THE LOGIC OF MIND-BODY IDENTIFICATION

BERNARD MOLYNEUX

The University of California, Davis

I explore an interesting but largely unappreciated logical difficulty in the attempt to identify phenomena that are *prima facie* different, namely that, due to problems involving regress, it may be arbitrarily difficult or impossible to explain away the differences, even for genuine identicals whose apparent differences are illusory. I show that the circumstances in which this occurs are approximated in the context of the problem of consciousness, and that this may explain why proposed identifications between mental and physical phenomena typically give rise to how-possibly questions.

I. Introduction

Attempts to identify the physical with the mental give rise to “how-possibly” questions. “How could a rush of pleasure possibly be nothing more than an electro-chemical flare within a piece of meat?” “How could an abstract functional state possibly be qualitatively horrible?” Those who would subsume the ontology of the mental within the physical must address these questions, either by answering them directly, or at least explaining why they erroneously arise.

Many contemporary physicalists take the second line. They hypothesize something about us and the way we conceive of our own conscious states that explains why we find these identifications puzzling. The dominant approach of this kind has been dubbed “the phenomenal concepts strategy” by Macdonald (2004) and Stoljar (2005). The idea is that when we think of our conscious brain states as we experience them “from within”, we use special “phenomenal concepts”, whereas when we think of them “from without”, we use a different, more ordinary kind of concept. These special concepts resist integration with our regular concepts in such a way that, when we try to identify something picked out by a phenomenal concept with something picked out with an ordinary concept, we experience puzzlement.

The details of these phenomenal concepts vary by account. On some they are special because their referents are identical to their mode of presentation (Loar

Contact: Bernard Molyneux <molyneux@ucdavis.edu>

1990); on others the referents are part of the concepts themselves (Balog 2012, Block 2007, Papineau 2002). Some assign them a special inferential, evidential, informational, recognitional or causal role, or count them as part of a special faculty (Aydede & Guzeldere 2005, Carruthers 2004, Hill 1997, Sturgeon 1994, Tye 2003); whereas others attribute to them special indexical or demonstrative features (Ismael 2007, Levin 2007, O'Dea 2002, Perry 2003). But regardless of the details it is generally these special features that account for our difficulties when we try to identify a conscious experience with some third person physical phenomenon.

I suspect that the problem of consciousness is probably, at root, a plurality of tangled difficulties, some of which may have to do with the peculiar demonstrative, indexical, recognitional, evidential or quotational qualities of our phenomenal concepts. However, I wish to isolate a different strand to the problem by focusing on how our sense of mystification may arise from peculiarities in the logic of identification, of which the problem of consciousness is a special case. Like those who appeal to special phenomenal concepts, I approach the problem by addressing ourselves as thinkers, asking why we encounter problems when theorizing about consciousness. However, the problems that I predict arise principally from logical considerations, not as the result of contingent psychological barriers. They are not restricted to creatures with our conceptual architecture, nor are they destined to arise only when thinking about consciousness. To get clear on all of this, this paper must be as much about the logic of identification as it is about consciousness. With this in mind, much of the discussion until section 4 will be about identification, and little will be said about consciousness.

The basic idea is that identification, in being governed by Leibniz's indiscernibility of identicals, has a regressive structure. To regard x and y as numerically identical, one must regard their properties as numerically identical; but to regard the properties as numerically identical one must regard *their* properties as numerically identical; and so on. On the face of it this could lead to problems, yet it generally does not. Our methods of identification—surveyed in section 2—generally lead us away from regress. Regress arises, as we'll see in section 3, when and only when certain conditions obtain. I show in section 4 how these conditions are approximated in the problem of consciousness, and how this might help explain the difficulty of that problem.¹ The remainder of the paper (section 5) is devoted to considering possible misunderstandings and objections.

1. I first suggest the potential for identity regress in (Molyneux 2011), and develop a regress problem for self-reflective robots in (Molyneux 2012). The present piece builds substantially on these ideas.

