental materials which constitute my prudential reflection are images more or less vivid of the movement of his head, of a sudden pain in my leg, of a state of terror, a swelling of the limb, a chill, delirium, unconsciousness, etc., etc., and the ruin of my hopes. But all these images are constructed out of my past experiences. They are *reproductions* of what I have felt or witnessed. They are, in short, *remote* sensations; and the *difference between the hemisphereless animal and the whole one* may be concisely expressed by saying that the *one obeys absent, the other only present, objects*. (PP 1890, 32; original italics)

This example of what James here calls "prudential reflection" is construed in terms of a special *kind* of *vis a fronte* choice, one that specifically employs *remembered*, and not merely presented, stimuli. It is not a choice between two presented objects, but a choice made partly by considering "absent" (remembered) objects. This capacity to consider absent objects is the second component of the sort of prudence James thought was distinctive to vertebrates with intact cerebral hemispheres.¹⁹

Now the word "prudence" did not appear in "Are We Automata?" But the concept is prefigured in that article nonetheless:

The addition of the cerebral hemispheres immediately introduces a certain incalculableness into the result, and this incalculableness attains its maximum with the relatively enormous brain-convolutions of man. In the beheaded frog the legs twitch as fatally when we touch the skin with acid as do a jumping-jack's when we pull the string. ... Even if all the centres above the cord except the cerebral hemispheres are left in place, the machine-like regularity of the animal's response is hardly less striking. ... If I pinch [a de-cerebrated frog] ... under the arm-pits, he will croak once for each pinch; if I throw him into water, he will swim until I touch his hands with a stick, when he will immediately stop. Over a frog with an entire

brain, the physiologist has no such power. The signal may be given, but ideas, emotions or caprices will be aroused instead of the fatal motor reply, and whether the animal will leap, croak, sink or swim or swell up without moving, is impossible to predict. In a man's brain the utterly remote and unforeseen courses of action to which a given impression on the senses may give rise, is too notorious to need illustration. (EPs 1879, 41–42)

Again, we see James contrasting the de-cerebrated frog, which reacts with “machine-like regularity” to stimuli, with intact vertebrates, which act from “remote and unforeseen” considerations. This is just the second condition of prudence that we have been discussing.

It may seem as though James believed the introduction of the hemispheres somehow broke apart the strict causal determinism that might be thought to govern a purely physical being. So before moving on, we must ask whether James was rejecting the very idea of a fully mechanistic physiology.

The answer is no, at least if one takes any physiological program that aims at lawlike generalizations about animal activity to count as mechanistic. James's interactionism would obviously violate a stronger form of mechanistic methodology according to which only *physical* mechanisms can appear in the explanans. But he saw his interactionism as fully compatible with a weaker form of mechanistic physiology.

One can see this in an 1879 letter to the psychologist J. J. Putnam. “Are We Automata?” had just been published:

... I did not pretend in my article to say that when things happen by the intermediation of consciousness they do not happen by law. The dynamic feelings which the nerve processes give rise to, and which enter in consciousness into comparison with each other and are selected, may in every instance be fatally selected. All that my article claims is that this additional stratum which complicates the chain of cause and effect also gives it determinations not identical with those which would result if it were left out. If a hydraulic ram be interposed on a water-course, a pendulum and escapement on a wheel-work the results are altered but still obey the laws of cause and effect. Free-will is in short, no necessary corollary of giving causality to consciousness. (CWJ 5.34)

James held that mental causes may permissibly appear in a strictly deterministic causal chain, and so appealing to consciousness need not undermine the goal of validating law-like generalizations in physiology. However, James did not take consciousness to be a physical phenomenon, as we shall see in step 3, so again his interactionism obviously would violate the stronger form of mechanism mentioned earlier, a form of mechanism that Huxley himself perhaps favored (Greenwood 2010, 277).

There is evidence in the *Principles* that James saw the science of physiology as rightly aiming to model all bodily motion in terms of reflex action, by the way: “[t]he conception

of *all* action as conforming to this type ['reflex action'] is the fundamental conception of modern nerve-physiology" (PP 1890, 35). So it seems that James did advocate a weakly mechanistic methodology for physiology, and apparently saw his own view of consciousness as compatible with this approach.²⁰

Step 2: (1) Is a Surprising Fact That Demands an Explanation, Since De-cerebrated Vertebrates Are Incapable of Prudence

Why did James think we should be surprised that vertebrates with cerebral hemispheres exhibit behavioral prudence? Put simply, the answer is that vertebrates that *lack* cerebral hemispheres only exhibit "hair-trigger" (EPs 1879, 43) responses to stimuli. For example, we have seen that a pinch under the de-cerebrated frog's armpits produces a croak "as fatally ... as do a jumping-jack's when we pull the string" (EPs 1879, 41).

So again, even though such frogs are capable of *vis a fronte* choosing, their behavior is characterized by a distinctive quiescence—by an apparent inability to act in response to anything *but* immediately presented stimuli. Goltz had reported that a de-cerebrated frog will sit in a warm bath with slowly increasing temperature even until it is killed by the heat (Danziger 1980, 100); and James takes this kind of case to show that de-cerebrated frogs have a diminished capacity for *spontaneous* action, action not directly generated by presented stimuli.

