

The Mental Measurement Problem: The Epistemological Consequences of Conceptual Dualism

Abstract

Conceptual dualists sharply distinguish between thinking about the mind in a first-personal, how-it-feels way and a third-personal, how-it-works way. Consequently, they run into the mental measurement problem: if the objectively measurable functional features of minds are conceptually separate from experience, how could we ever get a grip on the actual overlap between experiential properties and functional properties? To solve this problem, the conceptual dualist might appeal to ‘inner observation’ of their own mental states. But the mental measurement problem just re-occurs in the first-personal context: how could you ever come to know that some inaccessible mental state contributes to your own overall phenomenology? The conceptual dualist might, instead, try to choose some background commitment to help guide a science of consciousness in advance of any observation. But the *rational* selection of any such background commitment is not consistent with conceptual dualism. And so conceptual dualism

1. Introduction: Sonar Surgery and Phenomenal Overflow

Sonar Surgery

Imagine that you’re a test subject being studied by a lab of scientists. These scientists are considering the hypothesis that you have bat-like sonar experiences. In the same way that you have an intimate, first-personal acquaintance with what it is like to see a mural or smell a rose or touch a sticky surface, you are—according to this hypothesis—also intimately acquainted with a richly-detailed, dynamic perceptual array of sonar sensations.

How would we test this hypothesis? In general, what sort of evidence do we require in support of judgments about what subjects do/don’t experience? And if, by your own lights, you *don’t* enjoy bat-sonar experiences, could any scientific evidence ever persuade you to believe otherwise?

To address these questions, it can help to start with a case that clearly *wouldn't* be persuasive and see what changes might make it more persuasive:

1. Scientists point to a live bat at the other end of their lab as the 'source' of your sonar experience.

Under (1), you'd no doubt be incredulous. Sure, there's *some* experience associated with the bat's brain, but it's not *your* experience. And if can't rule out the experience associated with this bat brain as *your* experience, when could you ever rule out *any* experience associated with any neural (or non-neural?) activity anywhere in the world? Presumably, we have good reason to limit the potential sources of experience for us to states or processes within our own bodies.

Now consider the following series of tweaks to the scenario:

2. They cut out the bat's brain (plus ears) and keep it alive in a vat.
3. They perform surgery to place the live bat brain under the skin of your shoulder, keeping the bat brain in an isolated, self-sustaining mini-vat.
4. They perform further surgery, allowing the bat brain to be sustained, biologically, by your body. The bat ears, formerly envatted, are stitched to the tops of your shoulders.
5. Instead of shoving the bat brain into your shoulder, they shove it into your skull: you have a bat brain 'living' inside your skull, but it's still functionally isolated from your brain.
6. They perform additional surgery that allows the bat brain to interact or interfere with, say, visual perception. (e.g. When a particular sonar-detectable pitch is played, your vision goes fuzzy.)
7. Instead of altering a different perceptual modality, the sonar-processing system in the bat brain is integrated into your brain in a way similar to how your visual cortex is integrated into your brain.

When considering if/how these tweaks might impact your experience, it doesn't seem to much matter whether the brain activity in question is occurring inside your skin/skull. After all, there are any number of complex processes happening inside your body right now—like the functioning of your liver—of which you aren't consciously aware. While these processes typically aren't *neural* processes, it's not as if you automatically become consciously aware of any neural activity as soon as it occurs inside your body. For instance, a pregnant woman doesn't consciously experience being inside her own womb, although her fetus (which is contained wholly inside her body) probably does.

Insofar as we typically *do* restrict hypotheses about what can and can't underlie our conscious experience to processes occurring within our own bodies, this seems to follow from a *functional* constraint on which neural processing can conceivably underlie experience. Such processes, seemingly, need to make a certain kind of causal contact with certain other neural processes in our brains.

The Ever-Present Specter of Phenomenal Overflow

For the purposes of this paper, I want to focus on the cases described in (5) and (6), cases which I take to be structurally similar—at least in certain respects, which I’ll go on to detail—to less fanciful real-world cases. In scientific studies of blindsight, split brains, inattentional neglect, and more, it’s common for researchers to claim that the best explanation of the data¹ is that the experimental subjects had some kind of experience that was nevertheless quarantined off from the rest of their mental architecture. And such claims about *phenomenal overflow* are often granted a sort of legitimacy that few, I expect, would grant similar hypotheses about cases (5) and (6). And if we *do* take phenomenal overflow hypotheses seriously, then it’s not clear why we don’t take seriously similar hypotheses about case (1). Should we really take seriously the hypothesis that our mind is filled with strange and unfamiliar experiences that we can’t in principle, notice?

What makes phenomenal overflow hypotheses possible is *conceptual dualism*, the view that thinking about the mind in a first-personal what-it-feels-like way ought be sharply distinguished from thinking about the mind in a third-personal how-it-works way.

Under conceptual dualism, we can think about mental states as ‘conscious’ in two distinct, separate ways. We can think of mental states as *phenomenally* (p-) conscious: as ‘experiency’, as feeling a particular way, as having qualitative character. Or we can think of mental states as *access* (a-) conscious: as available for arbitrary use within our cognitive economy, as uniquely well-positioned to directly influence one’s thoughts and behaviors.² Importantly, conceptual dualists will claim that we can’t simply reflect on these concepts to learn about how the phenomenal properties of minds do/don’t actually overlap with their functional properties; there are no a priori constraints on the phenomenality (or lack thereof) of any functional states or on the functionality of phenomenal states. As proof of this, conceptual dualists will point to the conceivability of functional duplicates which lack experience (philosophical zombies) and experiences that play no functional role (which, elsewhere, I’ve dubbed strange experiences).³

I argue elsewhere that conceptual dualism ought to be rejected on conceptual

¹See for instance, Block XXXX.