2. Explaining Away the Differences

Sometimes, faced with the problem of explaining the identity of a physical and an experiential phenomenon, researchers will simply deny that identities need explanation (see e.g., Hill 1984: 101.) There is no need to explain how a thing is identical to itself, they say, and it is unintelligible to ask for an explanation. This answer, however, has not succeeded in ridding us of the problem of consciousness. That may be because, for psychophysical identifications, the claim that identities need no explanation often appears wrongheaded (see e.g., Papineau 2002: 150, Levine 2004: 81–82). Is this because psychophysical identifications are some sort of exception? I don't think so. For I think that *nearly all* interesting identity claims require *some sort* of “explanation”. Not because identity needs to be reduced—identity, I can agree, is metaphysically primitive. And not because identity claims need to be justified. They most certainly do, but that's beside the point. What's required, I suggest, whenever you identify some *x* with some *y* from which it seems distinguishable, is for you to *explain away* the apparent differences.² And since the identifications we actually make—e.g., of a celestial body that appears in the morning with one that appears in the evening; of weight with mass times gravitational force; of light with electromagnetic radiation between 390 and 740 nm—rarely present themselves as the trivial identification of patent indiscernibles, such an explanation is nearly always required.

In this section I develop some simple tools for thinking about such explanations. These allow me, in section 3, to show some useful results. I use these results in section 4 to suggest a regressive account of why we find mind-body identification so puzzling.

2.1. Identification Problems

The ancients conceived of Hesperus and Phosphorus as distinct. At some point, someone—perhaps Pythagoras—suggested they were not. That person would have faced a problem: That Hesperus and Phosphorus appear to have different properties. Phosphorus seems to appear in the morning but not the evening, whereas Hesperus seems to appear in the evening but not the morning.

Any answer would have to contain, in addition to the declaration of identity, an explaining-away of the apparent differences. We can represent this as a list of corrections to the prior view, where those parts of the view that are not corrected

2. Explaining away might not be the kind of “explanation” philosophers of science usually have in mind. I nonetheless suspect it's the kind you need to respond to how-possibly questions. For the best you can do with such questions is explain away the appearance of impossibility. I also think it is the only kind of “explanation” one can intelligibly request for an identity claim. Still, if you do not like this use of the term, not a whole lot in what follows depends on it.

(and are not affected by what is) are imagined to stay the same. I call this a “correction list” or “solution”. Here, in an abbreviated form, is the solution the ancients favored, where ‘H’ represents Hesperus, ‘P’ Phosphorus and ‘M’ and ‘E’, respectively, the properties of appearing in the morning and the evening:

$$[H + M, P + E, H = P]$$

Each element in this list represents one correction. The rightmost element is the principal correction. It represents that Hesperus and Phosphorus are now identified. The other corrections, to the left, facilitate this. ‘H+M’ represents that *appears in the morning* comes to be newly *added* to Hesperus, i.e., attributed as a property distinct from those that (when considered *as Hesperus*) it was hitherto thought to have. I call a correction like this an ‘addition’. ‘P+E’ is also an addition, and represents that Phosphorus is newly thought to appear in the evening.

2.2. Subtraction

Lois has hitherto regarded Clark, but not Superman, as human. If we let ‘C’ abbreviate ‘Clark’, ‘S’ ‘Superman’ and ‘H’ ‘the property of being human’, Lois can identify Clark with Superman as follows:

$$[C - H; C = S]$$

‘-’ represents that the property to the right is *subtracted* from the bearer to the left. In other words, the solution above explains away the differences by denying that Clark is human after all.

We might have tried the same strategy with Hesperus and Phosphorus, subtracting E and M instead of adding M and E. But that would have meant denying that Hesperus appears in the evening and that Phosphorus appears in the morning, which is why this approach has not been favored. It serves to illustrate, however, that there are multiple ways to explain away the differences. And that one’s choice of solution depends on one’s knowledge of the world.

