Introduction

Sebastian Montesinos

The most important lesson I learned in philosophy as an undergraduate was how to read papers. My intellectual mentors taught me that rather than jumping to criticize a piece or dispute its conclusions, one should read it very carefully, try to articulate its positions as charitably as possible, and then consider it only on the basis of that understanding. This is the standard practice in philosophy for good reason—it cultivates the essential philosophical virtue of charity. The second most important lesson I learned in philosophy was epistemic humility. A vast philosophical literature exists that no one person can entirely incorporate into their conceptual schema, and metaphysics is still in many ways shrouded in mystery, which is why it is always beneficial to be tempered and reasonable in one's conclusions.

The problem with Adelstein's response to our criticism of the psychophysical harmony argument is not our disagreement, but that the approach he took epitomizes precisely the opposite approach to philosophy than that outlined above. Any careful reading of our piece juxtaposed with Adelstein's demonstrates that he did not make even a first attempt to charitably interpret the arguments we made or carefully read what he wrote. He frequently misconstrues our arguments, skips over entire sections of our post, and makes points that were already addressed without mentioning our responses. This would be less egregious if Adelstein was more reserved in his conclusions. Instead, he blithely dismisses views out of line with his own on the basis that they are 'obviously,' 'clearly,' or 'manifestly' false, the views of a 'radical,' 'extreme minority,' and 'absurd' or 'utterly crazy'. He concludes that our arguments are 'lousy' and 'bad'. Therefore, he not only fails to display any charitability in his writing, he also lacks any epistemic humility. Positions that have sparked entire subfields within the philosophical literature and are advocated by highly intelligent, informed people are dismissed without even a first-pass investigation.

Adelstein's response to our post almost entirely falls into two camps: First, to claim that the views he is responding to are obviously false as revealed by intuition, unpopularity, or their radical implications (or some combination of all three). Second, misunderstanding, misrepresenting, or omitting key parts of the arguments that we make. Our aim in this piece is twofold: first, to present a broad criticism of that first, intuition-mongering strategy Adelstein utilizes ('The Use of Intuition in Philosophy'). We begin in a more surface-level way by discussing the dialectical relevance of intuitions (sections 1 & 2), and then we delve more deeply and explore what exactly this kind of strategy amounts to metaphilosophically, and why it is, in our view, so myopic (sections 3 & 4). Second, to point out the various ways in which Adelstein has failed to charitably read our piece, and how his objections are less compelling than he thinks (sections 1-6 of the 'Misunderstandings' heading).

Table of Contents

The Use of Intuition in Philosophy

1. *"It's Obvious bro."*

Benjamin (Truth Teller) and Sebastian Montesinos

A major methodological concern that permeates a great deal of Adelstein's post is what we will call 'intuition-mongering'. There are several points in Adelstein's article where he asserts "It is plausible that.." or "It is obvious that.." and goes on to provide no argument whatsoever for the claims in contention. This raises the following question: is Adelstein merely reporting his own psychological states? If that's the case, there's no contention here. Yet, he doesn't specify, "It is obvious to me that..."; instead, he broadly proclaims, "It is obvious that..." full stop. If Adelstein is merely reporting his own psychological states, it would do well for him to temper his assertions (Relatedly, here's everyone's favorite, Joe Schmid, on this). Further, if Adelstein is merely reporting his own personal sentiments, that isn't of much interest to us, or to a broader audience. That Adelstein is disposed to believe some proposition P is obviously true, does not provide us, or anyone else, with much of a reason to accept the claim that P is obviously true. If, however, Adelstein's intent is to suggest that most people find P obvious, he is venturing into empirical territory. As such, it becomes incumbent upon him to offer empirical evidence substantiating that most people take P to be obvious or plausible. Yet, this evidentiary support is consistently absent from his post. Additionally, even if it were true that most people take P to be obviously true, while this would be some evidence for the truth of P, it would not be decisive nor particularly strong evidence.

There are a couple points in Adelstein's article where one could interpret him as making an appeal to the intuitions of the majority of philosophers. For instance, he states that only 16% of philosophers accept or lean towards the inconceivability of zombies. One problem here is that these surveys do not get at the reasons *why* philosophers hold the views that they do. That a philosopher thinks zombies are conceivable does not show that the reason they think this is because it is 'intuitive' or 'obvious', nor that they feel particularly strongly about this seeming. As Collier will soon explain in section 3 & 4, the paradigms through which philosophers take themselves to be answering questions vary greatly, and are dependent on a large variety of theoretical methodologies that cannot be reduced to a clash of intuitions. Furthermore, there is significant evidence against the claim that, in this case at least, any of this is 'obvious' to philosophers. Only 60% of philosophers actively endorse the conceivability of zombies, and of that 60%, over half only lean towards, rather than flat-out accept, their stated view (Bourget & Chalmers, 2020). Overall, then, philosophers appear to feel rather ambivalently about this issue. Lastly, when you switch the target group to philosophers of mind—the demographic whose expertise is most relevant to the conceivability of zombies—the view that zombies are inconceivable goes up to 22%, and is slightly higher than the view (Adelstein's view we might add) that zombies are metaphysically possible! This suggests that more familiarity with philosophy of mind corresponds to an increased tendency to accept the inconceivability of zombies.

So, all in all, the assertion that there is a consensus or even majority view of 'obviousness' or 'strong intuition' in favor of the conceivability of zombies amongst philosophers is unsupported, and that is before even examining what exactly the implications of that *should be* in an argumentative context. A far less dubious argument would be just an appeal to the fact that a slim majority of philosophers accept the conceivability of zombies, which is some inductive evidence based on an appeal to expertise. What this is worth is unclear, at best it is minor evidence for the position, but hardly enough to rest a case on. Notably, by far the most accepted position here is the type B physicalist view that zombies are conceivable but not metaphysically possible, a position Adelstein totally rejects. Adelstein should be aware, then, that these kinds of appeals constitute at best *prima facie*, weak evidence for positions. However, that there is some weak, defeasible evidence for the conceivability of zombies is of course a far cry from Adelstein's much bolder claim that zombies are "obviously conceivable".

2. The Epistemic and Dialectical role of Bare Intuitions

Benjamin (Truth Teller)

What I suspect may be going on here is that Adelstein takes 'it is obvious that P' to be an intuition in the sense that this is a sui generis propositional attitude which supposedly plays the role of conferring justification without standing in need of justification. The upshot here is that having an intuition that P is obviously true provides him with prima facie justification that P is obviously true. However, intuitions and their role as justifiers (assuming they have such a role) are private and agent-relative—what appears obvious to one need not be so for another. Thus, even granting that Adelstein's intuition that 'P is obviously true' provides him with prima facie justification for P's obvious truth, this justification is not public, it will not transfer to those who lack the intuition. Recall also that this is a dialectical context in which the burden lies with Adelstein to positively defend the psychophysical harmony argument from criticisms, thus merely appealing to intuitions in a context where the intuitions clearly aren't shared won't do.

This is all to assume that intuitions, on their own, provide any non-theory laden justification for a belief with no independently motivated conceptual framework or inferential support to wear the trousers, something that I wouldn't grant. It's essential to clarify the terms. I am taking justification to be a kind of relation that a proposition or set of propositions P* bear, such that P* either entails or raises the probability (to a sufficient degree) of another proposition or belief's truth. It's also useful to distinguish here between "being justified in a belief B", and "justifying B". I want to leave open the possibility that epistemic externalism is true, wherein one might be justified in belief B despite lacking any conscious, reflective access to the reason(s) which justifies B, e.g a reliable cognitive process generates the belief B, or B being causally connected in the right-sort-of-way to the truth conditions of the proposition B is about. Are intuitions justifiers in this sense? Perhaps so, though we have reasons to be pessimistic in the case of intuitions about modal facts so disconnected from our ordinary practices of the kind Adelstein draws upon, such as the conceivability of zombies, or the possibility of inverted qualia. There does not appear to be any causal belief-generating mechanism appropriately linking our minds and our resources to the truth conditions of these sorts of propositions, and neither do we have any reason to think our brains would have evolved to have accurate beliefs regarding such things. Be that as it is, in a dialectical context, what interests us are justifications in the latter "justifying B" sense, that is, having and being able to provide accessible and articulable reasons to think a proposition is likely true. What this means is that it's not enough that intuitions are in fact reliable at generating true beliefs, it needs to be argued that they are so reliable. The proposition "Adelstein has an intuition that P is obviously true" does not entail that P is in fact true, probably true, or even more likely true than before. Indeed it is perfectly consistent with P being very probably false. An inference with a bridging premise is needed in order to cross this gap, which would of course require Adelstein to do the hard work of actually providing arguments for his claims and broader methodology, hard work he simply didn't do in more cases than not.