²There’s a good deal of disagreement over how to best characterize a-consciousness. For the purposes of this paper, such disagreements shouldn’t much matter. My arguments should apply to any conceptual dualist view that distinguishes between a functional, mechanistic understanding of mind and an ‘experiency’ understanding of mind.

³Such a conceptual distinction can be found, most prominently, in the work of Ned Block (who distinguishes between phenomenal-consciousness and access-consciousness) and David Chalmers (who distinguishes the third-personal notion of ‘awareness’ from the first-personal notion of ‘consciousness’). In my view, conceptual dualism, or something very much like it, plays a crucial role in shaping much of the philosophical discussion of the conscious mind—though how it does this and in what ways is, I think, can be difficult to tease out. What is clear: many of the arguments and positions taken in philosophy of mind rely, either implicitly or explicitly, on a commitment to conceptual dualism.

grounds⁴. The main aim of this paper is to draw out the unattractive epistemological consequences of conceptual dualism, to demonstrate how challenging it is for conceptual dualists to guard first-personal knowledge of our own mental life from the epistemological obstacles it faces in analogous third-person cases. The conceptual dualist has to think that it's *sorta hard* to gain knowledge of experiences and their functional characteristics because such things, according to their view, aren't revealed through a priori reflection. But the conceptual dualist must also think that it's *sorta easy* to gain knowledge of one's own experiences and one's own functional characteristics because, otherwise, we'd be alienated/isolated from our own experiences in a way that's similar to how we're alienated/isolated from the experiences of others. It's perfectly fine, of course, for the conceptual dualist to abandon any commitments to the special epistemic status conferred by first-personal acquaintance, but without independent positive motivation, such ad hoc maneuvers are bound to look unattractive and costly. Consequently, observers should be more willing than before to reject conceptual dualism.

2. The Methodological Puzzle, Laid Out

Since, under conceptual dualism, we can't learn anything about the overlap between p-consciousness and a-consciousness a priori, their pattern of co-instantiation will be discoverable only through empirical investigation. (Note that this *epistemological* stance is compatible with a wide range of *metaphysical* views⁵.) Let's adopt this scientific project as our own and see where it takes us.

Overlap Questions and Empirical Method

The target of our empirical inquiry is an *overlap question*: When, and under what conditions, do p-consciousness and a-consciousness overlap (or come apart)? Scientific investigations organized around overlap questions should be familiar. Consider:

⁴That argument goes roughly like this: While cases of a-consciousness without p-consciousness (philosophical zombies) have received a lot of attention, cases of p-consciousness without a-consciousness (which I've dubbed strange experiences) have received far less attention. It's not clear that we can clearly conceive of subjects having individual experiences (e.g. pains, itches, smells) or entire sensory modalities (e.g. vision, bat-sonar) that are completely inaccessible to them. Furthermore, it's not clear that we can conceive of a phenomenal subject who completely lacks a-consciousness—an inversion on the zombie case. These cases test the opposite direction of p-consciousness/a-consciousness entailment that zombies were meant to test, and their conceivability is, in my view, much less obvious.

⁵Metaphysical views like: p-consciousness and a-consciousness actually come apart (Ned Block's view), p-consciousness just is a-consciousness (a posteriori functionalism), p-consciousness and a-consciousness are metaphysically distinct but their overlap is guaranteed by some contingent natural law (Chalmers' naturalistic dualism).

- Are all Ravens black? (What is the extensional overlap of “is a Raven” and “is black”?)
- Are any non-aqueous planets inhabited by living organisms? (What is the extensional overlap of “is a planet inhabited by living organisms” and “is a non-aqueous planet”?)
- How many U.S. households pay for cable TV? (What is the extensional overlap of “is a household in the U.S.” and “is a household paying for cable TV”?)
- Are there any stable elements with 17 nucleons? (What is the extensional overlap of “is a stable element” and “is an element with 17 nucleons”?)

All such overlap questions share a similar logical form. We’ve got two features of the world, F and G, and we want to know their pattern of co-occurrence. We can proceed by going out and observing a bunch of instances of F and G under various conditions, in the hopes of collecting a representative sample, and, from there, we can inductively infer their overall overlap on the basis of the patterns we pick up on from observed instances.

A Measurement Problem?

But this sort of project can succeed only if we have some way of *independently* measuring F and G. If, for instance, we used a G-meter that tracked G-ness by way of tracking F-ness, we’d risk screening off any non-F Gs that we might have encountered. Observational data collected in this way, we think, couldn’t serve as a reliable ground for inductive inference—we wouldn’t be able to learn anything that we didn’t *already* know since we’d be able to predict, in advance, precisely what sort of F/G overlap we’d observe using this G-meter (i.e. a perfect overlap).