Where there is more than one difference to resolve they may each be treated differently. E.g., if Lois has hitherto regarded Clark, but not Superman, as both a human and a journalist, then (letting ‘J’ represent *being a journalist*) a good solution is:

$$[C - H, S + J, C = S]$$

In advancing this solution, Lois ceases to regard Clark as human, but comes to regard Superman as having a day job.

2.3. *Facilitatory Identifications*

To adapt a case from Aristotle, how can we identify the road (A) from Athens to Thebes with the road (T) from Thebes to Athens when one (U) slopes uphill whereas the other (D) slopes downhill? The following looks right:

$$[U = D, T = A]$$

The justification is that, quite generally, the property of sloping upwards from x to y and the property of sloping downwards from y to x are the same one property. There are not two distinct states of affairs here, even if there are two ways of describing the matter.

But note that this solution contains two identifications: The principal correction, i.e., the identification of the roads, and the *facilitatory identification* $U = D$. This facilitatory identification, like an addition or subtraction, explains away an apparent difference.³

2.4. *Higher Order Identifications*

Facilitatory identifications add complexity to our scheme. In some cases, to identify x and y , you may have to make a facilitatory identification between their properties. And to identify *those*, you may need to make a *further* facilitatory identification between *their* properties. This may go on indefinitely, pushing the required corrections up to arbitrarily high orders.

Most identifications do not involve such higher-order facilitatory identifications, and we do not often have to think about them. When it comes to the mind-body problem, however, theorists may feel the need to advance to the higher orders. For example, suppose that an experiential state E is regarded as having the property P of *being painful*, and suppose that the neural state N that we want to identify with E is not. Then:

1. If the physicalist is to avoid property dualism, she must not add P as a new property of N over and above those discoverable from the third person. And if she is to avoid eliminativism, she must not subtract P from E thereby denying that the experience is painful. Suppose, then, that she makes a first order facilitatory identification between P and some third person property N' of

3. However, the case offered is just a helpful illustration. In fact, I would deny that A and T are genuine discriminables in the sense of section 3.4. In fact, if this were a genuine example of historical identification, it would be one of those rare cases where one chooses to identify elements that do not seem different to begin with. But in fact it is a contrived philosopher's example. No-one ever really distinguished those roads.

- N. Then a problem is that P is regarded as having the property (U) of making the experience it modifies unpleasant, whereas no property of N (including N') is regarded this way. So:
2. If she is to be a good naturalist, she must not add U as a new property of N' over and above those discoverable from the third person. But neither does she want to subtract U from P, since that would mean denying that P makes E unpleasant. It seems better, then, to make a *second-order* facilitatory identification between U and some third person discoverable property N'' of N'. But then a problem is that U seems different from all third person discoverable properties, hence different from N''. If this is because it has different properties, it may take a third order identification to resolve these differences.

This is merely an example of where one might resort to higher-order facilitatory identifications. However, it also stands as a decent dress rehearsal for the deeper issues in sections 3 and 4, where it is shown that sequences of ever-higher-order identifications of the sort begun here are sometimes unable to ground.

Note something else that has happened in this section. We have gone from the problem of identifying particulars to the problem of identifying properties. Nothing changes, logically speaking, when we do this. Though it may be harder to think about the properties of properties, it remains the case that, if we are to identify *x* and *y* then, whether they are particulars or properties, we need to explain away their differences. And it remains the case that we can do that by either adding, subtracting or identifying the differences away.

3. Developed Notions

Most of the points thus far have been fairly obvious and, I hope, uncontroversial. In this section I want to introduce some newer, more technical notions, heading towards a general result that has implications for the problem of consciousness: That in certain cases of identification, it takes infinitely many steps to fully explain away the differences. And that if, as a consequence, the differences cannot be explained away by finite creatures like us then, instead, we must “quantify out” of the problem by insisting that there is *some* way of resolving all of the differences, without saying what it is.