Now, Adelstein may have sympathies for a kind of epistemic conservatism—one which allows his intuitions to provide him with reasons to think he is 'entitled' from his perspective to believe P is obviously true in the absence of defeaters—and that he isn't making an epistemic error in doing so. Maybe so, but the point here is only that, Adelstein merely reporting his intuition that "such-and-what is obvious, and/or a manifest fact," does not actually provide any justification, in the dialectically effective sense, whatsoever for the claim, not even for him. It is little more than rhetorical bluster, mere noise with no meat on the bones, and unless and until Adelstein does the aforementioned hard work, we are well within our epistemic rights to not take such intuition-mongering seriously.

3. Philosophy as a series of Moorean Shifts Between Two Guys That Can't Be Wrong

Lucas Collier

If Paul Churchland is right (bear with me), there were probably some linguistic communities from Ptolemy to Newton in which "movement" meant a "change in position relative to the Earth" (Churchland, 2007). Heliocentrism, which posited that the Earth itself moves, was thus nonsense. How can it be that the Earth moves relative to *itself*? The view did not merely ask of people that they endorse some propositions that they before rejected, it asked of them that they reconfigure their language and their conceptualization of the data at hand: the paradigmatic motion of the sun as it falls beneath the sea may be no motion of the sun at all. When the geocentrist begins to take their interlocutor seriously, meanings are changed, explananda reimagined, and what was once incoherent becomes real.

The histories of science, mathematics, and philosophy are wrought with similar instances of paradigm shifts and unimaginable truths. Putnam 1996 discusses the case of Euclidean geometry, which was shown to be a false description of physical space. He points out that prior to mathematicians like Riemann or Lobachevski, it's not clear that people *could* know a way for Euclidean geometry to be disconfirmed. They possessed a reservoir of concepts that was insufficient for imagining how they could ever be wrong. Putnam defers to Wittgenstein and compares such situations to difficult riddles, ones where you're not even sure what the riddle means until you know its answer. This calls for humility: we ought to temper our credence in our propositions together with paradigms, because what couldn't be right sometimes is in ways that we couldn't imagine.

One of my goals with these examples is to remind us of a very trivial truth. Despite the rationality of our arguments and the clarity of our intuitions, we are never free from our epistemic contexts: our times, our places, our neurons, our languages, our concepts, our bookshelves, our histories, our mothers, and so on. What is clear to us in one epistemic context might be opaque after our perspective has changed. One aim of good arguments is to bridge these epistemic contexts and allow us not only to solve a problem as currently understood, but reshape the problem itself. Maybe after reading an argument, one will look back and not just countenance a proposition they before denied, but feel as though the riddle takes on a new meaning.

This is why I am skeptical of the "intuition-mongering" outlined in the previous two sections which takes everything for granted and dismisses out of hand arguments and riddles alike. Its users loudly cut their way through philosophical cloth to leave perfectly intuition-shaped holes, becoming ever more sophisticated without ever seriously wondering whether they are metaphilosophically mistaken. This philosophical malady appears on a few occasions in Adelstein's response to our post, and I'd like to use these instances as examples to explain why the approach is so problematic.

Adelstein right out of the gate employs such a tactic in response to our "Concerns" section. The function of that section was to outline what the argument from psychophysical harmony takes for granted (a list that I now see as *much* longer since helping to write it). The idea was that there are plenty of live views (e.g. eliminativisms, various moral anti-realisms, directed naturalisms, content externalisms) or orientations (e.g. liberal naturalism) in philosophy that don't jive well with the argument, and that this should be taken into consideration when evaluating its dialectical efficacy. Adelstein doesn't really seem to appreciate this point. His first comment is this:

> [eliminativism, type A physicalisms, and liberal naturalism] require denying the manifest fact that zombies are conceivable, that inverted qualia are conceivable, and that Mary learns something when she sees red. On account of denying this, they are extremely implausible.

There are two things about this that I'd like to discuss. The first is the methodology implicit in this knee-jerk response. It's kind of innocuous at first glance: Adelstein is being epistemically conservative, exercising the maxim of minimum mutilation. That zombies are conceivable is more plausible to him than any of those views, thus they can be dismissed outright. But I think this response is insidious, it cloaks its hubris in Moorean grammar.

Let's zoom in a bit on the Churchlandian-style eliminativism brought up in the first section of the post as an example. The Churchlands are naturalists through and through. The difference between their points of view and someone like Chalmers' amounts to more than the rejection of a few propositions Chalmers accepts (and perhaps our original post could have done a better job of emphasizing this). *Qua* Quineans, they are extremely skeptical of Chalmersian conceivability and its Archimedian perspective, the two-dimensional semantics with which it interlocks, the modal machinery in which zombie arguments are couched, the conception of qualia at play, the "golden triangle" of necessity and priority and analyticity, conceptual analysis, the import of conceivability, the notion of "logical supervenience" with which Chalmers characterizes reduction, the second dogma of empiricism Chalmers has revived, and much more. The point is this: someone who views philosophy of mind through a Chalmersian lens cannot and should not estimate the plausibility of a viewpoint as disparate as Paul Churchland's on something as miniscule as zombies. Any serious look at the issue is holistic: his view of zombies is one brick in a house, and not one brick goes untouched by problems in the philosophies of language and science, epistemology, metaphysics, ethics, etc. To state, as Adelstein does, that Churchlandian eliminativism is "extremely implausible" because of its denial of one loaded proposition is to appraise the house's value based on the appearance of just that one brick. It amounts to nothing more than reiterating one's current epistemic context in the register of modus-tollens. What the scope of our inquiry into consciousness is, the conceptual resources we have to work with, what questions we ask, and what methodologies we may use to answer them are all contingent on our perspective, and I worry that this treatment of the Quinean-*cum*-Churchlandian and the liberal naturalist perspectives is conducive to a problematic epistemic obstinance.

He uses this maneuver again in his treatment of Nigel Thomas' "zombie killer" argument:

> "This argument has been subject to various criticisms… notably, it if true would rule out the conceivability of zombies—but zombies are clearly conceivable."

Let's briefly recap Thomas 1998. Thomas asks us to reevaluate zombies by pulling from resources in the philosophy of language. He builds a trilemma for the proponent of the zombie argument: you can either undermine your own argument with skepticism (the "Falsity" and "Truth" horns that it seems Adelstein would contest); espouse some implausible metasemantic views, thereby reducing the dialectical strength of the argument (one way to take the "Meaningless" horn); or give up that zombies are "conceptually possible," in Thomas' words. The goal is to show that denying the conceptual possibility of zombies is a better option than any of the views required to sustain it. If Thomas' diagnosis is accurate, then we should give up zombies. To Adelstein, as he says explicitly, this itself constitutes a *criticism* of the argument. It aims to change his mind on zombies, but his mind is right, so the argument fails. But surely philosophical discourse is only worthwhile when we sincerely allow it to challenge and change us, *particularly* our certain beliefs. Arguments help us form out of molten plastic our conception and resolution of the issues at hand, they are not merely stewards that deliver to us on silver platters all of the bullets we must bite to never change. Perhaps the conceivability of zombies really is so obvious as to warrant this kind of conservatism. My next section will explore this claim.

4. It Ain't Obviously So

Lucas Collier

I've already pointed out that certainty comes dirt cheap. But what I'd like to argue here is that for most people (especially people like me and Adelstein that lack any extensive philosophical qualification) it probably should not be touted as *obvious* that zombies are conceivable. David Chalmers would be the first to admit that "conceivable" can be understood in any one of a great number of different ways. He himself separates conceivability into *prima facie* and ideal conceivability, negative and positive conceivability, and primary and secondary conceivability. Conceivability has been subject to countless other recent analyses from philosophers like Menzies, Kripke, or Yablo. The connection between each of these modes of conceivability and possibility varies greatly. The relevance of this point will become clear with a brief look to the night sky.

Hesperus and Phosphorus, Greek names for the evening and night star, are now known to be identical. This is taken by most analytic philosophers to entail that it is metaphysically impossible for Hesperus and Phosphorus to be non-identical. So I ask Adelstein: in the same sense of "conceivable" in which it is *obvious* that zombies are conceivable, was it, prior to the discovery of Hesperus and Phosphorus' identity, conceivable that they were non-identical? Let's say the answer is yes. This implies that the notion of conceivability Adelstein is using here does not *entail* metaphysical possibility. One interpretation of this weaker strain of conceivability is a near-ordinary language reading that defines conceivability with regards to our imaginative capacities. What is conceivable is what we can entertain or think deeply about. This is not what most philosophers mean when they say zombies are (in)conceivable and it does nothing for a zombie argument. Perhaps Adelstein means by conceivability *a priori* epistemic possibility. But he himself thinks Nigel Thomas' argument represents a real challenge to conceivability, one that is subverted with the adoption of a non-physicalist reconceptualization of our cognition. If working out what is *a priori* epistemically possible requires us to fall back on and adjust our larger philosophical picture, then conceivability looks somewhat relativized to epistemic contexts.