The discovery that the cerebellum is crucial for jumping prompts the question of just how the cerebellum accomplishes this trick. Similarly, James is suggesting that the hemispheres are crucial for enabling spontaneous action—the outward manifestation of the second aspect of prudence. What is apparently needed is an explanation of how they do this—of how the hemispheres help produce genuine prudence.

Step 3. Phenomenological Claim: Consciousness Typically Involves Engaging in a (Nonphysical) Process of Evaluation

James is going to propose that the hemispheres give rise to consciousness, and that consciousness in turn is a key factor in producing behavioral prudence. To start to understand his proposal, we first have to explore some phenomenological considerations he offered.

James claimed that where we find consciousness, we typically find unrelenting evaluation, and vice versa—where we find genuine evaluation, we typically find consciousness (EPs 1879, 46–51). He made this claim largely on the basis of introspection, further arguing that genuine evaluation cannot itself be a physical process, and that therefore consciousness cannot be physical either. Hence we get the main assertion of step 3.

Let us first consider James's claim that wherever we find consciousness we find evaluation.

Looking back then over this review we see that the mind is at every stage a theatre of simultaneous possibilities. *Consciousness consists in the comparison of these with each other, the selection of some, and the suppression of the rest* by the reinforcing and inhibiting agency of Attention. The highest and most elaborated mental products are filtered from the data chosen by the faculty next beneath out of the mass offered by the faculty below that, which mass in turn was sifted from a still larger amount of yet simpler material, and so on. (EPs 1879, 51, my italics)

The italicized claim is quite strong. It suggests that one is phenomenally conscious *in virtue of* comparing “simultaneous possibilities” and continually using attention to select some for the basis of action. Suppose one allows this kind of selective attending to various possibilities, and perhaps to various aspects of the sensed environment, to amount to a form of evaluation (more on this point later). Then James’s claim seems to be that “[c]onsciousness consists in” evaluating.

There are other passages that suggest a weaker connection between phenomenal consciousness and evaluation, though:

There are a great many things which consciousness *is* in a passive and receptive way by its cognitive and registrative powers. But there is one thing which it *does*, *suâ sponte*, and which seems an original peculiarity of its own; and that is, always to choose out of the manifold experiences present to it at a given time some one for particular accentuation, and to ignore the rest. (EPs 1879, 46)

Here, James might be read as claiming that active choosing is *typical* of, but *not* essential to, consciousness. Evaluation is the only thing conscious states²¹ do “*suâ sponte*”—of their own accord²²—but they are also capable of passively registering a stimulus, according to this passage. So does James think we are conscious *in virtue of* evaluating, or not?

What James had in mind, I suggest, was that phenomenal consciousness is not *essentially* evaluative in any metaphysical sense, but that consciousness is *normally* evaluative, and indeed has evaluation as a proper etiological function.²³ This claim is compatible with the occurrence of the occasional conscious state that is not engaged in evaluation.²⁴

Now, the important passage I have been discussing from (EPs 1879, 51) mentions “selection” and “comparison,” but I have glossed these expressions as indicating a form of *evaluation*, which might seem stronger. Did James in fact think “selection” and “comparison” really amounted to a kind of *evaluation*? The answer is yes, and we can see this in nearby passages:

Good involves the notion of less good, necessitates comparison, and for a drop of water either to compare its present state with an absent state or to compare its total self with a drop of wine, would involve a process not commonly thought of as physical. Comparison requires a *tertium quid*, a *locus*—call it what you will—in which the two outward existences may meet on equal terms. This forum is what is known as a consciousness. (EPs 1879, 43, my underline)

Consciousness involves a “comparison” between items, and here James construes this sort of comparison as issuing in distinctions between the “good” and “less good.” As he would put it in the *Principles*, “consciousness” establishes value by “decree” (PP 1890, 144). Thus consciousness “evaluates” in the sense of *actively imposing* divisions between the more and less valuable *onto* the environment, for James. It would seem that evaluation can involve a comparison either between different things presented simultaneously in the environment or, crucially, between what is presented and “an absent state,” as when I contemplate jumping over the snake I see or turning around and planning some other route through the woods.

The underlined sentence gives an important clue about why James held that an evaluating consciousness must be non-physical. What does the work, for James, is a tacit principle to the effect that value relations (better and worse, more or less important, and so on) are not physical relations.²⁵ Elaborating on this central idea, he wrote that *better* and *worse* in no way “pre-exist” a “consciousness [that] only discerns them” (EPs 1879, 46). Instead, consciousness

in declaring ... superiority ... simply creates what previous to its fiat had no existence. The judge makes the law while announcing it: if the judge be a maggot, the suicide’s brain will be best; if a king, the chancellor’s. (EPs 1879, 45)

So, consciousness creates “by fiat” distinctions between “good” and “less good,” and in that sense makes law. This is the process I am calling “evaluation.”²⁶ But consciousness also enforces law in the sense of carrying out bodily action based on the “good” and “less good” divisions *it itself* has declared. Different consciousnesses (like the maggot’s as compared with the king’s) will harbor different interests, which is to say that each might assess the value of environmental conditions and possibilities—and then *act* on those conditions and possibilities—differently.²⁷

In short, James apparently held that to be *phenomenally conscious* of some object *o* (whether *o* is presented in an occurrent sensation or represented in memory) with respect to a background *b* (whether *b* is presented in an occurrent sensation or represented in memory) typically involves *valuing* *o* with respect to *b*. To *value* means (1) to create an evaluation according to which *o* deserves more attention than *b*, and (2) to attend to *o* and ignore *b* (or make *b* peripheral) accordingly. As James would argue in the 1880 “Feeling of Effort” (reprinted in EPs), attending to *o* will then naturally trigger some bodily response or other, all things being equal (Klein forthcoming, ch. 5).