3.1. *Full and Semi-Solutions*

Let me start by saying more about this notion of “quantifying out”. One way to think of this is as a fourth kind of correction, supplementing those of addition,

subtraction and facilitatory identification. For example, let $\lceil (\exists \varphi \psi) \rceil$ be shorthand for the claim that there exists a property of φ that is identical to ψ , and suppose we want to identify heat (H) with molecular motion (M) but we do not know how to treat the property (R) *radiates through empty space* that H but not M appears to have. If this is the only difference we have to account for, we might propose the following:

$$[(\exists MR), H = M]$$

‘ $(\exists MR)$ ’ says that *some* property of the molecular motion is identical to the property (R) of radiating through empty space. It does not say which. But that’s enough to permit the identification to proceed coherently. I call this kind of solution a *semi-solution*

Is this the same as *adding* R to M? No, since to add R to M is to continue to think of R as distinct from all the properties that M is hitherto held to have. But this semi-solution allows for R to be one of M’s previously known properties. In this way, $(\exists MR)$ is effectively a placeholder for *either* an addition *or* a facilitatory identification.

I use the term ‘semi-solution’ since, in not disclosing how exactly the differences are resolved, this solution does not provide us with all the information we would like. Semi-solutions may come in different forms, depending on how the question of interest is put off. One could, for example, include a maximally general element, using two quantifiers, that says that *each* property of H is identical to *some* property of M (and vice versa), but which offers no specifics. Rather than go through the various possible placeholders, I simply define a semi-solution as one that is not *full*, where a full solution is one that contains only additions, subtractions and identifications.

When might we offer a semi-solution? In two situations. First, when we are too rich—when there are lots of ways of adding, subtracting and identifying away the differences, and we do not know how to choose between them. Until we know which one is correct, we restrict ourselves to the quantificational claim that there is some way of explaining away the differences, going no further. Second, when we are too poor—when we have no plausible full solutions at all. If we have a powerful argument, perhaps from coincidence of time, place and causal role, that $x = y$, but no way to explain away the differences, we must tell ourselves that there is some way to do so and hope that, later, we think of one.

For an example of the latter kind of case, suppose once again that the physicalist is identifying (P) painfulness with some property (N) of a brain state. As before (i) she must deal with the fact that (U) *qualifies its bearer as unpleasant* is an apparent property of P but not N (ii) she refuses the dualist option of *adding* U to N as a new property over and above its third person properties and (iii) she chooses

not to *subtract* U from P and deny that the pain makes the experience unpleasant. She must therefore identify U with some third person property of N, except that none of N's properties look suitable. Hence she makes do with a semi-solution, assuring herself that U is identical to *some* property of the brain activity, without saying which.

There is, however, something odd about providing a semi-solution when one can think of no full solution. For it is effectively to insist that all *is* well while having no idea how all *could possibly be* well. This foreshadows how it is that how-possibly questions arise in the context of the problem of consciousness, and counts as a second dress-rehearsal for the problems to come.

3.2. Coherence

I stipulate that for a list to count as a genuine solution, it must be *coherent*, meaning that it does not identify any two entities without resolving their differences. In other words, if one's correction list contains $x = y$ where one has hitherto regarded x and y as having differences, then the list should also contain either a subtraction, addition or identification that explains those differences away, or a quantificational placeholder, of the sort discussed in the prior section, that removes the difference without providing the specifics. If, however, it has none of these things then it is not coherent and hence, the way I shall talk, not a solution at all.⁴

3.3. Preservative Solutions

Let a *Preservative* solution be one that contains no additions or subtractions. For example, suppose we wish to identify (L) light with a kind of electromagnetic radiation, which we will call 'E'. Suppose the only difference is that L possesses (C) *comes in many colors* whereas E possesses (W) *comes in many wavelengths*. Then the following preservative solution looks good:⁵

$$[C = W, L = E]$$

After pursuing this solution, we are now inclined to *say* that L has W, where we did not say such things before. But, in doing so, we are not attributing a *new* property to L; we are identifying W with one of L's *old* properties. Preservative

4. More precisely, let v be a view, s a list of corrections to that view and v_i the view that results from making these corrections. Then s is a coherent solution with respect to v only if for all x and y such that x and y are represented as discriminable in v , if $x = y \in s$ then x and y are not represented as discriminable in v_i .