Our investigation into the p-consciousness/a-consciousness (P/A) overlap question seems plagued by just such a measurement problem. While it’s relatively easy to collect data on a-consciousness using standard scientific methods, it’s unclear whether any such objective, third-person method could serve as a reliable measure for p-consciousness, since p-consciousness is supposed to be conceptually distinct from any and all physical or functional features of minded organisms or systems. Of course, if we knew the relationship between p-consciousness and a-consciousness (or any other objectively observable feature of the environment) in advance of observation, we wouldn’t encounter such a problem. But the nature of this relationship is precisely what we’re now trying to sort out. So our investigation into P/A overlap seems to have gone round in a circle. How might we structure a science of consciousness to break out of this circle, to make the P/A overlap question a tractable problem? This is what Ned Block has called the *methodological puzzle*⁶ of consciousness research:

⁶I’ve adopted Ned Block’s terminology, here, but the same basic problem pops up—under a various headings, framed in slightly different ways—in both the philosophical and scientific discussions of the methodology of psychological science. These discussions, in my view, are less

“It does not seem that we could find any evidence that would decide one way or the other, because any evidence would inevitably derive from the reportability[/accessibility] of a phenomenally conscious state, and so it could not tell us about the phenomenal consciousness of a state which cannot be reported[/accessed].” — Block, p. 483, “Consciousness, Accessibility, and the Mesh. . .”

Real-World Challenges

In getting a science of consciousness off the ground, one can feel tempted to push aside ‘merely philosophical’ concerns and instead focus on more pressing ‘practical’ issues. After all, we typically don’t expect scientists to resolve Cartesian skepticism before making any progress on scientific projects that assume the existence of the external world. Why not take a similar approach here? A careful examination of actual experimental cases can help demonstrate the real-world challenges we run into as a result of the methodological puzzle.

[I still need to make this section more robust. But some basic empirical data. . .]

Visuo-spatial extinction: subjects can identify objects presented, without competition, to one eye, but fail to identify objects presented to left eye when different object is simultaneously presented to right eye. Importantly, when a face is presented to their left eye, the fusiform face area still lights up.

Either the subject has a face-experience but is unable to cognitively access experience and report on it. Or the subject has no face-experience in the first place.

Other Examples: blindsight, split brains, attentional neglect, memory bottlenecks (Sperling 1960)

The Central Dilemma

[DRAFT NOTE: See the appended “Three Different ‘Knowing Minds’ Problems” for a section I had initially planned on including somewhere around here. I now worry it’s distracting and introduces unnecessary complications.]

So here’s the central epistemological dilemma for the conceptual dualist: how do we square the epistemological difficulties of the methodological puzzle with the epistemological non-difficulty of our everyday P/A overlap judgments?

clear and less well-organized than they should be and, as a result, make it hard to appreciate the precise nature of the methodological challenge. So part of what I want to do in this paper is to lay out the issue in the clearer way, in a way that exposes the essential structure of the problem, in a way that can help us unify these various discussions.

Here, we can articulate two different hurdles that the conceptual dualist needs to overcome:

1. The low bar: how are we every justified in rejecting phenomenal overflow hypotheses?
2. The high bar: how are *any* beliefs about the P/A overlap judgments justified?

Our current epistemological troubles can arise only once we're committed to the conceptual independence of experience from the sorts of 'third-personal' data to which scientific observation is restricted. Notice that analytical functionalists—who take all mental talk to reduce, analytically, to functional talk—easily get over both the low bar and the high bar, since they claim that what we can observe objectively/third-personally is not, conceptually, at a remove from experience. Furthermore, if there are the sorts of a priori functional constraints on experience for which I argue elsewhere, we easily get over the low bar, since evidence for a deficit in access would, straightforwardly, be evidence for a deficit in experience. (There's a separate story to tell about how a low bar solution can be turned into a high bar solution, but I'll leave that story for another time. But for present purposes, I just want to note how, in rejecting a clean conceptual break between p-consciousness and a-consciousness, we seem to immediately gain a sort of traction on the P/A overlap question that currently seems to elude the conceptual dualist.)

Conceptual dualism clears neither the high bar nor the low bar. It lacks the resources to establish a posteriori what it maintains can't be established a priori. The methodological puzzle is not some hurdle the conceptual dualist must overcome, but rather a direct expression of their core commitment. The view removes the sorts of conceptual hooks required for getting a empirical grip on the P/A overlap question.

But, presumably, the conceptual dualist won't want to just 'give in' to the methodological puzzle like this and affirmatively embrace a skepticism about a wide range of everyday judgments. So how can the conceptual dualist avoid this *prima facie* unattractive consequence? In the next two sections, I'll take a look at the two most promising strategies.

First, we might appeal to 'inner observation' of first-personal states. That is, we might try to use a regularity observed in our own case—between our own p-conscious states and a-conscious states—to support a broadly-applicable principle for guiding a third-personal science of consciousness.

Second, we might try to challenge the conception of science that seems to be motivating our present concerns. Under this approach, the methodological puzzle can be seen as an instance of a more general worry about the ultimate groundlessness of scientific methods. But once it's recognized that empirical investigation, essentially, requires *some* set of background assumptions or other, the idea that such background assumptions stand in need of prior justification can start to look silly. So unless we want to reject the very possibility of scientific

progress, it should be harmless to make some assumptions that get us around the methodological puzzle.

As I'll demonstrate, neither of these strategies will make conceptual dualism any more attractive than the already-unattractive bullet-biting strategy.

3. The Inner Observation Strategy

The Appeal of First-Person Data

To overcome the challenge of the methodological puzzle, conceptual dualists might appeal to 'inner observation' of first-personal states. That is, we might try to use a regularity observed in our own case—between our own p-conscious states and a-conscious states—to support a broadly-applicable principle for guiding a third-personal science of consciousness.