Let me make this point in another way. I take it that by "*a priori* epistemic possibility," most people have in mind something like this definition of conceivability from Balog 1999:

> A statement S is conceivable if it is consistent with the totality of conceptual truths, that is, if -S is not a conceptual truth. ("Conceivability, Possibility, and the Mind-Body Problem," p. 498)

The consequent of this conditional should raise some eyebrows about the claim that zombies are obviously conceivable. So long as we take one of the primary pursuits of analytic philosophy to be conceptual analysis, we take one of the primary pursuits of analytic philosophy to be *working out what the conceptual truths are*. What the conceptual truths concerning consciousness (and every concept it interacts with) are to Chalmers are not what the conceptual truths are to Ned Block, and these aren't Churchland's conceptual truths. In fact, many naturalists don't even accept that "conceptual truths" picks out a distinct class of truths–this has been a popular view at least since Quine's attacks on the analytic-synthetic distinction. If Adelstein has in mind something like Balog's statement of conceivability, then the claim that it is *obvious* that zombies are conceivable looks myopic or atomistic.

Let me give a little thought experiment to support this point. Suppose we start an investigation of consciousness as a Chalmersian and think there is a conceptual gap between phenomenal and physical facts such that an identity theory is *a priori* epistemically impossible. That is, it is not consistent with the conceptual truths. If our investigation eventually leads us to a drastic reconceptualization of our notions of phenomenality and physicality, and to an identity theory, has something *a priori* impossible become possible? If so, then *a priori* possibility seems to depend on the conceptual scheme or epistemic standpoint from which an idea is being considered. What is obviously consistent with the totality of conceptual truths changes with our concepts. If instead we were just *mistaken* about the *a priori* epistemic possibility of identity theory, then we do not necessarily have access to the *a priori* epistemic (im)possibility of things. In this case, our Chalmersian conceptual scheme would have misled us. I think this second option is incoherent when fully explained, but it suffices to say that it similarly makes claims of *a priori* epistemic (im)possibility revisable as we change our epistemic contexts, and so once again it seems irresponsible to claim any contentious *a priori* epistemic possibility is *obviously* so.

Chalmers' primary conceivability is also a mode of conceivability that does not necessitate metaphysical possibility, but it may strongly suggest it when certain conditions are met by the intensions involved (Chalmers 2009). One such marriage of primary conceivability and the right intensional properties can be found in the case of zombies. Does Adelstein mean to say that it is a "manifest fact" that zombies are ideally positively primarily conceivable? This is certainly much less intuitive. And primary conceivability is a face of Chalmers' generalized two-dimensional semantics. By whatever metric we use for obviousness, it is not *obvious* that Chalmersian generalized two-dimensional semantics is right. Construed in this way, zombies come downstream from much larger commitments, and it might be that the disagreement over zombies can be recast as a disagreement over philosophy of language and epistemology.

Maybe Adelstein means zombies are "obviously" conceivable in Yablo's sense of the word. But to Yablo this only gives defeasible *prima facie* justification for possibility–justification that may be overturned when we learn that Hesperus is identical to Phosphorus or that some mental state is identical to some brain state. This kind of conceivability is not strong enough for the zombie argument, so I'm not sure it is what Adelstein intends. Furthermore, it is at best unclear what our "extremely implausible" viewpoints would have to say about the conceivability of zombies in Yablo's sense.

Let's say the same sense of conceivability in which it is "obvious" that zombies are conceivable, it was not conceivable that Hesperus and Phosphorus were non-identical. "Conceivable" in this sense entails metaphysical possibility and fits into the pants it must wear in the zombie argument. But conceivability says nothing of our capacity to imagine when it is placed under the governance of metaphysics. Notions of conceivability which entail possibility have a problem of *modal error*: the non-identity of Hesperus and Phosphorus really was impossible before their identity was discovered, and people at the time were wrong to think they could conceive of it. If the realm of conceivable things is responsive to our (eventual) metaphysics, why should questions of conceivability be used to dismiss entire philosophical pictures such as the Churchlands'? The physicalist could just as well say that zombies are metaphysically impossible and so they cannot be conceivable in this stronger sense. We find ourselves in a modus-ponens modus-tollens tug of war—an athletic competition held only to celebrate our dialectical shortcomings. The conceivability-in-*this*-sense of zombies just redirects us back to considering philosophical pictures, and so it will have been at best a new way of characterizing an old disagreement.

I hope the above paragraphs make it clear that conceivability is very nebulous. The term is estranged from ordinary language and its relationship with possibility is a volatile one. The conceivability of zombies in particular has come under intense scrutiny. Robert Kirk summarizes a few of these objections in Kirk 2005 and Kirk 2023. For example, the conceivability of zombies has been challenged by neutral-monism (a view that Chalmers admits complicates his semantic analysis), views about reference and epistemic contact, anti-epiphenomenalism (or in Kirk 2005, opposition to a very particular epiphenomenal scenario), plausible views about intentionality/semantics, philosophies of perception, views about conditional analysis, and different explications of conceivability. Kirk notes that among philosophers there is live debate about how much the conceivability of zombies depends on these factors. He also reviews the arguments that Chalmers gives for the conceivability of zombies and shows how even those often rely on antecedently having the right set of intuitions or philosophy of language. Whatever we take conceivability to mean, it should be recognized that the conceivability of zombies is not cut-and-dried, and it should give us pause when someone views the issue as so black and white. As merely a small brick within our doxastic peripheries, zombies look like little more than spoils to the victor.

Misunderstandings

The primary aim of this section is simply to point out where Adelstein uncharitably rendered our arguments or ignored parts of our post. There are certainly some cases where we can charitably assume that this skewed presentation is due either to a lack of clarity in our original post, or an innocuous mistake on Adelstein's part. However, when one considers every example we outline here, it becomes difficult to believe that Adelstein's post does not overstep the reasonable line into the realm of intentional lack of charity and misrepresentation. In fact, it's probable that Adelstein failed to read much of Cutter & Crummett's paper and our post to begin with. The secondary aim of this section is to provide responses to his objections, insofar as they can be fairly constructed as best we can muster.

1. The 'Concerns' Section

Sebastian Montesinos, Benjamin (Truth Teller), and Joseph Lawal

Adelstein's overarching response to this section of our piece is to point out that the most of the views we list are either unpopular in philosophy, 'radical', or have highly unintuitive implications. We have already presented our case against intuition-mongering in sections 1-4 of the previous heading, and need not repeat it here.

However, this is not even the main issue with what Adelstein wrote. The main problem with his response is that he entirely omits the primary point of the section in question. We harbor no illusions that some of the views we listed are considered radical or are unpopular. In fact, we all reject at least one of (and usually more than one of) the views we listed! The point of our original piece was, in part, to examine what makes an argument effective. Specifically, we were interested in the extent to which psychophysical harmony should be a threat to naturalists given different interpretations of what makes a philosophical argument successful. That is why we devoted an entire section to explaining the dialectical relevance of these alternative views in the context of the question of what makes a convincing argument. It is also why we split our first post into two broad sections, only the second of which was called "Objections," and used our introduction and conclusion to explain the *contextual goal* of each respective section.

