

# James's Evolutionary Argument

**William S. Robinson**  
Iowa State University

DOI: 10.2478/disp-2014-0014

BIBLID [0873-626X (2014) 39; pp. 229-237]

## **Abstract**

This paper is a commentary on Joseph Corabi's "The Misuse and Failure of the Evolutionary Argument", this Journal, vol. VI, No. 39; pp. 199-227. It defends William James's formulation of the evolutionary argument against charges such as mishandling of evidence. Although there are ways of attacking James's argument, it remains formidable, and Corabi's suggested revision is not an improvement on James's statement of it.

## **Keywords**

Epiphenomenalism, IBE, physicalism, pleasure, William James

In "The Misuse and Failure of the Evolutionary Argument", Professor Corabi offers a critique of James's argument against automatism (what today we call "epiphenomenalism"), and develops what he claims to be an improved version of the argument, even though, in the end, he believes it is a failure. In this response, I will try to bring out the strength of James's argument, and explain why I think Corabi's version is not an improvement.

I had better say at the outset that I am not persuaded that rejection of epiphenomenalism is forced upon us by James's evolutionary argument (hereafter, the EA). I have explained why in Robinson 2007 and shall not repeat that material here.<sup>1</sup> However, I do think that the EA is a formidable argument, and that James expressed it quite well. In the first part of these remarks, I will recast James's argument in a

<sup>1</sup> Professor Corabi has replied to my 2007 article in Corabi 2008. I have not published a response to this paper, and I will not comment on it here, except to say that I stand by my 2007 paper.

style that will ease our discussion. (Readers may compare my version with James's paragraph quoted at the beginning of Corabi's paper.) Later sections will comment on Corabi's view of the EA.

## 1 James's argument

In what follows I will use 'approach' (to a stimulus) to include not only reduction of spatial distance, but also prolongation or repetition of proximity to a stimulus. Similarly, tendencies to 'avoid' will include tendencies to increase spatial distance, and tendencies to shorten or avoid repetition of proximity to a stimulus.

'Smooth correlation' (between utility and hedonic valence) will refer to the following facts. (a) For the most part, approaching stimuli that are beneficial to us (that enhance our fitness) is pleasant. For example, eating when hungry, and drinking water when thirsty, are pleasant. (b) For the most part, approaching stimuli that are detrimental to us is unpleasant. James cites consumption of alcohol to the point of drunkenness as an exception, but holds that such exceptions "relate to experiences that are either not vital or not universal" (James 1890, v. 1: 144). Neither Corabi nor I make any point that hangs on the exceptional cases.

There is, of course, also a high correlation between utility and behavior. For the most part, we do not approach the detrimental, and we do not avoid the beneficial. The heart of James's argument, however, turns on pleasure and pain, and I will use 'smooth correlation' for the match between the (positive or negative) utility of stimuli and the pleasure or pain that we get from their presence. I will reserve 'high correlation' for the utility/behavior correlation.<sup>2</sup>

With these understandings in place, we can set out James's argument as follows.

J1 There is a smooth correlation between utility of stimuli and their hedonic valence.

J2 If the pleasantness and painfulness of stimuli had no effects

<sup>2</sup> In the chapter that contains the EA, and in the first sentence of the key paragraph, James writes of "consciousness". However, the argument itself is entirely about pleasures and pains. My formulation of the argument reflects this fact.

in behavior, then there would be no explanation for the fact in J1.

J3 If the pleasantness and painfulness of stimuli do have effects in behavior, we can explain the fact in J1.

The envisaged explanation invokes natural selection. Organisms lacking the smooth correlation would soon die out, if pleasure and pain are efficacious, because they would approach the detrimental and/or avoid the beneficial. Organisms that did have the smooth correlation would be more likely to live long enough to pass their traits — including whatever grounds the smooth correlation — on to the next generation. On the other hand, if pleasure and pain are not efficacious, “one does not see” why there should not be serious mismatches between utility and hedonic value, and it would require an unscientific embrace of “*a priori* rational harmony” to explain the fact in J1.