David Chalmers has suggested such an approach:

“Still we have a good idea from the first-person case about states in which we have experiences and states in which we do not, and an analysis of what is going on in these cases usually allows us to characterize those states in functional terms. So reflection on the relationship between experience and function in familiar cases gives us considerable leverage.” (*The Conscious Mind*, p. 226)

Under Chalmers' approach, we can reliably fill in the data about p-consciousness in our own case and pair it with standard third-person data about ourselves to better understand how the two mesh. And this might give us an unobstructed view of the overall P/A overlap map—or at least a particular region of it. (We then might sketch in the rest of the map based on what we observe in this region of the map.)

Here, it's important to be careful in characterizing what exactly makes first-person data *first-personal*. We're not, here, distinguishing between data *about* others and data *about* ourselves. For we can also collect third-person data about ourselves—that's, in fact, a crucial component of this strategy. Only because we can collect both third-person and first-person data about the same subject can we calibrate our third-person methods to track inner experience. So what's, instead, distinctive about first-person data is either the first-person *method* of data-collection (some sort of introspective method), or the first-person *format* (or conceptual content) of the data collected (the uniquely first-personal character of the p-consciousness concept with which this data is 'encoded'), or both.

This sort of strategy has a two-step structure:

1. Observe the pattern of P/A overlap in our own case.

2. Take this P/A overlap data as a guide for interpreting third-person data in cases for which we don't have first-person data.

Criticism of this approach (especially when applied to both the problem of other minds and the inverted spectrum problem⁷) often focuses on the second step of this procedure, the step—so it's claimed—that is founded on the unearned assumption that one's own mind is similar to other minds in the ways that are in question. It's thought that such an assumption is dubious for the same reasons that any inductive generalization based on a limited sample is dubious.

While I think such criticisms ought to be taken seriously, I think they misidentify the source of the problem, at least when this approach is applied to address the methodological puzzle. I want to, instead, focus my criticism on step 1: How does one *ever* establish the pattern of P/A overlap in one's own case?

The Broad Challenge of the Methodological Puzzle

Discussion of the methodological puzzle is often restricted to a fairly limited range of 'problem cases' involving subjects who have deficits with respect to 'normal' a-consciousness (e.g. blindsight, inattentional blindness, split brains, etc)⁸. But if the basic problem has to do with the missing a priori connection between a-consciousness and p-consciousness, it's hard to see what justifies limiting the discussion in this way. The reasoning that illuminates our poor epistemological situation in problem cases seems to be equally applicable in 'everyday' scenarios:

A puzzle for everyday negative experience judgements

- Everyone I know reports not having bat-sonar experiences. So everyone I know lacks access to bat-sonar states.
- But we can't take lack of access as evidence for lack of experience, according to the reasoning behind the methodological puzzle.
- So it looks like we can't know whether or not everyone/any particular person is or isn't having bat-sonar experiences.

A puzzle for everyday positive experience judgments

- Everyone I know reports having visual experiences. So everyone I know has access to visual states.

⁷I'll be specifically focused on how the 'inner observation' strategy might help in overcoming the methodological puzzle, but it's worth paying attention to the general form of this method. I expect many readers will be more familiar with its application to the problem of other minds, but while the strategy is similar, the context of application leads to crucial differences. Wherever applied, this general approach always has a two-step structure. The first step is to observe something about your own situation, and the second step is to take this first-personal data as a guide for interpreting third-personal data in cases for which we don't have first-personal data.

⁸[Add citations, if not in main text.]

- But we can't take access-facts as a guide to experience-facts, according to the reasoning behind the methodological puzzle.
- So it looks like we don't know whether or not everyone/any particular person is or isn't having visual experiences.

We can run through this same pattern of reasoning given any particular experience judgment that we make. And so it looks like an endorsement of the methodological puzzle subsequently requires that we endorse a similar puzzle for everyday beliefs about experience.

(Here, we might appeal to non-problematic 'basic' assumptions that get us out of this puzzle in everyday scenarios. We might think, "Surely it's not problematic to assume that 'normal' people have 'normal' experiences." This may be right. But if it is right, the very same sort of assumptions could be brought to bear on the standard methodological puzzle cases. I'll discuss these sorts of assumptions in section 4.)

But doesn't this reasoning apply equally well in one's own case? How do I know that I'm not having bat-sonar experiences right now, at this very moment? And how would I know that I'm having visual experiences except by being a-conscious of them? So it looks like conceptual dualism might require that inner observation of our own experience is also called into question for precisely the same reasons it calls into question 'outer' observation of experience in problem cases.

In any case, what is it about our epistemological standing such that the P/A overlap judgements for our own mind are not subject to the kinds of skeptical challenges mentioned above? If we can't give a satisfying answer to such a question, the inner observation strategy can't possibly succeed.

The Appeal to Acquaintance

At this point, some will want to appeal to a special kind of knowledge—knowledge by acquaintance—which is to be contrasted against less reliable knowledge by perception/observation. Under this approach, it was a mistake for us to assimilate knowledge of our own mental states to ordinary perceptual knowledge of the external world. Perception, it's thought, requires a sort of two-step process whereby we interpret the data coming in via perception to reach conclusions about the external world. When we're directly acquainted with a mental state, however, there's no such two-step process: we make direct epistemic contact with the object of our judgments. As a result, such judgments are immediately justified in ways that other sorts of judgments are only mediately justified.