See the following excerpts from our post (and see the full section for the reason we come to these conclusions):

"We now examine what the dialectical implications of these positions are…*in the strong sense of 'effective' in which an argument can move someone's credence such that they adopt a new position*, psychophysical harmony is inert…*if a naturalist* is presented with psychophysical harmony and doubts no part of the argument, it will almost certainly *be more rational* for them to remain a staunch atheist, and simply adopt one of the many naturalistic or non-naturalistic hypotheses consistent with atheism on which psychophysical harmony is not a problem… Of course, [C&C] could argue that every one of these views is false on the basis of other arguments, but that would make psychophysical harmony *as an argument for theism* reliant on the success of highly controversial arguments in philosophy of mind, metaphysics, metaethics, philosophy of language, and the philosophy of religion, which is something Cutter & Crummett seem to want to avoid…[however], even if psychophysical harmony does not work *as an argument for adopting theism*, it may still show that we ought to assign a much lower credence in naturalism than previously thought, and thus assign a correspondingly higher credence in theism"

Adelstein does not try to reconstruct the dialectical point we were trying to make, instead he fails to even mention it and says things like "[these views] are all extremely implausible," "if the atheist is forced to axiarchism…this argument has still accomplished a lot," and later when responding to a statement in our first objection that references the concerns section, "The fact that an extreme minority view, whose opponents often find it utterly crazy, explains [PH] should be of little comfort to most people." Put aside the fact that, if Adelstein is implying that the credence most atheists will assign to the inclusive disjunction of the views we listed is extremely low, we certainly disagree. Our point was that we suspect that when most *naturalists* look at their web of beliefs, they will almost always prefer one or some of these views *over theism*, and that this has implications for a certain kind of *dialectical claim* about the psychophysical harmony argument. Note that this is consistent with a naturalist finding these views enormously implausible in general, so long as they find theism as or more implausible. He even reiterates a claim that we make—that there is a sense in which the argument has still "accomplished a lot"—as if it were a response to what we said, despite the fact that *we said it ourselves*, and it is therefore clearly not in tension with our own aim in this section.

Now, Adelstein does say the following: "Whether [these views are] more implausible than theism will depend on one's credence in theism…but I find [them] incredibly absurd." This is the closest Adelstein gets to referencing our point, and if Adelstein intended merely to point out that these alternative views do not help *him* because *his* subjective credences cut overwhelmingly against them relative to theism, there would be no problem with what he said. However, it is then inappropriate for him to claim that the concerns about PH adduced in our piece are 'lousy' or 'bad'. What he should say is that they are no help to him because they are in tension with his personal distribution of credences. Since we freely admitted that they would not help some people, this has nothing to do with our point in that section. What Adelstein does (and this is a pattern throughout his response) is to equivocate between the extent to which a counter argument matches his own subjective intuitions and credences and the extent to which a counter argument has any merit *in general*. These two things are obviously not the same thing. Adelstein therefore needs to either make it clear that his claims were merely about his own psychological states, or he needs to address the relevant dialectical context in which our argument was offered.

Most importantly, it was inappropriate for Adelstein to not even mention or try to reconstruct that point we were making, and to offer responses to this section as if they engaged with our point when they did not. A naive reader would make the wrong assumptions about our claims in that section if they only read Adelstein's piece, because he omitted any discussion of our dialectical aims, and what he did say implies that our arguments were different than the ones we made. They likely would have erroneously assumed that a) our argument in this section relied on the views we listed being probably true and b) that we were trying to offer them as an objection to the PH argument given a strong understanding of 'objection' that we clearly did not intend.

Having put aside that criticism of Adelstein's approach to this section, here are a few specific misunderstandings:

Adelstein lumps together eliminative materialism, liberal naturalism, and type A physicalism and says they are all implausible because of their denial of the conceivability of zombies and inverted qualia, and their denial that Mary learns something new. Putting aside the fact that, as we addressed in the section on Moorean shifts, this is an inappropriate approach to rejecting these views, *it is not even an accurate account of the implications of these views*. Paul Churchland, the godfather of eliminative materialism, is also one of the godfathers of the physicalist response to Mary's room wherein Mary *does acquire knowledge* upon seeing red. Nothing about liberal naturalism *per se* commits a liberal naturalist to any of the views which Adelstein attributes to them. We provided two examples of liberal naturalists—Hilary Putnam and Donald Davidson—who do present reasons to doubt the conceivability of zombies. But the point of mentioning these two philosophers was not to make some general point about liberal naturalists, but to show that one need not be a type A physicalist to reject certain moves on which the psychophysical harmony argument depends. We also gave no reason to think either Davidson or Putnam respond to the knowledge argument by saying Mary doesn't learn anything; it's unclear why Adelstein makes this attribution. Additionally, Adelstein makes no mention of Kripkean conceivability as a challenge to his reliance on the conceivability of zombies.

Adelstein argues against error theory, but the argument he gives is essentially identical to the one Cutter & Crummett provide, and that we explicitly respond to. Cutter and Crummett say that pain's badness is "self-evident" and Adelstein says that the statement "pain is worse than pleasure" is "obviously true." Our counter was that these kinds of responses "trade on the ambiguity between stance-independent badness and badness." About a year ago, Adelstein and one of the authors of this post had a back and forth about moral realism where a similar point was made. Whether or not this is a good response (we are no longer even sure about the version of it that we gave, at least with respect to error theory specifically), the fact that Adelstein does not even address it again demonstrates the carelessness with which he approached our piece. As an additional note, our point was that moral anti-realism in general may pose a problem for normative harmony, a much larger umbrella than error theory alone.

Adelstein goes on to claim that since theism entails that error theory is false, theism predicts that pain leads to aversion behavior. This misses the point. It is not being disputed that on theism there are normative facts, and that theism predicts, insofar as there are sentient creatures whose consciousness and dispositions are governed by psychophysical laws, that such laws will be normatively good. The point is that the psychophysical harmony argument proceeds with the assumption that normative harmony (our behavioral states line up with our psychological states in a normatively valuable way) is a datum in need of explanation. Yet, that this datum obtains is precisely what the error theorist denies. For the datum of normative harmony at least, there is nothing to be explained for the error theorist. Nonetheless, we anticipated that there may be a way to state the psychophysical harmony argument that is consistent with error theory, and what we say about this looks quite similar to what Adelstein said (see footnote 1 in the concerns section). Why would Adelstein bring up this objection but fail to state that we were aware of it?

2. Normative Harmony

Sebastian Montesinos

Adelstein's response to my argument omits the most important part of the second horn of the dilemma I presented, and, as a result, he ends up giving a non-response to it. Let's look at what Adelstein said:

> "[Montesinos] disputes that we can have reliable judgments about this because our judgments are just based on pain in the real world. But even though our judgments are based on pain in the real world, we can see the obvious fact that our behavioral disposition isn't the reason that it's bad. That disembodied minds being tortured is bad is the most obvious thing ever! If you found out that you were a brain in a vat for your entire life, your pain would still have been bad."

Adelstein says that my view is that our judgements about pain are just based on pain 'in the real world'. That is not an adequate presentation of my view. The view I presented is that our judgements, beliefs, attitudes, and the concepts therein are all embedded in a broader 'conceptual framework', a background theory for explaining a phenomenon of interest that governs the very way we talk about it. In the case of our mental states, we all inherit a folk psychological framework for understanding and explaining these states. Any propositional attitude we form about pain will inevitably be sculpted through this contingent folk theory. To understand this point, it will be helpful to add that this view is usually holist in nature: the idea is that the conceptual framework is web-like, and everything that features therein is shaped by, and only made sensible in virtue of, its relative place within that web.

Turning now to pain, one of the judgments we make about it is that it is bad. However, as per the view sketched above, this judgment can only be made sensible *in virtue of the background conceptual framework in which it is embedded*. So too for our understanding of pain, since *the very concept of pain inherits its meaning from its place in the conceptual web*. What this implies is that our judgements about pain are contingent on the particular structure of the conceptual framework in our world. In a world in which all of the implicit folk psychological laws were reversed, our conceptual framework would be reversed, and our judgments about pain would also be reversed. Therefore, our judgments about pain are crucially dependent upon the conceptual framework in our world, and we should not take for granted that they would hold true in worlds with radically different conceptual frameworks, or worlds that lack such frameworks entirely. In other words, on this account, justification is a relationship that holds on the basis of a judgment's place in a contingent web, such that this relationship is world-indexed.

Depending on how one interprets what Adelstein said, he either misrepresents my point or fails to interact with it. If he means to imply that my view relies on the claim that when people think about pain internally, they come to have the belief that pain is bad because of its behavioral effects rather than how it feels—then this is a misinterpretation. The point is that our judgments about pain are dependent on the conceptual framework they are filtered through: alter that framework, and our judgments will change. *This is not a psychological claim about the way people reason to pain's badness, and it does not require that we reason to it in any particular way.* All this claim requires is that our understanding of pain and judgments about it derive their meaning and reliability from their place in a broader folk theory. Our judgments about pain are based on its functional profile *only in the sense that* that profile plays a role in shaping the conceptual framework through which our judgments are filtered. This does not entail that we will actually *consciously hold the belief* that pain is bad because of its functional profile.