This neat formula undoubtedly puts James’s position more abstractly than he would have liked, so perhaps it is better simply to say that for James, phenomenal consciousness typically involves *valuing*.

Step 4. Quasi-Mechanistic Hypothesis: Phenomenal Consciousness Produces Prudent Behaviors by Enabling Evaluation

We can see a happy synergy now between step 3 and the two that preceded it. If James is right that healthy, intact vertebrates exhibit a surprising ability to behave with prudence, and if we accept his claim that phenomenal consciousness typically involves evaluation, then it is natural to entertain the following hypothesis. Perhaps consciousness is what *produces* the cerebrated vertebrate's prudent choices, and perhaps it does so precisely by enabling evaluation.

We can get some purchase on the process James envisioned by briefly examining how the hemispheres might provide a locus for the supposed causal link between non-physical consciousness and prudent behavioral patterns. He gives evidence that in humans at least, the hemispheres alone produce consciousness (PP 1890, 74). He also argues that the hemispheres play a key causal role in producing tendencies to react to physical stimuli in habitual (PP 1890, 112) or instinctive (PP 1890, 32) ways, not all of which actually benefit the organism.²⁸ He was proposing that consciousness then *selects* which of these temptations, if any, to act on or to perceptually focus on.

Sometimes this selection might involve foregrounding and backgrounding various aspects of the “blooming, buzzing confusion” that our sense-organs constantly register; and sometimes this selection might involve picking which of several reaction-temptations to actualize given a particular stimulus. Either way, Jamesian consciousness acts as a kind of *filter* or *gate* on our barrage of sensory stimulation, channeling an organism's active responses in *prudent* directions.²⁹

I see no indication that James took himself to have directly *proved* that consciousness enables prudence. Recall the passage quoted earlier: this is a “hypothetical statement” only. Together with the adaptive claim I shall consider next, this statement is being recommended because it *would* help explain the physiological surprise we discussed in steps 1 and 2.

Step 5. Adaptive Hypothesis: Phenomenal Consciousness Is an Adaptation for Producing Prudent Behaviors

We can now see why James would suspect consciousness to be an adaptation.³⁰ *Prudent behavior* would likely help an organism survive and reproduce, and if he is right that consciousness is what *produces* this behavioral trait, then it stands to reason that consciousness might *originally* have proliferated because it produced this helpful behavior. In other words, it stands to reason that consciousness is an adaptation for producing prudence.

We can anchor this last step in the text. First, it is clear that James thinks *consciousness itself* is an adaptation (in the Darwinian sense), a trait that produces differential reproductive success by producing prudence:

[G]ive to consciousness the power of exerting a constant pressure in the direction of survival, and give to the organism the power of growing to the modes in which consciousness has trained it, and the number of stray shots³¹ is immensely reduced, and the time proportionally shortened for Evolution. It is, in fact, hard to see how without an effective superintending ideal the evolution of so unstable an organ as the mammalian cerebrum can have proceeded at all. (EPs 1879, 54)

James was suggesting that if consciousness has physical efficacy, then it might evolve as just one more, functional part of an organism's *Bauplan* (very loosely construed). And this passage need not be read in a Lamarckian fashion, by the way.³² He may simply mean that the organism *grows* to fit consciousness in the way that eyelids grow to fit eyeballs—in other words, the sense of “growing” here would be phylogenetic, not ontogenetic. Indeed, this fits with James's claim that consciousness “has been slowly evolved in the animal series, and resembles in this all organs that have a use” (EPs 1879, 40–41).

The next question is *why* James expected prudence to have survival value. We get a hint in the earlier suggestion that the cerebrum is an inherently “unstable ... organ.” James expected there to be an evolutionary trade-off between perceptual acuity and behavioral stability, and thus that a more articulated brain *should* simply produce a more varied, but less stable, array of perceptual reactions. He described nature's “dilemma” this way:

[A] high brain may do many things, and may do each of them at a very slight hint. But its hair-trigger organization makes of it a happy-go-lucky, hit-or-miss affair. It is as likely to do the crazy as the sane thing at any given moment. A low brain does few things, and in doing them perfectly forfeits all other use. The performances of a high brain are like dice thrown for ever on a table. Unless they be loaded, what chance is there that the highest number will turn up oftener than the lowest? (EPs 1879, 43)

James associated prudence with behavioral regulation—with the ability to stabilize the superabundance of “hair-trigger” behaviors that might otherwise result from a highly articulated cerebrum.³³ As I read him, James thinks that a creature with a capacity for prudence is adaptively better off *because* it is better able to regulate its own behavior.