This sort of approach can be traced back to Bertrand Russell, who first introduced the notion of knowledge by acquaintance. David Chalmers, at times⁹, adopts

⁹[Need to fill this out, but the basic idea: In short, Chalmers' discussions of bridging principles and the principle of structural coherence are inconsistent with one another. To address the paradox of phenomenal judgment, he appeals to an a priori condition that

this Russellian line:

“I think the answer to this is clear: it is having the experiences that justifies the beliefs. For example, the very fact that I have a red experience now provides justification for my belief that I am having a red experience. . . . there is no way to construct a skeptical scenario in which I am in a qualitatively equivalent epistemic position, but in which my experiences are radically different. My experiences are part of my epistemic situation, and simply having them gives me evidence for some of my beliefs. . . . To have an experience is automatically to stand in some sort of intimate epistemic relation to the experience—a relation we might call ‘acquaintance’ ” (196)

The major issue with this approach: in the context of the methodological puzzle, this response doesn’t bear directly on the issue at hand.

Acquaintance can help us settle the question of P/A overlap only if acquaintance and access are constitutively or conceptually related. If acquaintance is just a synonym for ‘access’ or if it entails access, the question just becomes: Is it *a priori* necessary that subjects are acquainted with their experiences? If yes, then conceptual dualism is false. If no, the challenge remains. If, instead, acquaintance and access are distinct from one another, it’s hard to see how an appeal to acquaintance helps us address the P/A overlap question in our own case.

Just to be clear, the appeal to acquaintance may, in fact, be able to provide us with a certain set of justified P/A overlap judgments in our own case—judgments regarding p-conscious states that are also a-conscious. But such data will be just as useful as the data collected by a G-meter that can’t detect non-F Gs. That is, these judgments can’t possibly be used to gain traction on the *overall* P/A overlap in one’s own case. For such judgments can’t help us figure out whether it’s reasonable to believe that we have bat-sonar experiences. So the central dilemma remains.

The Appeal to Rationality

A related objection goes like this: It’s not exactly an *a priori* truth that subjects are directly acquainted with their own experiences. Instead, subjects can’t *rationality* reject that they’re acquainted with their own experiences. If there *were* a barrier between oneself and one’s experiences—the primary source of rational belief—one’s own rationality would be undermined.

guarantees our acquaintance with our own experiences. In proposing bridging principles to help get around the methodological puzzle, he seems to think that P/A overlap commitments are pre-theoretical and can’t be chosen on the basis of any data we collect. And to support his principle of structural coherence, he seems to think that we can directly observe the correlation between awareness (a-consciousness) and experience (p-consciousness) in our own case.]

On the one hand, this appeal to rationality seems to weaken the position currently under consideration, because it seems to admit that one doesn't really *know* that one is acquainted with one's experiences. Rather, it's just that it's safe to *assume* an acquaintance with one's experiences. And this, perhaps, risks consequentially undermining the justificational status of the judgments in question.

But, the appeal to rationality may be intended to have more force. If we're not rational, there's no additional harm in assuming we're rational—this assumption can't change our already-compromised epistemic standing. And if we *are* rational, how can an assumption of rationality be anything other than harmless? But, here, we again run into the risk of using a priori reasoning to ground the conceptual dualist's P/A overlap judgments.

However, none of this addresses what I take to be the most important reason to reject the appeal to rationality here: it's hard to see how a failure of acquaintance (or access) would result in a lack of rationality. One is rationally criticizable only if one fails to rationally integrate all the evidence one has access to. But if you're not a-conscious of some experience, it looks like crucial information about your experience is being withheld from your rational consideration. And it's not irrational to get experience judgments wrong if the experiences in question aren't available to you. Consequently, failures of access don't seem to threaten one's standing as a rational agent.

The Appeal to Externalist Justification

Both of the above objections can be understood along 'externalist' lines: regardless of your own 'inner awareness' of such facts, it's the bare *fact* that you're rational/acquainted with your experiences that gives your P/A overlap judgments superior justificational standing.

For this to work, one would have to find independent support for the kind externalist epistemology that makes this objection possible. But putting that aside, I think we already have good reason to reject this particular instance of externalist analysis since it's inconsistent with the present inner observation strategy for justifying P/A overlap judgments. The original idea we were pursuing was that the way in which P/A overlap judgments were justified in our own case was somehow special when compared against other P/A overlap judgments. But the sort of externalism now under consideration threatens to undermine this distinction. If the bare fact that one is acquainted with one's own experiences justifies the corresponding P/A overlap judgments, wouldn't the bare fact that others are acquainted with their own experiences justify the relevant judgments in the same way?

Presumably, the intuition we were following was that there's some difference in the way these judgments are justified, a difference that's transparently available to the subject making the judgments. While the judgments about other minds might *happen* to be true, judgments about our own minds are on firmer footing—or

so the thought goes. But this appeal to externalist justification seems make the ultimate source of justification *opaque* in both the first-personal and third-personal cases. And if there are *ever* cases where justification isn't opaque, judgments about one's own mind seem to be prime candidates.

If we step back from the specifics of each of the above solutions, I hope it's clear that there's a general structural issue afflicting the conceptual dualist in the first-person case. They can try any number of ways to justify P/A overlap judgments in one's own case, but any such attempt will run into the following dilemma: If the conceptual dualist *does* discover some kind of justification for P/A overlap judgments, that kind of justification can't be a priori, since that would conflict with conceptual dualism. But if it's not a priori, then the corresponding P/A overlap judgments will be subject to skeptical challenges.

4. The Background Commitment Strategy

On behalf of the conceptual dualist, we might just want to accept the wide nature of the methodological puzzle and look for a broad solution to a broad challenge. In doing so, we're rejecting that first-person data plays any special role in resolving the methodological puzzle. We'll, instead, look for a common justificational story for *all* P/A overlap judgments.