On the other hand, if Adelstein means to argue that our belief about pain's badness in non-actual worlds is reliable just because we hold the belief that pain's badness does not rely on its functional profile, then he has not interacted with my argument. The very point in contention is whether that belief (and others like it) is reliable when it is pulled out of our conceptual scheme and foisted onto worlds where that scheme is absent, skewed, or inverted. It simply falls out of the view I gave earlier that it is not. That belief is also part of the conceptual framework in our world, it also inherits and earns its keep due to its place in a contingent web, and it would also be inverted in the world where 'pain' and 'pleasure' are swapped in our conceptual frameworks. This is why the point about disembodied minds doesn't fly: in order to make a judgment about the badness of their pain, we have to think that judgments that are only made sensible in light of our contingent conceptual framework would hold without them.

In sum, the judgment that our behavioral dispositions are not what makes pain bad, when this is construed as *a belief we have about why pain is bad*, makes no difference to my point. On the other hand, if one construes the judgment that our behavioral dispositions are not what makes pain bad *as a denial that we possess have a contingent conceptual framework that shapes our judgments, or a denial that this framework is shaped by functional laws,* it does make a difference to my point, but it amounts to a either a rejection of the very premise of the horn in question, or a rejection of the way the psychophysical harmony argument works to begin with.

The above is why the way to object to my argument is just to deny the very premise of a 'folk psychological web' that is used to filter our judgements, and claim that we directly access our mental states without any theory-laden filtering. This is, as I understand it, the whole idea behind phenomenal introspection. But, again, this would be to deny the very premise of the horn in question. As I mention in the original piece, I think there are powerful reasons to suspect that our judgements about our own mental states are just as theory-laden as our external observations, and that phenomenal introspection is incoherent or false, depending on how it is interpreted. Those who reject this will not be able to utilize my response to normative harmony. However, this view, or something like it, is very much mainstream and plays a critical role in post-Quinean and naturalized views of knowledge. Therefore, it is a solid place for naturalists to stand against the psychophysical harmony argument.

I will not comment extensively on Adelstein's response to my first horn, because I no longer think it is a strong place to stand, albeit for reasons very different from those that Adelstein provides. In particular, I have become increasingly skeptical both of the 'direct introspection' of phenomenal states and of the kind of intuitive methodology that motivated the argument in that section, and I would therefore rest my chips almost entirely on the second horn of my argument. That being said, if I were to defend that horn, I would point out to Adelstein that the intuitiveness of pain being intrinsically linked with aversive behavior is grounded in our reflecting on the idea that a feeling so horrible as pain could just as easily lead to enjoyment, content, and seeking behavior. This is what is meant to look very bizarre introspectively.

3. Semantic Harmony, Causation and Explanation

Joseph Lawal

In our original response to psychophysical harmony, I suggested that the argument from semantic harmony was insufficiently attentive to the question of content individuation, and that incorporating arguments originally from the philosophy of language, such as Putnam's original Twin Earth case and Burge's modified Twin Earth cases, casts doubt on the coherence of the claim Cutter & Crummett are making. Adelstein makes two responses to this objection. The first reads as follows:

> Semantic harmony is not about judgments about consciousness, it's about reports about consciousness. Even if you couldn't think you were having a reddish experience without having a reddish experience, you could still obviously say you were having a reddish experience. Why that doesn't occur cries out for explanation.

This objection is puzzling, not least because it is demonstrably false. In fact, it is difficult to see how one could arrive at this conclusion given not only what I wrote in the original objection, but given what Cutter & Crummett themselves write. Here is the summary statement of the nature of semantic harmony by Cutter & Crummett, which I quoted at the beginning of my objection:

> In many cases, the psychophysical laws pair phenomenal states with physical states in a way that generates a semantic correspondence between our *judgments/reports* and our phenomenal states. (Cutter & Crummett, forthcoming, p. 13, emphasis added)

Cutter & Crummett here (and throughout the section on semantic harmony) do not distinguish between judgments and reports as far as the argument is concerned. The harmony is between our judgments and/or reports and our phenomenal states. This is explicit in the original paper. In the section of my response which Adelstein quotes, I drop reference to reports for convenience, following Cutter & Crummett in treating judgments and reports as, for all intents and purposes, interchangeable. But it is clear both in the earlier part of my objection and, more importantly, in Cutter & Crummett's piece, that semantic harmony is intended to apply both to judgments and reports.

Setting Cutter & Crummet's original point aside, can Adelstein's response to my objection stand on its own terms? No. Adelstein's claim is that, even granting that a person in a disharmonious world couldn't think about having a reddish experience, she could still *say* she was having such an experience. But of course, that is exactly what the externalist denies. In fact, Adelstein seemingly ignores externalism about mental content and focuses on semantic externalism, but it is semantic externalism which is relevant to verbal reports. My objection was that, if externalism is right, it does not seem that someone in a disharmonious world could make the relevant report (or judgment). Of course, by stipulation, the person in the disharmonious world is *making the sounds* which correspond to the sounds I would make in reporting "I am having a reddish experience." But if Adelstein thinks that this is equivalent to making the report I make, then he has not understood the objection.

Part of the problem here is that Adelstein does not seem to understand the position I'm advocating, or at least is incautious in his restatement of it. My objection concerns externalism regarding content. As I note in my original objection, Twin Earth thought experiments were initially employed by Putnam to advocate for externalism with respect to semantic content - that is, the content of the things we say. Only later, thanks in large part to the work of Tyler Burge, were these thought experiments extended to apply, not just to the things we say, but to the things we think, to mental content. Adelstein seems at times not to be entirely clear about this distinction. At one point, he notes, correctly enough, that I am suggesting that "what a thought is about depends in some way on external reality." But he then goes on to describe the Twin Earth case as one in which my twin and I are thinking about different things "even if we have the same thought." But this is precisely what is being denied - my twin and I *don't* have the same thought, on an externalist account, because my thought *constitutively depends* on something on which my twin's thought does not. The whole point is that we have different thoughts - my aim isn't to be pedantic here, but to note that Adelstein at times seems to miss the entire point of the objection.

Adelstein's second objection fares little better. Here it is:

> Most worlds that don't have semantic harmony don't have people having beliefs about their mental states. Most of them just have no ordered consciousness—maybe they'll have people only with the phenomenology of eating Cheetos, for instance. So even if you can't have mistaken judgments about having reddish experience, this doesn't explain the vast improbability of having such judgments in the first place.

Again, this response seems to be at odds with what Cutter & Crummett are actually trying to argue. For instance, they say:

> If there hadn't been psychophysical laws correlating our physical states with distinct, non-physical states of consciousness, we would have made the same reports and judgments, but they would have been false. (p. 14)

Cutter & Crummett *stipulate* that the reports and judgments remain the same in the relevant scenarios, but that they come out false in those scenarios. No mention is made of psychophysical laws rendering us unable to make judgments at all, or unable to form any beliefs at all. Our mental lives are not just made up of conscious experiences. Even in a world, to take Adelstein's example, in which people only have the phenomenal experience of eating Cheetos, they still have lots else going on, according to Cutter & Crummett, including forming all the same judgments and beliefs which we form (they just happen to be false, at that world). To even raise the problem of semantic harmony, Cutter & Crummett need there to be two things *which are harmonious*; they are not trying, as Adelstein apparently is, to call into question whether one of the things which is supposed to be in harmony with another exists at all at disharmonious worlds.

In our response to Cutter & Crummett, I also argued that the psychophysical harmony argument was not well suited to pose a problem to anyone who accepts that the mental is causally efficacious:

> The point here is that for anyone who denies epiphenomenalism, it is at best questionable whether the core of the argument from psychophysical harmony has any force at all. The argument depends on my being able to conceive of pain causing me to smile, etc., but the mental's being causally efficacious seems to undermine the conceivability of such a state of affairs.

My claim was that it is far from clear that the authors managed to present a genuinely conceivable state of affairs in describing a world in which, for instance, pain and pleasure are inverted and the pleasurable mental states cause avoidance behavior etc. This scenario is *prima facie* conceivable (there's nothing *obviously* wrong with just stipulating that my pain and pleasure states are switched), but upon reflection, the conceivability looks, at best, strained. The main problem with Adelstein's response to this objection is one that plagues much of his piece: he appeals to what are, to him, obvious truths to dismiss rather than properly engage with the actual argument.

Adelstein presents three responses. The first is an especially egregious example of the problem I just mentioned:

> If interactionists are committed to this, then they should give up on interactionism! Because inverted qualia, even radically inverted qualia is obviously possible. It's not hard to imagine a world where you act as you do but have radically different experiences.

The problem with this kind of response is that the whole point of my objection is to explain why this is hard to imagine. Waving a hand and saying "well it is easy to imagine" doesn't advance the dialectic in any way, since my argument acknowledges that some people think that an inversion of pain and pleasure is easy to imagine, and attempts to show that that is an illusion. There's little for me to say here because Adelstein didn't actually address, e.g., the case I raise concerning a person who experiences extreme pain from a backrub, thinks "this is painful!" and determines to avoid backrubs in the future, and then is horrified to discover that his body goes on seeking it out in spite of that determination. This kind of case raises serious doubts about the coherence of the scenario Adelstein thinks is obviously conceivable, and Adelstein has not a word to say about it.