Let me illustrate the trade-off James envisioned between perceptual acuity and behavioral stability. On the one hand, some organisms survive and reproduce in relatively simple environments. For instance, yeast is a single-celled organism that metabolizes organic compounds in its environment. A yeast cell floating in a solution relatively homogenous with respect to nutrients has a limited need for either mobility or perceptual acuity—it can simply float free, taking up whatever nutrients happen to come into contact with its cell wall.³⁴

But contrast this to the case of a field mouse that, let us suppose, eats a leaf whose characteristic pattern is mimicked by the skin of a local snake. If these mice react automatically and reflexively to occurrent stimuli—much like the armpit pinch elicits a croak with-

out fail in the de-cerebrated frog—the mice might too quickly pursue every leaf-like color pattern they see, sometimes falling prey to the snake’s mimicry.³⁵

In other words, mice must have a capacity to react to a wide range of stimuli if they are to avoid predators and hunt for food. But if they respond instantly and unfailingly to every stimulus, they will go off half-cocked, so to speak, at every suggestion. In contrast, an organism perceptually attuned to a narrower band of stimuli might react “infallibly and certainly,” but only at the cost of not being able to handle much environmental complexity, James would say (EPs 1879, 42–43). In short, James thought one would expect vertebrates with complex cerebral hemispheres simply to show ever more varied behavioral capacities, but to show increasing behavioral instability as well.

So we arrive at the second of James’s conjectures. Suppose that the hemispheres give rise to an evaluating consciousness. Suppose such a consciousness in turn enables prudent behavior—that is, *purposive* behavior that takes account of *remote sensations* (like the mouse considering what *would* happen if it jumped at the color pattern). This kind of prudence would amount to a form of behavior regulation that might have a positive survival value, particularly for creatures with high perceptual acuity. Hence consciousness may have proliferated among ancestral vertebrates in virtue of performing this kind of regulating function, James was proposing—in virtue, that is, of enabling prudence.

Against Epiphenomenalism

We have just seen that James developed an interactionist conception of consciousness that would explain some puzzling physiological experiments. Thus his defense of his own interactionism has the form of an argument to the best explanation. Such arguments can be strengthened by demonstrating that *competing* explanations of the same phenomenon are likely to be false. And that is exactly what James did—he strengthened his case by providing independent reasons for rejecting Huxley-style epiphenomenalism.

It is useful to distinguish two related objections James raised. I will call the first objection “broad” and the second “narrow.” The broad objection targets the notion that any conscious states could have been shaped by selection if epiphenomenalism were true. The narrow objection targets the notion that epiphenomenal pleasures and pains associated with “fundamental vital processes” could have been shaped by selection. The latter issue is now the more widely discussed.³⁶

Here is the broad objection. Automaton theory depicts consciousness as something that could make no causal difference to an organism’s behavior. But then consciousness “would be useless” from an evolutionary standpoint, in that it could not itself have contributed to any animal’s reproductive success, and so could not have evolved via natural selection (EPs 1879, 41; also see PP 1890, 142).

The usual epiphenomenalist response (especially more recently; see, e.g., Jackson 1982) has been to speculate that consciousness could be an evolutionary by-product—a “span-

drel" in the sense of Gould and Lewontin (1979)—rather than a trait that has been directly selected. As I read his narrow objection, however, James already forestalled this response by pointing to specific aspects of our conscious experience that bear hallmarks of natural selection's handiwork.³⁷ The particular band of experience in question is a sub-set of our phenomenal pleasures and pains—those associated with our "fundamental vital processes," as we shall see.

In "Are We Automata?," after considering some of the evidence concerning pithed vertebrates we have just discussed, James writes that there is "yet another set of facts which seem explicable on the supposition that consciousness has causal efficacy" and not explicable by epiphenomenalism (EPs 1879, 55; PP 1890, 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. He writes:

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. ... 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 ... why the most noxious acts, such as burning, might not give thrills of delight, and the most necessary ones, such as breathing, cause agony. (PP 1890, 146-147, italics original)³⁸

Consider a sub-set of our phenomenal pleasures and pains—namely, those that natively go along with "fundamental vital processes." James's examples of vital pains include the experiences of being burnt, suffocated, or physically injured; his examples of vital pleasures include eating, drinking, and resting. Epiphenomenalists accept that these life-essential, phenomenal pleasures and pains are *effects* of the beneficial and harmful brain states with which they are natively associated. But epiphenomenalists *cannot* consistently say that they are *adapted* effects of those beneficial and harmful brain states. In other words, if epiphenomenalism were true, there can have been no selection pressure on any of the underlying bodily states to produce the particular, life-essential phenomenal pleasures and pains with which they are associated. This is because if epiphenomenalism were true, such pleasures and pains could have no "efficacy" and so (recall James's broad objection) could have made no difference to reproductive success.

But these phenomenal pleasures and pains have three features that *do* suggest that they were shaped by selection, for James: they are natively patterned³⁹ (they have a characteristic "*distribution*," he says); those patterns are systematically linked with underlying bodily states (this is the "*well-known fact*"); and the patterns are "universal" among humans. James concludes that epiphenomenalism cannot make sense of how our native distribution of life-essential, phenomenal pleasures and pains (with their systematic connections to underlying bodily states) could have evolved.

Concluding Remarks

James wrote such colorful descriptions of experience that he has earned a reputation as more poetic than empirical. But his literary flair can distract contemporary readers from the fertile, scientific context of his actual research.