Standard Empirical Background Commitments

The methodological puzzle might just be an instance of a more general worry about the ultimate groundlessness of scientific methodology. Cartesian skepticism about external world beliefs represents one version of this worry. Humean skepticism about belief in unobserved matters of fact is another. In general, you can always discover some defeasible background commitment(s) upon which any empirical project is founded, whether that be the commitment that one's perceptual faculties are generally reliable (in the case of Cartesian skepticism) or the commitment that unobserved aspects of the world will resemble observed aspects of the world (in the case of Humean skepticism).

But to worry about the groundlessness of such commitments, many think, is misguided. We typically don't expect science to proceed in the manner of deductive proof. Instead, scientific inquiry can proceed only after we've made a certain number of pre-theoretical assumptions. This gets at a distinctive feature of science: its conclusions never enjoy the kind of certainty that we attach to the conclusions of deductive proofs. But this is no defect—that's just how science works.

So once it's recognized that empirical investigation, essentially, requires *some* set of background commitments, the idea that such background commitments stand in need of prior justification can start to look silly—it just expresses a basic unwillingness to accede to the legitimacy of any and all scientific inquiry. Unless we want to reject the very possibility of scientific progress, it should be harmless to assume some background commitments that get us around the methodological puzzle and get a science of consciousness off the ground.

The Revisability Constraint

What's distinctive about the commitments that ground empirical inquiry is that they, themselves, are recognized as contingently true. Consequently, we can specify, for any given commitment, some series of observations that would overturn or otherwise threaten that commitment. For instance:

- *For a commitment to the reliability of perception:* observing your visual world pixelate, and give way to the visual experience of an evil scientist's lab.
- *For a commitment to the uniformity of nature:* observing your pen float up to the ceiling after letting it go.

Perhaps a better example are the everyday cases in which we suspend such commitments in particular cases:

- Belief in External World: You get lost in VR goggles and have to remind yourself that you're in a video game.
- Belief in Causal Laws: Some lab measures a particle traveling at a speed faster than light, but they then discover the source of instrument malfunction.
- Belief in Other Minds: You call up a customer service line and realize only after a while that the verbal responses you're getting are actually canned recordings/a computer.

In these cases, an inconsistency within the set of all observations—or between between what's been observed and the assumptions you'd made—needs to be reconciled, and that's what triggers revision. Importantly, there isn't an appeal to some independent authoritative resource that stands outside of empirical observation which forces this revision. Rather, such updates to your empirical framework result of some kind of internal tension.

The sorts of assumptions or commitments that underlie empirical investigation must meet a pretty minimal constraint: they must, in principle at least, be open to update and revision in the face of 'unexpected' observational data. More precisely, it must be conceivable that empirical evidence can come into rational tension with the background commitments that had been guiding the empirical investigation.

When we make assumptions about reality, we typically think that reality can push back, that the world can provide us feedback on our theorizing. To presume otherwise would, seemingly, challenge the overall scientific project. It might not be that such opportunities for revision always—or even typically—uniquely pick out some particularly problematic assumption as the source of error. But no empirical assumption can be presumed safe from reality’s feedback—they’re *all* subject to revision. Indeed, some may even think that such assumptions can accrue abductive support as a result of the continued success of an empirical project—if reality *could* push back but doesn’t, that suggests that one made the right assumptions.

The Immunity of P/A Overlap Commitments

It strikes me that any commitment to P/A overlap could not be revisable in this way. In particular, the gathering of non-confirming evidence seems impossible here where it was possible in the other cases. Background commitments generate predictions, and we get feedback on background commitments when the predictions they generate do or don’t pan out. But the predictions generated by P/A overlap commitments will, necessarily, pan out.

Is it even *conceivable* that you observe the coming apart of p-consciousness and a-consciousness in the way it’s conceivable that you observe the suspension of natural laws? It looks like p-conscious states, in so far as you ‘encounter’ and ‘observe’ them, have to be a-conscious—there was no real possibility of observing otherwise. Either you have direct access to them yourself or you can indirectly observe the effects of a-conscious states in others. Importantly, what’s inconceivable here is not the *having* of, say, inaccessible bat-sonar experiences. Rather, what’s inconceivable is that such inaccessible states could ever make a rational impact on your epistemology.

For instance, consider the case that p-consciousness and a-consciousness are, in fact, perfectly correlated, but you start from the assumption that they can come apart. What sort of evidence could *possibly* support the revision of your mistaken assumption?

By your own hypothesis, you might sometimes have phenomenal states that you don’t access. But then when you go on to never encounter non-a-conscious p-conscious states (as the actual pattern would require), things won’t turn out any differently than your hypothesis predicted. For your hypothesis, itself, postulated p-conscious states that were (both third-personally and first-personally) unobservable. A comparison to similar hypotheses that postulate things which cannot, in principle, impact observation (e.g. The hypothesis that there’s an invisible fairy always floating behind my head) should tip us off to something fishy going on here.

Now consider the other direction: p-consciousness and a-consciousness can, in fact, come apart, but you start from the assumption that they always go together.

In order to bring your commitments in alignment with reality, you'd need to pick up evidence of some non-p-conscious state of a-consciousness, or vice versa. But it's unclear how, say, your own p-conscious bat-sonar experience could possibly make an impact on your epistemology if you it weren't already a-conscious.

To sum up, it appears that any assumption about the overlap between p-consciousness and a-consciousness 'fits the data' just as well as any other, *no matter what the data is*. This is because every assumption predicts the same thing: you'll never observe a non-a-conscious p-conscious state. So while such states may conceivably *exist* under conceptual dualism, it looks like the *observation* of such states is inconceivable.