Adelstein does have a fallback: even if my point is right, he thinks, the problem of psychophysical harmony persists. Perhaps the kind of scenario I just described can't be thought of as one in which pain causes me to seek out a stimulus, but presumably we can still imagine an epiphenomenalist pain-pleasure inversion scenario. If so, the argument goes, then

> the fact that the actual world is genuinely interactionist becomes miraculous! Whether the disharmonious worlds count as interactionism [sic] is irrelevant to whether they're possible worlds, and if they are possible, their probability utterly swamps that of the harmonious worlds.

Since Cutter & Crummett are interested in epistemic and not metaphysical possibility, I assume that Adelstein's appeal to possible worlds here is intended to involve epistemically possible worlds. This response then broaches complex issues which it would be difficult to deal with entirely here. The key question is this: should committed non-ephiphenomenalists think that epiphenomenalism is epistemically possible? It is hard to evaluate this question without a clearer idea of what Adelstein (and perhaps Cutter & Crummett) have in mind in talking of epistemic possibility and epistemically possible worlds. It is clear that Cutter & Crummett do not intend to take the route Adelstein here takes of insisting that non-epiphenomenalists are susceptible to the argument from psychophysical harmony on the grounds that the number of disharmonious epiphenomenalist worlds swamps that of harmonious, non-epiphenomenalist ones.

Finally, Adelstein takes issue with the general account of causality on which my argument depends:

> Lawal's view commits him to an implausible view of causality. In the case where whenever one experiences pleasure from an activity they act to avoid it in the future, the pleasure still causes their aversion—if they didn't experience the pleasure, they wouldn't have done it. This is certainly true on counterfactual accounts, and on pretty much all other causal views.

To be frank, I hesitate in even interpreting this response because I struggle to see any charitable interpretation. Views on causation vary widely, and the counterfactual account of causation which Adelstein explicitly mentions is at least tentatively accepted by only 37% of philosophers according to the 2020 philsurvey (Bourget & Chalmers, 2020). The bare fact that the account of causation on which my argument depends *conflicts with the counterfactual account* hardly seems like good grounds to call it implausible, whether by appeal to a consensus (since there clearly isn't one) or because the counterfactual account is so clearly superior to alternatives despite the lack of consensus (the view has major problems like any other contentious view in philosophy, and Adelstein has given us no reason at all to prefer that account). Presumably, most accounts of causation aim to capture the data we get from examining how we think about cases of (purported) causation. If we don't judge that one thing has caused another in a particular case, like the pain-pleasure inversion scenario described above, then that may be relevant to an account of causation, often as a counterexample to an attempted analysis in the way that Gettier cases are meant to provide a counterexample to certain attempted analyses of knowledge. Exactly the wrong move in this kind of context is to say the example is clearly not a problem because it conflicts with a favored account; when the analysis of causation is grounded in judgments about our application of the concept of causation, this move clearly begs the question.

4. Disharmonious Doubles

Sebastian Montesinos and Lucas Collier

Adelstein's response to this section is excessively unclear. He says very little, and what he does say constitutes an indirect response, leaving us to reconstruct from scattered hints an interpretation of his counter-argument.

Our original argument was that there is no good way to interpret the mental states and reports of people in disharmonious worlds that leaves the psychophysical harmony argument in-tact. Our piece was inspired by a paper from philosopher Nigel Thomas called *Zombie Killer*. In that paper, Thomas presents a trilemma for advocates of 'philosophical zombies'—and in our piece, we argue that a similar dilemma exists for those who advocate the psychophysical harmony argument. Instead of explaining what horn of *our* argument he is attempting to object to, he links to another post on his blog that he says refutes Nigel Thomas' argument. That post never mentions Nigel Thomas. However, we assume that he must be referring to the final section in that article, which references yet another piece, this time a paper from philosopher Helen Yetter-Chappell, in which she discusses an argument that looks similar to an argument Thomas makes. To sum up, Adelstein provided an indirect response to one part of the article that inspired our argument—and he makes no attempt to explain exactly what part of that argument his linked article responds to, let alone how that critique is supposed to extend to the argument that we make. It would have been helpful if Adelstein could have made this explicit.

Nonetheless, let's take the section in the piece Adelstein links to that we believe is meant as an indirect response to Nigel Thomas, and combine that with the one short paragraph he does write about our argument. Does his counter to our argument now become clear? Unfortunately, it remains extremely confusing. Here is what Adelstein says:

> "[Lucas and Sebastian's argument] assumes erroneously that if we have non-inferentially justified beliefs then our inverted twin would have the same non-inferentially justified process because they form their beliefs as part of the same cognitive mechanism. But all non-physicalists should deny this assumption—only conscious beings have beliefs and those constitutively depend on their mental states, rather than their physical brain states."

As a preliminary note, remember that the psychophysical harmony argument does not ask us to imagine zombies who retain our intentional states but lack phenomenal experiences, but instead disharmonious doubles who retain our intentional states yet have discordant phenomenal experiences. At the beginning of this paragraph, Adelstein is talking about our 'inverted twins', who would have conscious experiences. Yet, he then goes on to say "only conscious beings have beliefs," as if he were talking about zombies. But no one was talking about non-conscious beings. Our inverted twins would be conscious. Also, Adelstein starts off by talking about the process that forms the beliefs held by our inverted twin—yet, by the end, he is talking about their beliefs. Is he saying that the process would be different, or that the belief itself would be different, or that both would differ?

Adelstein says that our beliefs "constitutively depend" on our mental states. But Adelstein never says what the notion of constitutive dependence amounts to for him: could the phenomenal belief "I just saw redness" follow an experience of greenness, or does the belief have a necessary connection to actual quale which it is supposed to be about? If so, what is the notion of constitutive dependence at work here that grants that kind of modal security? Would the consequent difference in phenomenal beliefs between disharmonious and harmonious worlds require that there is a distinct cognitive mechanism? Additionally, Adelstein is an epiphenomenalist, so he thinks mental states are causally inert. Therefore, the relationship that he claims exists between our beliefs and our prior mental states is utterly mysterious. It certainly can't be causal, and it can't be the kind of constitutive dependence Joseph espouses, lest Adelstein wants to render semantic disharmony inconceivable.

In the paper Adelstein links to, he essentially reiterates the claims made by philosopher Helen Yetter-Chappell in a paper on epiphenomenalism. Adelstein quotes a section from that article which claims that the mechanisms that lead to our beliefs are different from the mechanisms that lead to a zombie twin's beliefs. But this is unclear. Mechanistic explanation is characterized first and foremost by its explanation of an outcome in terms of its interacting parts, and secondarily by referring to a higher level system than be decomposed and localized, specific causal detail, and an emphasis on the force, action, and movement in causal relationships (Ross, 2021). How this is supposed to work with, let alone be compatible with, a non-physical kind of causation is left unexplained. Furthermore, mechanistic explanation is causal: but Adelstein denies the causal efficacy of conscious states! So how can he claim any mechanism is involved to begin with?

Please note that, at this point, we are not attempting to provide any *counter-argument* to Adelstein. We are simply pointing out how what Adelstein says is so frustratingly unclear, and leaves far too much unsaid. He punts to other authors, but any reconstruction of what they are saying in the context of our argument is not even attempted. Since Adelstein failed to make where and how his objection worked unclear, and since to the extent that it is stated it remains opaque, we are tempted to just leave our response here. However, we have instead attempted to charitably reconstruct what he is trying to do as best we can. Here is what we *think* Adelstein's counter is supposed to be:

One horn of our argument is that our doubles in disharmonious worlds are simply mistaken when they make false utterances about their phenomenal states. One of the points we make about this option is that if it is true, it appears to remove our warrant for concluding that our world is semantically harmonious. The reason for this is that because our world and the double's worlds are physically-functionally identical, the correlate or cognitive process used to associate our double's judgment and report about their experience with their actual experience is the same as it is in ours. We argue that this is one factor (along with others) that implies that we lack warrant in concluding that our world is harmonious. What we think Adelstein is arguing is that, in fact, the process that leads to the belief about the nature of our experiences in both worlds need not be the same. If dualism is true, this process might be non-physical and differ between the actual world and the disharmonious world, despite them being physically-functionally identical. Given this reconstruction, we respond as follows:

Adelstein never actually explains how the possible existence of these different mechanisms diffuses the worry of skepticism. We're not sure it does. This will hinge on how exactly we ought to interpret the nature of a non-physical mechanism, which as we already noted, is unclear in Adelstein's piece. Let's use an analogy often used in philosophy of mind: the conceptual separation of C-fibers firing from the quale of pain—a possibility that forms the basis for the argument from psychophysical harmony. Assume that physicalism and functionalism are true. In the harmonious world, a person touches the hot stove which activates the *mechanism*, C-fibers, which outputs a pain quale. In the disharmonious world, when a person touches that stove, *the very same mechanism* outputs a different quale, say, pleasure. If this interpretation of a mechanism is analogous to the dualistic version, then the same is conceivably true of the dualistic mechanism in question: the very same non-physical process that takes one from their experience of greenness to their belief that they experienced greenness could conceivably take one from an experience of greenness to the belief that they experienced redness. Since Adelstein does nothing to explicate the notion of a non-physicalist mechanism or process, he ironically leaves the conceivability of this scenario more open than the conceivability of the physicalist equivalent upon which the psychophysical harmony argument rests. Here's is the key point: Adelstein hasn't diffused our argument by stating that there are disharmonious worlds that contain different belief-producing mechanisms than those in harmonious worlds, because there are also conceivable disharmonious worlds where *the same* belief-producing mechanisms lead to our beliefs—and how do we rule out that we are in *those* worlds. He must make his claim stronger—that it is *inconceivable* that the same mechanism at work in the harmonious world is at work in its disharmonious double.

But, even the above is not enough. So far, we have only considered the possibility that there are conceivable *dualistic* worlds in which the mechanism remains the same. Suppose that Adelstein establishes that this is inconceivable. One of the assumptions upon which the psychophysical harmony argument is predicated on to begin with is that we can imagine disharmonious *physicalist* worlds. In these worlds, the mechanisms in question are physical and therefore *must* remain the same as those in physicalist harmonious worlds. So, even if we can rule out that we are not in the disharmonious dualistic world where the underlying mechanism is different, so long as there are epistemically possible *physicalist* worlds where the mechanisms remain the same between harmonious and disharmonious worlds, our skeptical argument stands.

Suppose that Adelstein is able to diffuse both of these points by establishing that it is inconceivable that the mechanism that produces red beliefs in the disharmonious world is the same as the one that produces green beliefs in the harmonious world. In that case, it is still unclear how the difference in these non-physical processes rebuts our reason for skepticism. Our disharmonious double would also believe that they are in the world where the underlying mechanism led to them thinking that their world is harmonious. So, how exactly are we meant to know, even introspectively, that *we* are not mistaken? In the harmonious world there is a cognitive mechanism M_1 that takes us from our experience of redness to our belief that we experienced redness, and in the disharmonious world there is a different cognitive mechanism M_2 that takes us from our experience of *greenness* to the belief that we experienced redness. How could we tell which cognitive mechanism we, in fact, possess? The faulty cognitive mechanism M_2 would produce the same doxastic states as M_1, and it is precisely these doxastic states which would inform our analysis of our world. You may stomp your foot and say "but unlike my disharmonious double, I *had* the experience of redness." But your double is just as adamant that they did too. Adelstein will need to elaborate on exactly what view of justification & introspection he is taking here, and how this view is supposed to avoid our skeptical concern.

It is worth noting how far Adelstein would need to come at this point to refute our argument. First, he would need to establish the inconceivability of physicalism. Second, assuming he does that, he will need to commit to some kind of 'direct, non-theory laden introspection of our mental states' theory that will avoid skepticism—assuming such a view even does help avoid skepticism to begin with. If *these* are the views needed to save the psychophysical harmony argument, our post certainly did its job.

Adelstein also says that a belief about a phenomenal experience must "constitutively depend" on the experience which it is about. Perhaps this can bail him out? But it seems to prove too much. If our belief that we experienced redness must bear some relation to the redness of our experience, it's hard to see how this doesn't render semantic harmony inconceivable for much the same reasons as the kind of content externalism discussed by Joseph, who also espoused the view that our judgements about our experiences bear this kind of constitutive dependence to our experiences themselves. If the belief about the qualities of our experience must "reach out" to the experience itself, then our disharmonious doubles cannot have the belief that they experienced redness if they did not.

In our view, the above constitutes a response to the most plausible reconstruction of Adelstein's point. The other ways we attempted to understand his argument looked much more confused, and we therefore dismissed them. For instance, when Adelstein responded to Joseph, he incorrectly suggested that semantic harmony was merely about verbal reports. In other words, in a disharmonious world, we experience greeness, this experience leads to a correct belief about greenness, and yet we report that we saw redness. Did he mean the same when responding to us? Since what he writes about our section references the *beliefs* our disharmonious doubles have and the processes that lead to them, we don't think so. Nonetheless, if that is what he meant, he was not talking about semantic harmony as it is understood by Cutter & Crummett—and it is not a version of psychophysical harmony that is consistent with physicalism, since it implies that while reports are constants across worlds, our internal intentional states are not, which implies a physical difference between these worlds.

5. Understated Evidence

Sebastian Montesinos and Benjamin (Truth Teller)

Adelstein provides two responses to this section. His first point implies that he did not read this section carefully, the second entails that he did not read it at all.

His first response is that any theodicy that explains evil will also explain psychophysical disharmony. This is clearly untrue. We cited a wide array of different kinds of disharmony and most of it is not even *prima facie* explained by any theodicy we know of. To illustrate, let's use just one example of disharmony we cited: that our perceptual beliefs about the richness of color perception in the periphery are false. The soul-making theodicy says that evil exists because it allows us to gradually build our ability to manifest important virtues. There is no obvious virtue that having incorrect beliefs about color in the periphery builds. The free will theodicy says that evil exists because it is a necessary by-product of having free choice. There is nothing about having wrong beliefs about color in the periphery that promotes free choice. Anyone can read our original piece and ask themselves the same questions about the rest of the disharmony we cited, and it will be obvious that no standard theodicy clearly explains the vast majority of it. In fact, the reasoning behind many mainstream theodicies *undermine* our expectations for disharmony because many instances of disharmony that we cite a) frustrate our ability to utilize our cognitive mechanisms for good (see our first and second points in the original piece), and b) display the inherent limitations in humans that prevent our ability to build virtue in the first place (see the final two points in the original piece).

Adelstein's second response to this section is egregious. He makes the following counterargument:

> “Even if there isn’t perfect harmony, there’s more harmony than almost all psychophysical laws. Virtually all sets of psychophysical laws will produce radical disharmony—the fact that the world is almost entirely harmonious is quite miraculous. If we’re in the top .000…1% most harmonious laws, then this is strong evidence for theism, even if we’re not sure why we don’t have the most harmonious laws.”

This response is essentially the same as Cutter & Crummett’s response, and we dedicated not just some minor paragraph to responding to it, but an entire section of the post (“Why Cutter & Crummett’s Response to Understated Evidence is Insufficient”). Here is the opening sentence in that section:

> “At this juncture, a friend of the psychophysical harmony argument might retort that while theism does not predict our particular distribution of psychophysical mappings, theism predicts those set of distributions which are orderly enough, coherent enough, normatively good enough…whereas naturalism (or indifference) is so bad at antecedently predicting orderly, coherent and morally good psychophysical laws, that theism still gets the advantage.”

We go on to explain why this response is insufficient in detail. We will not reiterate this response here, other than to point out that we explained exactly why what Adelstein says is not enough to motivate the kind of prediction the PH argument requires. What we say is especially important in Adelstein’s case because his entire article, not just the PH section, is riddled with the very error we point out there: like many theists, he does not think carefully about why something should be expected on theism other than that it is ‘good’. This is not an adequate way to determine whether some set of phenomena is expected on theism (see the upcoming version of ‘Why I’m an Atheist’ on the NaturalismNext blog for a discussion for how to make predictions on theism). By the way, this is not a comment on our disagreement with Adelstein’s conclusions—in the view of one of the authors of this post (Montesinos), two of the three pieces of data Adelstein cites in his post constitute the best evidence for theism—the issue is the lack of care in getting there.

One additional point worth making about our counter is that it is more devastating in virtue of the possibility of views other than theism or naturalism: of all the views on the table, directed and theism-adjacent views that are neither indifferent with respect to harmonious distributions (like naturalism) nor epistemically tuned towards only the very most harmonious worlds with special restrictions on certain kinds of disharmony (like theism) are clearly the best hypotheses on offer with respect to the disharmony we cite. Adelstein does not address this problem in his post.