I have argued that James sought to explain differences between the observed behaviors of intact versus pithed vertebrates, differences that had been recorded in a series of controversial experiments by Pflüger, Lewes, Goltz, and others. James suggested that what pithed vertebrates really lacked was a capacity for *prudent* behavior. He hypothesized that if the hemispheres give rise to phenomenal consciousness, and if consciousness is fundamentally an evaluating agency, then it would make sense of how intact vertebrates (but not their de-cerebrated peers) achieve prudent behavior. He further contended that since prudent behavior is stable behavior, and stable behavior might have positive survival value, consciousness might be an adaptation for this kind of behavior regulation.

James's hypothesis relied on an intriguing observation about an aspect of phenomenal consciousness that is rarely emphasized in contemporary discussions: when one is phenomenally conscious, James contended, one is continuously *evaluating* what is in one's environment, typically. In effect, he suggested that consciousness evolved as an appraisal system that helped our ancestors sift through the "blooming, buzzing confusion" of environmental details with which our perceptual apparatus puts us in touch.

In James's view, Huxley-style epiphenomenalism was the main rival to some form of interactionism like his own. But James argued that epiphenomenalism is likely false because it cannot give a satisfactory evolutionary explanation of phenomenal pleasures and pains.

So, James developed an account according to which *evaluating* is the proper etiological function of phenomenal consciousness. And he supported his account by arguing that it provided the best available explanation of some surprising experimental results in physiology.⁴⁰

Bibliography

Brunton, T. Lauder. 1898. *Lectures on the Action of Medicines: Being the Course of Lectures on Pharmacology and Therapeutics Delivered at St. Bartholomew's Hospital During the Summer Session of 1896*. London: Macmillan.

Capek, Milic. 1954. "James's Early Criticism of the Automaton Theory." *Journal of the History of Ideas* 15: 260–279.

Carruthers, Peter. 1998. "Natural Theories of Consciousness." *European Journal of Philosophy* 6 (2):203–222.

Carruthers, Peter. 2001. "Consciousness: Explaining the Phenomena." *Royal Institute of Philosophy Supplement* 49: 61–85.

- Chalmers, David. 1997. "Facing Up to the Problem of Consciousness." In *Explaining Consciousness: The "Hard Problem,"* edited by Jonathan Shear, 9–32. Cambridge, MA: MIT Press.
- Clifford, William Kingdon. 1874/1886. "Body and Mind." In *Lectures and Essays, by the Late William Kingdon Clifford*, edited by Leslie Stephen and Frederick Pollock, 244–273. London: Macmillan.
- Corabi, Joseph. 2008. "Pleasure's Role in Evolution: A Response to Robinson." *Journal of Consciousness Studies* 15 (7): 78–86.
- Corabi, Joseph. 2014. "The Misuse and Failure of the Evolutionary Argument." *Disputatio: International Journal of Philosophy* 6 (39):199–227.
- Danziger, Kurt. 1980. "The Unknown Wundt: Drive, Apperception, and Volition." In *Wilhelm Wundt and the Making of a Scientific Psychology*, edited by R. W. Rieber and David K. Robinson, 95–120. New York: Plenum Press.
- Dennett, Daniel. 2001. "Are We Explaining Consciousness Yet?" *Cognition* 79 (1–2): 221–237.
- Fearing, Franklin. 1930/1964. *Reflex Action: A Study in the History of Physiological Psychology*. New York: Hafner.
- Flanagan, Owen. 1997. "Consciousness as a Pragmatist Views It." In *The Cambridge Companion to William James*, edited by Ruth Anna Putnam, 25–48. New York: Cambridge University Press.
- Goltz, Friedrich Leopold. 1869. *Beiträge Zur Lehre Von Den Functionen Der Nervencentren Des Frosches*. Berlin: A. Hirschwald.
- Gould, Stephen Jay, and Richard C. Lewontin. 1979. "The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme." *Proceedings of the Royal Society of London, Series B* 205 (1161): 581–598.
- Greenwood, John. 2010. "Whistles, Bells, and Cogs in Machines: Thomas Huxley and Epiphenomenalism." *Journal of the History of the Behavioral Sciences* 46 (3): 276–299.
- Hodgson, Shadworth Hollway. 1865. *Time and Space: A Metaphysical Essay*. London: Longman, Green, Longman, Roberts, and Green.
- Hodgson, Shadworth Hollway. 1870. *The Theory of Practice: An Ethical Enquiry*. 2 vols. London: Longmans, Green, Reader, and Dyer.
- Huxley, Thomas Henry. 1870. "Has a Frog a Soul? And of What Nature Is That Soul, Supposing It to Exist?" *Papers Read before the Metaphysical Society* (Privately Published): 1–7.