Just to be clear, I'm not claiming that P/A overlap commitments are distinctive because one can always find *some* way holding onto them in the face of new observational data. Rather, they're distinctive because they seem *completely immune* to rational revision of one's background framework. Such commitments don't seem to be tangled up with the web of commitments and beliefs that typify empirical inquiry. But now it looks like P/A overlap commitments are answerable neither to empirical data nor to a priori reflection. And that, I think, is unpalatable.

Instead, our 'assumptions' about P/A overlap look similar to 'assumptions' about systematic mathematical illusions. We can always assume that our past or present calculations were, for some reason, off, and that might undermine the certainty with which we treat judgments like " $2+2=4$ ". Similarly, we can always appeal to the failure of P/A overlap in our own case to undermine judgments about the boundaries of our experience (e.g. we can always assume that humans have non-a-conscious sonar experience, and that assumption will never be contradicted by the data). But Philosophers tend to think that skepticism regarding mathematical knowledge can't *really* be coherently entertained. And I think we should take a similar attitude towards our knowledge of P/A overlap.

5. Conclusion

If all this is correct, it puts the conceptual dualist in an awkward position. They've claimed that we can't just reflect on "is p-conscious" and "is a-conscious" to discover any constraints on P/A overlap. But a survey of the options for discovering the nature of P/A overlap empirically (under the assumption of conceptual dualism) turned up no good alternative for grounding beliefs tied up with P/A overlap—beliefs like "humans don't enjoy bat-sonar experiences". So unless we're prepared accept the complete groundlessness of such beliefs, we must abandon conceptual dualism instead.

ADDENDUM: Three Different “Knowing Minds” Problems

[DRAFT NOTE: I’ve had the following serve as the final subsection of Section 2, but I thought it introduced a lot of unnecessary terminology and distinctions, making the paper more cumbersome. Let me know if you think it’s useful enough to add back into the main text.]

Squint at the methodological puzzle long enough and it can start looking like the problem of other minds. After all, both seem to stem from the inadequacy of third-personal observation as means for gaining knowledge about first-personal mental states. Especially since access-facts aren’t especially difficult to establish, it looks like the methodological puzzle consists entirely in the difficulty we have in mapping out the pattern of instantiation for p-consciousness. But then the methodological puzzle is just reduced to the problem of figuring out where in the world, if at all, p-consciousness shows up—and isn’t this just equivalent to the problem of other minds?

While the methodological puzzle and the problem of other minds may *ultimately* collapse into one another under some positive conceptions of the mind, we can’t assume their equivalence up front. Initially, they each show up in quite distinct epistemological contexts.

We encounter the problem of other minds in our attempt to furnish the external world with minded beings. In such a context, the existence of our own mind is treated as a settled issue, and what’s in question is whether any other minds exist at all. On the other hand, we encounter the methodological puzzle when try to draw the boundaries of minds which we already take to exist, when try to determine which mental states of some phenomenally conscious subject contribute to their overall phenomenology and which mental states are phenomenologically inert. Here, the mind’s boundaries can be fuzzy not only in third-personal cases but first-personal cases as well—the methodological puzzle threatens knowledge of one’s own mind in a way that the problem of other minds seemingly doesn’t. (I’ll be saying more about this below.)

In short, the problem of other minds amounts to the question: How can I tell which things out there are minded? And the methodological puzzle amounts to the question: How can we determine the relationship between a given mind’s p-consciousness and a-consciousness? I’ll say more about this shortly, but I’ll also add a third related question, the sort of question behind the inverted spectrum problem: How can we reliably determine the precise phenomenal character of any p-conscious state? These are, logically, separable questions and need not be answered together. However, I worry that they sometimes get run together, and, as a result, the unique character of each question isn’t preserved.

To bring this into sharper focus, it’ll help to distinguish between three different classes of basic judgments towards which our epistemological worries are aimed.

Basic *existence judgments* have to do with which things in the world are subjects who enjoy p-conscious mental states. Any time we pick out some person, organism, or system as an ‘experiencer’, we’re making an existence judgment. *Overlap judgments* are more fine-grained, concerning whether individual states are just p-conscious, just a-conscious, neither, or both. We make overlap judgments when we pick out the boundaries of some subject’s phenomenology. And *character judgments* are the most fine-grained, having to do with the specific qualitative nature of p-conscious states. We make character judgments anytime we go beyond claiming “there’s *something* it’s like to be in this mental state” and start to fill in the details of ‘what it’s like’: “it’s like *this!*”. To get a feel for these sorts of judgments, consider the following examples:

Existence judgments

- “I am the subject of p-conscious states” - “Antonia is the subject of p-conscious states” - “Antonia’s zombie twin is not the subject of p-conscious states” - “My laptop is the subject of p-conscious states” - “This sesame seed is not the subject of p-conscious states.”

Overlap judgments

- “Nassim has a p-conscious, but not a-conscious, visual experience of the letter F.” - “Zombie Nassim is a-conscious of, but not p-conscious of, the letter G.” - “Antonia is both p-conscious and a-conscious of a pain in her lower back.” - “I am neither a-conscious nor p-conscious of the regulation of the functioning of my liver.” - “I have neither a-conscious nor p-conscious of bat-sonar experience.”

Character judgements

- “Nassim has a p-conscious visual experience as of a *red* bell pepper.” - “Antonia has a p-conscious visual experience as of a *green* bell pepper.” - “Antonia and Nassim have the same kind of color experience when they visually experience the bell pepper.” - “Antonia has the same p-conscious color experience of the bell pepper today as she did yesterday.”