6. ***The Revenge Objection***

Benjamin (Truth Teller)

The Revenge problem is the last objection to the PH argument presented in our piece. Adelstein construes the problem as saying that theism doesn’t actually explain the phenomenon of psychophysical harmony, it merely assumes it by locating it in God. This is not entirely correct, at least, it isn’t how I run the objection. The objection can be more accurately stated as saying that the same reasons that motivate a central premise of the psychophysical harmony argument equally motivate the claim that theism should be assigned a correspondingly lower prior probability before conditionalizing on the data in our Bayesian machinery. The idea is that if it is radically improbable that our psychological states line up with our intentions and behavioral dispositions in strikingly harmonious ways, so too, is it radically improbable that God's psychological states and causal powers link up with states of affairs external to Him in strikingly harmonious ways. This objection was well-known to Crummett and Cutter prior to our writing the post, so the main thrust of my section on the revenge problem was to address some of the counter-objections that could be adduced, including the one Crummett and Cutter are sympathetic to, and a couple others found in Apologetics Squared's video.

The first objection to the revenge problem discussed is the idea that theism might be conceptualized in a manner that negates its *a priori* improbability as indicated by the problem. This perspective posits that the theistic model, when appropriately understood, is intrinsically simple. My reply in the article was that the kind of resources used to show that theism is intrinsically simple (e.g God's being omnipotent, God's being perfect) cannot be used to explain God's psycho-divine harmony (PDH), as God would, in the first place, need to have very complex PDH in order to have those properties. In response, Adelstein does nothing to dispute my argument in the original post that PDH must be explanatorily prior to God's perfection (and other divine attributes). Instead, Adelstein claims it can be argued that while the metaphysically thin property of perfection obtains only in virtue of the presence of more complex sets of properties which are jointly intrinsically very unlikely, it can still explain them. I can't make sense of this response. In my understanding of the nomenclature, relations of explanatory dependence are asymmetric. It is either the case that God's PDH is explanatorily prior, or explanatorily posterior to God's perfection, it cannot be both, as that would be a vicious circle. Adelstein gives an analogy:

> "An apple pie may constitutively depend on the atoms arranging it, but the reason why they are arranged as they are may depend on features about the whole."

However, this only further muddies the waters for me. If we accept that an apple pie is dependent on its constituent atoms, then I have no idea what it is to further add that the atoms, in turn, constitutively depend on the pie as a whole. It's opaque what the relation of 'constitutive dependence' even picks out such that the statement is a sensible one (Again, as I understand it, it is an asymmetric relation) and Adelstein doesn't disambiguate it. Perhaps instead, Adelstein is trying to convey that while an apple pie constitutively depends on its atoms, the pie's functional or teleological features can explain the atom's arrangement. Leaving aside that this type of explanation strikes me as dubious, even if we were to accept it, the analogy's relevance to my main discussion remains unclear. Afterall, if 'perfection' is supposedly a simple, metaphysically thin property, as proponents of this rejoinder would insist, how can it have functional or teleological features rich enough to explain PDH?

Adelstein next responds to my response to the counter-objection to the revenge problem, which states that the kinds of laws needed to explain psycho-divine harmony are very simple. I argue that they aren't very simple, and appear to be quite complex, or at least, even if God's actual distribution of psycho-divine laws are simpler than other a priori epistemically possible distributions, there are so many other a priori epistemically possible distributions of psycho-divine laws each of which get some share of the probability space, that the prior of the theistic hypothesis is probably still astronomically low. Adelstein's first reply is that 'the criteria I give is clearly unworkable' because it implies that omnipotence is equally as improbable as omnipotence minus the ability to perform a random action. It ought not be of any surprise to the reader that Adelstein doesn't give any argument for why we should believe that omnipotence is a priori more probable than omnipotence 'minus one', let alone that we should think that the particular distribution of laws governing God's powers and mental life is significantly more likely than any other particular distribution of laws. As such, I'm well within my rights to simply point out that the claim is unmotivated and leave it there. However, as it turns out, in fact the point I made doesn't turn on God's laws being just as probable, or not greatly more probable than any given epistemically possible distribution of laws. God's laws can be a million times more probable than any given distribution of laws, but, once again, there are just so many, likely infinitely many, epistemically possible, disharmonious ways the psycho-divine laws conceivably could have been (e.g, God tries to will a frog into existence, but he causes a set of dominoes to fall over instead etc. etc.) that (by parity of reasoning for the improbability of psycho-physical harmony) so long as God's psycho-divine laws being harmonious is not maximally or close to maximally more a priori probable than other conceivable distributions, that God's laws are harmonious should be a very surprising fact and astronomically improbable indeed.

Adelstein also claims that since psychophysical harmony is just so shocking and improbable on naturalism, and we shouldn't be too confident about the prospects of the revenge objection, psychophysical harmony is still evidence for theism. What this misses is that the point of the revenge objection is not to dispute that we should update in favor of the theistic hypothesis when taking into account the data of psychophysical harmony. The point is that in the process of Bayesian updating, taking the priors into account is crucial. It doesn't matter if the data is a googolplex times more likely on H1 than H2, if H1 has a correspondingly lower prior, then the posterior may not come out overall in favor of H1 when we update. As an example—suppose Fred wins the lottery, clearly, that's evidence for the hypothesis that aliens rigged the lottery numbers in Fred's favor. It's incredibly unlikely that Fred ends up winning the lottery by chance, and not unlikely that Fred wins the lottery on the assumption that aliens rigged it, but the hypothesis that aliens rigged it has such a low prior that we still shouldn't believe the 'aliens rigged it' hypothesis even after conditionalizing on the data. The idea behind the revenge problem is that we are in a situation analogous to Fred's. Psychophysical harmony, it can be granted, is evidence for theism in the sense that the probability of theism raises once we conditionalize on it, but evidence is cheap, and the prior for theism is just so low that it's at the very least not clear that we should be confident that the theistic hypothesis has an overall non-negligible probability after taking the data into account.

Finally, Adelstein asserts that since there are at least somewhat plausible views on intrinsic probability on which theism would be simple, we should give some credence to such views, and factor that into our probabilistic calculations. As tempting as a response like this may be, it doesn't work. The reason this doesn't work is that the view we use to calculate intrinsic probability is the very resource that informs how we determine the prior probability of the hypothesis under discussion in the first place, which, as stressed, is a crucial step in Bayesian updating. I've nowhere claimed that my view on assigning the priors is maximally plausible, but we have to assign priors somehow otherwise our posterior probability will be inscrutable and thus whether we should adjust our beliefs in favor of theism given psychophysical harmony will be indeterminate. Plausibly, whatever method we use to show psychophysical harmony has an astronomically low prior will justify us in assigning an astronomically low prior to theism in light of revenge considerations. What's sauce for the goose is sauce for the gander. I must conclude that Adelstein's response to my section on the revenge argument is a failure.

Sources Cited

Adelstein, M. (2023). *For Theism: Part 1 - by Bentham's Bulldog*. Bentham's Bulldog. https://benthams.substack.com/p/for-theism-part-1

Balog, Katalin. (1999). Conceivability, Possibility, and the Mind-Body Problem. *The Philosophical Review*, *108*(4), 497–528.

Bourget, D., & Chalmers, D. (2020). *The 2020 philpapers survey*. PhilPapers Survey 2020. https://survey2020.philpeople.org/

Chalmers, David (2009). The Two-Dimensional Argument Against Materialism. In Brian P.McLaughlin & Sven Walter (eds.), Oxford Handbook to the Philosophy of Mind. Oxford University Press.

Churchland, Paul M. (2007). The evolving fortunes of eliminative materialism. In Brian P. McLaughlin & Jonathan D. Cohen (eds.), Contemporary Debates in Philosophy of Mind. Blackwell.

Cutter, Brian & Crummett, Dustin (forthcoming). Psychophysical Harmony: A New Argument for Theism. Oxford Studies in Philosophy of Religion.

Kirk, Robert

(2005). Zombies and Consciousness. Oxford, GB: Oxford University Press UK.

(2023) “Zombies.” *The Stanford Encyclopedia of Philosophy* (Fall 2023 Edition). Edward N. Zalta & Uri Nodelman (eds.).

Putnam, Hilary (1996). “Rethinking Mathematical Necessity.” In James Conant (ed.), *Words and Life*.

Ross, L. N. (2021). Causal concepts in biology: How pathways differ from mechanisms and why it matters. *The British Journal for the Philosophy of Science*, *72*(1), 131–158.

Thomas, Nigel J. T. (1998). Zombie killer. In Stuart R. Hameroff, Alfred W. Kaszniak & A. C. Scott (eds.), Toward a Science of Consciousness Ii. MIT Press.