- Huxley, Thomas Henry. 1872. *Lessons in Elementary Physiology*. 6th ed. London: Macmillan and Co.
- Huxley, Thomas Henry. 1874/1894. "On the Hypothesis That Animals Are Automata, and Its History." In *Collected Essays: Method and Results*, 199–250. New York: Appleton.
- Jackson, Frank. 1982. "Epiphenomenal Qualia." *Philosophical Quarterly* 32 (127): 127–136.
- James, William. 1879. "Are We Automata?" *Mind* 4 (13): 1–22.
- Klein, Alexander. 2016. "Was James Psychologistic? ." *Journal for the History of Analytical Philosophy* 4 (5): 1–21.
- Klein, Alexander. 2018. "The Curious Case of the Decapitated Frog: On Experiment and Philosophy." *British Journal for the History of Philosophy* 26 (5): 890–917.
- Klein, Alexander. 2019. "William James's Objection to Epiphenomenalism." *Philosophy of Science* 86 (5): 1179–1190.
- Klein, Alexander. 2020. "The Death of Consciousness? James's Case against Psychological Unobservables." *Journal of the History of Philosophy* 58 (2): 293–323.
- Klein, Alexander. Forthcoming. *Consciousness Is Motor: Warp and Weft in William James*. New York: Oxford University Press.
- Lewes, George Henry. 1859. "The Spinal Chord a Sensational and Volitional Centre." In *Report of the Twenty-Eight Meeting of the British Association for the Advancement of Science; Held at Leeds in September of 1858*, 135–138. London: John Murray.
- Lewes, George Henry. 1859–1860. *The Physiology of Common Life*. 2 vols. Edinburgh: W. Blackwood.
- Lewes, George Henry. 1873. "Sensation in the Spinal Cord." *Nature* 9: 83–84.
- Lewes, George Henry. 1877. *Problems of Life and Mind, Second Series: The Physical Basis of Mind*. London: Trübner & Co.
- Lloyd, Elisabeth A., and Stephen Jay Gould. 2017. "Exaptation Revisited: Changes Imposed by Evolutionary Psychologists and Behavioral Biologists." *Biological Theory* 12 (1): 50–65.
- Mattson, Mark P. 2014. "Superior Pattern Processing Is the Essence of the Evolved Human Brain." *Frontiers in Neuroscience* 8: 265.
- Myers, Gerald E. 1986. *William James: His Life and Thought*. New Haven, CT: Yale University Press.

Perry, Ralph Barton. 1935. *The Thought and Character of William James*. 2 vols. Cambridge, MA: Harvard University Press.

Pflüger, Eduard. 1853. *Die Sensorischen Functionen Des Rückenmarks Der Wirbelthiere, Nebst Einer Neuen Lehre Über Die Leitungsgesetze Der Reflexionen*. Berlin: Hirschwald.

Price, Elfed Huw. 2014. "George Henry Lewes (1817–1878): Embodied Cognition, Vitalism, and the Evolution of Symbolic Perception." In *Brain, Mind and Consciousness in the History of Neuroscience*, edited by C. U. M. Smith and Harry Whitaker, 105–123. Dordrecht: Springer.

Reck, Andrew J. 1972. "Dualisms in William James's *Principles of Psychology*." In *Knowledge and Value: Essays in Honor of Harold N. Lee*, edited by Andrew J. Reck, 23–38. Dordrecht: Springer Netherlands.

Robinson, William S. 2007. "Evolution and Epiphenomenalism." *Journal of Consciousness Studies* 14 (11): 27–42.

Robinson, William S. 2014. "James's Evolutionary Argument." *Disputatio: International Journal of Philosophy* 6 (39): 229–237.

Robinson, William S. 2015. Epiphenomenalism. *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. Fall. Stanford, CA: Stanford University. <http://plato.stanford.edu/archives/fall2015/entries/epiphenomenalism/>.

Spencer, Herbert. 1873. *The Principles of Psychology*. 2nd ed. 2 vols, *System of Synthetic Philosophy*. New York: Appleton and Co.

Van Gulick, Robert. 1995. "What Would Count as Explaining Consciousness?" In *Conscious Experience*, edited by Thomas Metzinger, 61–79. Paderborn: Ferdinand Schöningh.

Wright, John. 2015. "On James's Argument against Epiphenomenalism." *William James Studies* 11: 69–85.

Notes:

(1.) The idea that consciousness needs to be "explained" has been central in recent philosophy of mind (e.g., see Chalmers 1997, 9). Even those who are suspicious of the "hard problem" see consciousness as a proper *explanandum*, and the debate nowadays tends to focus on what kinds of things can serve as *explanantia* (e.g., see Dennett 2001, Carruthers 2001). For analyses of various senses in which people have sought to "explain" consciousness, see Van Gulick (1995), Carruthers (1998).

(2.) James says the science of mind should not try to explain "how or why" conscious states and brain states "hang indubitably together," but should only attempt to document an "empirical parallelism" between the two, and to do so "provisional[ly]." This way "our psychology will remain positivistic and non-metaphysical..." (PP 1890, 182).

(3.) Parts of the essay were later used in chapters five and nine of James's opus, *The Principles of Psychology* (1890).

(4.) The essay can be found at (MEN 1872, 247–256). For a recollection of the aims of this essay, see James's letter to C. A. Strong (October 21, 1889), quoted in Perry (1935, II.26).

(5.) See his 1874 review of William Carpenter's *Principles of Mental Physiology* (esp. ECR 274) and his review of Wundt's *Grundzüge der physiologischen Psychologie* the following year (esp. ECR 303). I thank Trevor Pearce for calling my attention to these reviews.