James takes the implication of these premises to be obvious, but let us state the conclusion explicitly.

J4 The hypothesis that pleasures and pains are efficacious is a better hypothesis than the hypothesis that they are inefficacious.

James does not mention physicalism or interactionism, but both of these views accept efficacy of pleasures and pains. So, in the contemporary climate of opinion, it is reasonable to extend James's own conclusion in the following way.

(J5) Physicalism is a better hypothesis than epiphenomenalism.

(J6) Interactionism is a better hypothesis than epiphenomenalism.

And it is a consequence of these that

(J7) [Physicalism v Interactionism] is a better hypothesis than epiphenomenalism.

However, the key point that underlies (J5) through (J7) is the acceptance of efficacy for pleasures and pains; and this is clearly stated in J4.

It is well known that physicalism and interactionism have difficulties of their own. So, it may well be that (J5) and (J6) are false — false because, for reasons not mentioned in James's argument, they

turn out to be either worse hypotheses than epiphenomenalism, or equally problematic hypotheses, all things considered.

Plausibly, it is worries of this kind that lead Corabi to a desire to give an improved version of the EA. His discussion of interactionism can, I believe, be summed up as the point that natural selection will not explain why the laws from pleasure to approach, and pain to avoidance, are what they are. If laws are contingent, then there are alternative possible scenarios that will include the high correlation of utility and behavior, but lack the smooth correlation of utility with hedonic valence — namely, scenarios in which both the valence and the laws connecting valence to behavior are inverted. This point seems to me to be correct.

Corabi's discussion of physicalism is harder to summarize. I will return to it briefly after we have seen more of Corabi's framework. The point to notice for now is just that since (J5) and (J6) have problems of their own, it cannot be clear that James's argument to the best (or, at least, the better) explanation is altogether decisive. With regard to this point, Corabi and I are in agreement.

Our fundamental disagreement concerns the following claims made in the introductory section of Corabi's paper.

- C1 James's argument "mishandles" the evidence.
- C2 Corabi's "canonical" version of the argument is an improved version of the EA.
- C3 The central mistake of both versions is accepting an assumption that is unjustified and almost certainly false. This assumption is that "all precise versions of physicalism will posit just this [i.e., the actual] connection between the physical and the phenomenal".

## 2 Difficulties about "the evidence"

In my view, trouble begins early in the section on "Some preliminary matters and the canonical formulation", when Corabi asks "What is this evidence [on which the evolutionary argument is based]?" Corabi's answer is that it consists of correlations between distal stimuli

and qualia, and between exposure to stimuli and behavior. Corabi then adopts the principle that we should always use the most determinate evidence available. Thus, we should avoid formulations in terms of sharp cuts, avoidance behavior and unpleasant qualia, and instead “use evidence like ‘sharp cuts to the arm of determinate type  $t$  result in avoidance behavior of determinate type  $b$  and are mediated by qualia of determinate type  $q$ .’”

This account of the “evidence” used in the EA seems to me to be profoundly mistaken. The evidence to which the EA appeals is the smooth correlation — the fact stated in J1. The gist of the argument is simply that efficacy for pleasures and pains allows for a better explanation of J1 than would any theory that denies such efficacy.

Let us introduce ‘determinate stimuli-behavior-qualia statements’ (hereafter, ‘determinate SBQ statements’) to refer to statements of the form illustrated by Corabi’s example quoted a few lines above. What might determinate SBQ statements be evidence *for*?

It might be suggested that they are evidence for J1. This cannot be right. The smooth correlation asserted by J1 holds between utilities and hedonic valences. But determinate SBQ statements do not classify stimuli as beneficial or detrimental, and they do not classify qualia as pleasant or unpleasant.<sup>3</sup> Without such classifications, they are *irrelevant* to J1. Moreover, natural selection plausibly explains the smooth correlation, as stated in J1, much better than it explains fully determinate instances. Once a detrimental stimulus has been connected with displeasure sufficient to cause avoidance, there would not likely be further selection pressure to home in on an exact value. One might well expect that, in general, severely detrimental stimuli should be more unpleasant than milder threats, but again, most-determinate values would not be likely predictable from natural selection considerations.