I’ve used uncommonly rigid, non-everyday linguistic constructions in these examples so that each class can be more easily picked out and to avoid confusion or misunderstanding. But I take these classes to mark something fairly intuitive and familiar, something pretty close to the surface of everyday thought and talk about the mind. So we shouldn’t need any sophisticated theory in hand in order to recognize such distinctions. Rather, such distinctions should serve as a constraint on any good account of our concept of mind.

The sorts of epistemological challenges we’ve been discussing can be helpfully understood as centering around how such judgments can be justified. Such *justificational challenges* can be either *narrow* or *wide*.

A narrow justificational challenge targets only certain subset of judgements that are thought to be particularly problematic, perhaps judgements made in a certain context or made about a particular kind of thing, leaving the rest of the class of judgments intact and unchallenged. Narrow justificational challenges have the form: “How can you reliably make X-judgments under such-and-such

conditions?”. (e.g. “How do we justify color judgments about objects in the dark?”)

A wide justificational challenge isn’t constrained by such conditions, calling into question the entire category of judgments being targeted. Wide justificational challenges have the form: “How can we *ever* be justified in making X-judgments?” (e.g. “How do we *ever* justify color judgments?”)

To take a wide justificational challenge seriously is to think there’s a *completely general* problem for the kind of judgments in question—it’s meant to threaten the very idea of making such judgments in a grounded way¹⁰. Rather than being a problem about how we can extrapolate from some limited set of judgments to some broader set, it’s a problem about how we’re able to make any such judgments *at all*. Because of this, wide justificational challenges seem to call for a conceptual solution: we need to identify the epistemological hooks for the concepts that feature in the judgments in question. Without locating such epistemological hooks, the application of such concepts won’t be properly constrained—or at least that’s the sort of concern that a wide justificational challenge is supposed to raise. Alternatively, narrow justificational challenges seem to be addressable through standard empirical methodology, since the judgments that are left fixed and unchallenged can potentially guide our concept application after gathering information on their similarity/dissimilarity to the problem cases.

Applying these distinctions to the three epistemological problems picked out earlier yields an interesting disanalogy: Whereas the problem of other minds and the inverted spectrum problem seem to take the form of *narrow* justificational challenges (concerning existence judgments and character judgments, respectively), the methodological puzzle seems to take the form of a wide justificational challenge (concerning overlap judgments).

The problem of other minds leaves at least one existence judgement unchallenged: “I am the subject of p-conscious states”. And the inverted spectrum problem, too, seems to keep fixed a number of character judgments, judgements like: “Antonia is having a *visual* experience, not a *pain* experience of the green pepper”¹¹, “I’m visually experiencing the bell pepper as *red*, not *green*”. The problem of other minds is thought to be so troubling because it arises *even if* we knew the world contained experiential subjects. And what makes the inverted spectrum problem so challenging is that it pops up in cases where the mostly-successful functional constraints seem to give out, when we have two or more candidate phenomenal states share a certain kind of abstract structure but nonetheless differ in character.

The challenge behind the methodological puzzle doesn’t seem to be similarly

¹⁰One’s *response* to the general challenge need not, itself, be completely general—one can answer with a particularist or pluralist solution.

¹¹Though I take Lewis’s example of mad pain as a challenge to keeping these sorts of judgments fixed, that should still leave first-personal character judgments fixed.

limited. The challenge emerges precisely because our overlap judgments seem to have no a priori constraints whatsoever. At least initially, any arbitrary overlap judgment (e.g. “Antonia has bat-sonar experiences that are not a-conscious.”) seems to be on a par with any other (e.g. “Antonia is neither a-conscious nor p-conscious of bat-sonar experience”). Our challenge, then, is to figure out what makes any of these judgments better-justified than all the other candidate judgments.

This way of framing things can also help us identify two sorts of mistakes that might derail our project.

First, we should try to avoid blurring the distinctions between existence, overlap, and character judgments. This risks obscuring the potentially unique difficulties brought out by the problem of other minds, the methodological puzzle, and the inverted spectrum problem. And we should also be careful not to slide back and forth between wide and narrow versions of the justificational challenge for overlap judgements. As I’ll be using the term, “the methodological puzzle” specifically picks out the wide version of this challenge (though it’ll also be useful to talk of the “wide version” and “narrow version” of the methodological puzzle when I want to draw an explicit contrast). But plenty of discussions seem to frame “the methodological puzzle” as a narrower challenge involving particularly difficult ‘problem’ cases, leaving uncontroversial ‘clear’ cases unchallenged. Putting aside labels, a discussion of the narrower challenge simply can’t address the concerns raised by the wider challenge until it’s made clear precisely which class of overlap judgments are immune—and *why* they’re immune—to the sort of justificational challenge in question. (I discuss in more detail in section xx, below.)¹²

¹²To say a little more about how I see these two mistakes coming into play: I think the inner observation strategy looks promising, at least in part, because we’re liable to confuse first-personal subject judgments for first-personal overlap judgments. We may have sufficient introspective evidence that we are, ourselves, conscious—though, I think even this is dubious under conceptual dualism—but that can’t serve as good evidence for judgments about the overlap of experience and access in our own case. The second mistake, the elision of wide and narrow versions of the methodological puzzle, also plays a role in obscuring the the problems of the inner observation strategy—by . But it can also mislead us into thinking the inductive enterprise [of the second strategy] is more stable than it actually is.)