(6.) What remains of this manuscript can be found at ML (16–43). Perry suggests James's lectures at Johns Hopkins in February of 1878 were similar to his Lowell series, citing a letter from Francis J. Child (who had hosted James during his stay in Baltimore) to the effect that the final Hopkins lecture had “offered reasons for not accepting the theory that we are automatons unreservedly” (RBP 1935, II.27). The documents that survive of that lecture series are also incomplete (see ML 3–15).

(7.) See letter to Augustus Lowell, January 23, 1879, at CWJ (5.37).

(8.) There is surprisingly little literature on James's early defense of interactionism. For instance, this is one of the few major issues in James's writing barely touched on in the otherwise exhaustive Myers 1986 (see 54–58 for his brief discussion). Perry's chapter on the topic (Perry 1935, II.ch. 53) offers some revealing documentary evidence but is short on critical analysis. There is an idiosyncratic essay in the 1950s that mistakenly (see the 1879 letter quoted below, from CWJ 5.34) runs automatism together with determinism (Capek 1954, esp. 274–276). Two more recent essays (Flanagan 1997, Reck 1972) are both concerned to reconcile James's early interactionist dualism with what he would later call his “philosophy of pure experience” (or what Russell termed “neutral monism”). As such, neither author makes much effort to unpack James's actual evidence for interactionism—Flanagan simply says James rejected epiphenomenalism on “‘common-sensical’ grounds” (Flanagan 1997, 31), and Reck makes a similar claim (Reck 1972, 29). James did regard interactionism as in accord with “common sense” (e.g., at PP 1890, 139), but as we shall see he thought his view also accorded with extensive empirical evidence. Finally, there is a more recent literature on epiphenomenalism, some of which takes up one of James's key *objections* to this position (see fn. 36, later), but none of which considers James's own positive view.

(9.) Indeed, James's usage is the first cited by the Oxford English Dictionary. I note that at PP 1890 (133) James actually puts “epiphenomenon” in quotation marks, seeming to attribute the term to (Hodgson 1865), but I cannot find Hodgson using the term himself. In contrast, at (VRE 1902, 390), James attributed the term to Clifford. He presumably had in mind Clifford (1874/1886), but the word does not appear there, or anywhere else in Clifford's writings I can find. The OED cites a primary, older usage of the word that comes from medicine, and dates to the early eighteenth century. In medicine an “epiphenomenon” is a secondary symptom of an underlying condition, such as a fever caused by an infection. James had completed his physiology training in Harvard's medical school

and was presumably adapting this medical usage in his discussion of consciousness. Finally, James actually did use “epiphenomenalism” in an *unpublished* notebook, dated 1903–1904 (MEN 52).

(10.) Hodgson articulated this view in 1865 (see Hodgson 1865, I.278 ff.), and James quoted from this work (at PP 1890, 133). But Hodgson only actually espoused the view himself in Hodgson 1870 (I.416 ff.).

(11.) I discuss these experiments at length in Klein 2018.

(12.) The frog brain was often depicted as a set of sub-structures connected serially, through the brain stem, to the spinal cord. For a diagram from the era, consult figure 1, later.

(13.) Pflüger says the frogs used in the experiment I describe in the text were pithed below the medulla oblongata (Pflüger 1853, 18).

(14.) James writes that reflex actions are “the result of outward discharges from the nervous centres” when “these outward discharges are themselves the result of impressions from the external world, carried in along one or another of our sensory nerves” (WTB 1897, 91). A standard example would be wincing at an object moving toward the eye.

(15.) For an overview of Lewes’s life and intellectual work, see Price 2014.

(16.) By calling his view a hypothetical statement of “direct personal feeling,” he did not mean that the *evidence* for his hypothesis was personal feeling, but rather that he was offering a scientifically credible hypothesis that *agrees* with personal feeling (a point he made with some frequency; see earlier, fn. 8). This is clear from the sentence following the passage I quote in the text: “But the intricate analysis by which it [viz., James’s own hypothesis] has been reached gives it great plausibility” (EPs 1879, 52). The “intricate analysis” is not a simple appeal to common sense, but the physiological and evolutionary considerations I am analyzing here in the present section.

(17.) James did not give a citation, but he presumably had in mind Goltz 1869 (70).

(18.) James was especially indebted to Goltz for the point that, despite their remarkable capacity for coordinated action, de-cerebrated frogs rarely act at all unless prompted by some overt stimulus. See Danziger 1980 (99–100) for a brief discussion, and for more on the controversy surrounding the pithing results generally, see Klein 2018 (897–898).

(19.) Let me emphasize that Jamesian prudence is not to be taken as an exclusively subjective concept. Both components of prudence involve objective, behavioral phenomena. Thus, James cashes out *acting on considerations of what is absent* as something observable in the intact frog, as when we find it “impossible to predict” (for this phrase, see immediately following) a response to a stimulus. And the same goes for *vis a fronte* choosing—one can simply observe whether this capacity is present, e.g., in a frog emerging for air that tries various means to reach its obstructed goal.

(20.) His subsequent work on will complicates this story somewhat, but I cannot explore this issue here. For more on how James's accounts of consciousness and will fit together, see my (Klein Forthcoming).