Perhaps, then, determinate SBQ statements are evidence for interactionism or physicalism. But this is not right either. Physicalism says that experiences are physical. By itself, that does not imply any

<sup>3</sup> I used ‘painful’ rather than ‘unpleasant’ in giving James’s version, because that is the term James uses. But Corabi uses ‘unpleasant’, and the argument’s structure seems better served by this term. After all, many things that would be detrimental for us to ingest have unpleasant tastes and smells; but these are not properly called ‘pains’. Hereafter, I will use ‘unpleasant’.

SBQ statement, determinate or not. It just says that *whatever* SBQ relations we may find, the experiential component will be physical (and thus able to have effects). That is compatible with, e.g., arsenic being delicious. More generally: The falsity of a determinate SBQ statement is not evidence against physicalism. Parallel considerations hold for interactionism.

Perhaps determinate SBQ statements are evidence for the conjunction of interactionism or physicalism with natural selection. But, for reasons recently explained, natural selection plus either view will not predict determinate SBQ statements. What will be predicted, on either physicalism or interactionism plus natural selection, is that detrimental stimuli will generally be correlated with unpleasant experiences, and beneficial stimuli will generally be found pleasant. That is, they will predict J1.

J1 says nothing directly about behavior: it asserts the smooth correlation between utility and hedonic valence. So, perhaps we ought consider just the SQ part of SBQ statements. It will be evident on inspection that the preceding remarks apply just as well to determinate SQ statements.

Later in the same section, Corabi writes: “A good (albeit idealized) way to think of the confirmation process is to envision each general hypothesis (e.g., epiphenomenalism) as a disjunction of highly determinate versions of that hypothesis, each of which specifies the history of the world in maximal detail.” But this does not seem to me to be a good way to think at all. The general hypotheses (physicalism, interactionism, epiphenomenalism) make no predictions about which determinate stimuli will be found correlated with which determinate qualia. Two of them predict that utility will smoothly correlate with hedonic valence, but none of them predict correlations between specific S and specific Q. Finding that there is a specific SQ correlation will thus not be helpful in deciding among the views.

Later in the paper, Corabi explicitly recognizes this point (in the section titled “The problem with the key assumption”). When we observe (or fail to observe) a particular determinate SQ correlation, none of the three hypotheses (physicalism, interactionism, epiphenomenalism) will gain or lose any ground. But if this is so — if determinate SQ correlations cannot be used to discriminate among rival views — then their claim to be the relevant form of “evidence”

is undercut.

Corabi's conclusion differs: it is that an improved version of the EA still fails. In my view, however, his revision of the EA is not an improvement. Instead, the insistence on maximum determinateness merely introduces irrelevant detail that obscures the heart of the EA. This heart is the smooth correlation stated in J1. Fragmenting this correlation into a plethora of determinate possibilities does nothing to strengthen the EA, and focusing on determinate characteristics rather than on the utility of stimuli and the hedonic valence of the experiences they cause removes the relevance of the "evidence".

### 3 Results for C1 – C3

The previous paragraph explains my rejection of C2 (the claim that the proposed "canonical" version is an improvement). Regarding C1, the case for "mishandling" of evidence rests on the fact that James is content with a general statement of smooth correlation, and does not require determinate SBQ statements. We have seen, however, that these statements are a distraction rather than an improvement, so we have no reason to accept that James mishandled evidence. On the contrary: J1 accurately states something that needs to be explained, and the sense of the argument is that efficacy for experiences better explains it than does epiphenomenalism. This remains a formidable argument, even for those who find grave difficulties in physicalism and interactionism, and even for those who may think that, when all the difficulties for the latter views are taken into account, the claim that the EA is a successful inference to the best explanation is undermined.

Regarding C3, it should by now be clear that James's version of the EA does not rely on any assumption about physicalism. So, C3 is false, as applied to it.<sup>4</sup>

<sup>4</sup> Neither does the EA say anything about physiological transitions in the brain (contrary to what Corabi says in the third paragraph of the section on "The central traditional confusion and the key assumption"). Any physiological hypothesis that is compatible with efficacy for pleasure and unpleasure will do.

#### 4 Physicalism

Corabi's discussion of physicalism, in the section on "The problem with the key assumption", is couched in terms of "alternate determinate physicalist hypotheses". But at the end of section 2 above, we saw that these determinate hypotheses are not the right candidates for "evidence" regarding the EA. So, the problem that Corabi finds for the use of such hypotheses cannot show that there is a weakness in the EA.

However, I agree with Corabi that, even on the assumption of physicalism, statements of identity between properties that are specified in neural terms and properties that are not specified in neural terms are epistemically contingent (despite being metaphysically necessary, if they are true). And I agree that this fact presents a problem for those who wish to use the EA as an argument for physicalism. In what follows, I will briefly explain how epistemic contingency relates to the formulation of the EA given in the first section of these comments.

The focus of that formulation is pleasure and unpleasure. The claims of interest in discussing physicalism are thus claims of identity between physical (presumably neural) event properties, and pleasure or unpleasure. If we let 'NE1' and 'NE2' stand for the neural event properties that will (according to physicalism) be discovered to be identical with pleasure and unpleasure, respectively, we can express the crucial claims as follows.

P+ Pleasantness is identical with NE1.

P- Unpleasantness is identical with NE2.

Now, since these statements are not known to us *a priori*, it is intelligible to us that they might have been false (even though they are metaphysically necessary, if they are true). A plausible consequence of that intelligibility is that we cannot *explain* why the smooth correlation holds. For, as far as we can see, it might have turned out that NE1 was identical with pain, even while the laws of neurophysiology remained what they actually are. In that scenario, we would approach the painful, and doing so would enhance our fitness. Or, it might have turned out that alternative arrangements of brain parts other

than those involved in NE1 and NE2 would have established a condition in which the behavioral effects of NE1 and NE2 were reversed. In that case, if P+ and P- are true, our fitness would have been served by approach to the unpleasant and avoidance of the pleasant.

The Jamesian complaint is that epiphenomenalism cannot explain the smooth correlation. According to the line of thinking just described, physicalism cannot explain it either. So, if that line of thinking is correct, then physicalism is no better off in the face of the EA than is epiphenomenalism. Expressed in terms of the argument in the first section above, there is some reason to doubt J5.

The upshot of these reflections is this. We should agree with Corabi that the EA is not as powerful an argument for physicalism (or for [physicalism v interactionism]) as it is sometimes thought to be. But James's formulation of the argument remains formidable, because it does require a response from epiphenomenalists. They must either offer some alternative explanation of the smooth correlation, or cast doubt on the ability of rival views to explain it. James's formulation makes this burden clear. Reframing the matter in terms of determinate SBQ (or SQ) statements introduces irrelevancies that obscure such strength as the EA has.<sup>5</sup>

William S. Robinson  
Iowa State University  
402 Catt Hall  
Ames, Iowa 50011-1306 USA  
wsrob@iastate.edu

### References

- Corabi, Joseph. 2008. Pleasure's Role in Evolution: A Response to Robinson. *Journal of Consciousness Studies* 15:78-86.
- James, William. 1890. *The Principles of Psychology*. Cambridge, MA: Harvard University Press.
- Lipton, Peter. 2004. *Inference to the Best Explanation*. Second edition. London and New York: Routledge.
- Robinson, William S. 2007. Evolution and Epiphenomenalism. *Journal of Consciousness Studies* 14:27-42.

<sup>5</sup> There is no rejection here of Bayesianism, properly applied. For explanation of compatibility of Bayesianism and IBE, see Lipton 2004.