(21.) I read James's use of "consciousness" as shorthand for "conscious mental states." James held that psychology should resist the urge to postulate a *thinker* somehow standing behind our passing thoughts—the thinker *just is* the passing thought, for James (PP 1890, 328). So like in much contemporary philosophy of mind, when James uses the noun form "consciousness," this should be taken as shorthand for a property that only mental states (not whole creatures) can instantiate.

(22.) This phrase is usually used in connection with actions a court of law takes on its own accord, as opposed to actions it might take on motions brought by interested parties, as when a court by itself moves to dismiss a case (say, because of a lack of jurisdiction), rather than because an interested party has moved to dismiss.

(23.) On this terminology, see subsequent discussion, fn. 30.

(24.) A trait that enables x-ing need not always be involved in x-ing in order to be an adaptation for x-ing. For instance, spatial pattern separation (the ability to remember distinct but similar spatial patterns) may be an adaptation for foraging (Mattson 2014, 5), even though we sometimes use this skill when not foraging (e.g., when playing various kinds of games). Similarly, consciousness could be an adaptation for evaluating, even though consciousness sometimes occurs without evaluation (perhaps in states of meditation, say). What is more, foraging is neither necessary nor sufficient for spatial pattern separation, obviously, even though the latter may still be an adaptation for the former; similarly, evaluation may be neither necessary nor sufficient for consciousness, even though the latter may still be an adaptation for the former.

(25.) In another memorable passage making a similar point, James wrote that an "electrical machine [n]ever get[s] restless because it can only emit sparks, and not hem pillow-cases like a sewing-machine" (PP 1890, 22). I take it the non-physicality of evaluation is James's central reason for resisting any account that portrays consciousness as *nothing but* prudent behavior, by the way, which is a purely physical activity. For James, consciousness is non-physical, so it may *cause*, but cannot therefore be identical with, behavioral prudence.

(26.) James sometimes writes about selecting objects "suited" to one's "private interests" (EPs 1879, 50).

(27.) It is not so much that all value is relative, for James held that *every* interest created by consciousness produces a prima facie demand to be satisfied (WTB 1897, 148). Deliberation about value, for James, is not about figuring out what the *right* set of interests is. In some sense, *every* interest is "right," and none absolutely overrides any other. He was particularly keen to deny a kind of flat-footed evolutionary ethics that claimed survival to be an absolutely overriding interest, a view he associated with Spencer (Klein 2016, sec.

3). Instead, James thought ethical deliberation is about sorting out conflicts that will inevitably arise between the *mutually incompatible* interests different conscious creatures will create—as he put it, ethics is about choosing *which* “interest out of several equally coercive shall become supreme” (EPs 1879, 51). This idea prefigures “The Moral Philosopher and the Moral Life” (1891), where James offers a naturalistic account of ethical evaluation in terms of the reconciliation of competing values. For discussions of James’s treatment of value, see chapters by Marchetti, Bush, Talisse, and Kloppenberg and Throntveit in the present volume.

(28.) James wrote that the “materialist,” or presumably anybody who denies the efficacy of consciousness, “is immediately confronted by the notorious fact that the strongest tendencies to automatic activity in the nerves often run most counter to the selective pressure of consciousness” (EPs 1879, 59).

(29.) Hence, “[c]onsciousness produces nothing, it only alters the proportions” (EPs 1879, 52).

(30.) I will 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).

(31.) The longer passage includes reference to “Mr. Darwin,” who “regards animated nature as a sort of table on which dice are continually being thrown. No intention presides over the throwing, but lucky numbers from time to time fortuitously turn up and are preserved.” For more on the dice metaphor, see the passage from EPs 1879 (43) reproduced immediately below.

(32.) Consult Pearce, this volume, on James and the evolutionary biology of his day.

(33.) A similar suggestion can be found at ML 1878 (26).

(34.) James would have been familiar with this example, which I adapt from Spencer 1873 (I.295), and which I discuss in another connection in Klein 2016 (4).

(35.) James considers the related example of fish like “cunners and sculpins” that are easy for humans to catch because they “lack ... all thought by which to weigh the danger against the attractiveness of the bait” (PP 1890, 34).

(36.) Some recent literature that takes up James’s objection to epiphenomenalism includes Robinson 2007, Corabi 2008, 2014, Robinson 2014. Another recent treatment is Wright 2015, which in some respects defends James’s perspective.

(37.) I develop this reading in detail in Klein (2019).

(38.) The passage also appears at EPs 1879 (55–56) with minimally different wording.

(39.) By “native” I mean *inheritable*—I use the former term for linguistic convenience. The “distribution” of pleasures and pains at issue for James must be inheritable, since only inheritable traits can evolve.

(40.) I would like to thank Gary Hatfield for commenting on this chapter, along with Trevor Pearce and Elisabeth Lloyd, who both commented on early versions. I gave portions of this chapter as part of a lecture series at École Normale Supérieure in Paris. I would like to thank Mathias Girel, CAPHÉS, Labex, and the Centre de Cavailles for hosting me, and also the US-UK Fulbright Foundation for financial support during the 2016–2017 year when I was working out some of these ideas. That year I was in residence at the University of Sheffield, where conversations with Luca Barlassina and Jeremy Dunham particularly shaped the direction this research would eventually take.

Alexander Klein

Alexander Klein, Associate Professor of Philosophy, McMaster University