5. It is not, because it does not resolve the differences between the properties. E.g., that W, but not C, has something to do with waves. But it serves well enough as an example.

solutions, then, are so-called because they are, in a certain sense, anti-revisionist. In containing no addition of new properties and no subtraction of old, they preserve our *de re* commitments about which entities have which properties.

We might have alternatively offered the following semi-solution:

$$[(\exists LW), (\exists EC), L = E]$$

This affirms that *W* is identical to some unspecified property of *L*, and that *C* is identical to some unspecified property of *E*. Despite being only a semi-solution, the list contains no additions or subtractions. Hence, by definition, it too is preservative, and will continue to count as such until the placeholder is replaced by an addition or subtraction.

3.4. Differentials and Discriminability

It's time to get more precise about the notion of a "difference". From here on, let entities count as *discriminable* if and only if they are distinguished by their differing properties. The properties that distinguish *x* and *y* will be referred to as their *differentials*.

Note that to say that *x* and *y* are discriminable is to say not only that they have distinct properties, but that they can be distinguished *by* those properties. What this requires, in addition to their properties being distinct, is that their properties themselves be discriminable. For unless the color or shape or size . . . of *x* is discriminable from the color or shape or size . . . of *y*, *x* and *y* themselves will be indistinguishable. (A precise non-circular definition is footnoted.⁶) Accordingly, for a property *x'* of *x* to be a differential with respect to its bearer and *y*, it must be discriminable from every property of the latter.

It might be thought that 'discriminability' is just another word for the more familiar notion of discernibility, where *x* and *y* are discernible if and only if they have distinct properties. It turns out that there are lots of interesting connections between the two notions, but I lack space to get into that here. Suffice it to say that the new notion does indeed reduce to the old if it is granted that, among proper-

6. Let *x* and *y* be "discriminable" iff they are not *in* discriminable. Let *x* and *y* be "indiscriminable" iff they are "*n*-indiscernible", for some natural *n*, where 'n-indiscernibility' is inductively defined:

(**n-Indiscernibility**) *Base Condition:* *x* and *y* are 0-indiscernible iff they are numerically identical. *Inductive Condition:* *x* and *y* are *n+1*-indiscernible iff each property of *x* is *n*-indiscernible from some property of *y*, and vice versa.

This grounds the notion of discriminability non-circularly in facts about numerical identity.

ties, there can be no distinct indiscernibles.⁷ Readers who are willing to grant that much may substitute ‘discernibility’ and ‘indiscernibility’ for ‘discriminability’ and ‘indiscriminability’ throughout what follows. Importantly, if it is granted, then where we said that if x and y are discriminable, their properties are discriminable, we can now say that if they are discernible, their properties are discernible. For if they are discernible they have distinct properties, and distinct properties, the reader is granting, must be discernible properties.

3.5. *Representing as Discriminable*

Though two elements cannot be discriminable without having differentials, two elements can be *represented* as discriminable without being *represented* as having differentials. A novel may inform us, for example, that Smith and Jones look quite different without telling us how. In a similar way, a viewpoint or theory may represent x and y as discriminable without providing the details.

In dealing with such cases, we will make reference to:

(Absence of Differentials) *If x and y are represented as discriminable, yet no specific differential is represented, then any correction list containing $x = y$ is a semi-solution.*

The motivation for this principle is pretty clear: If our view provides no specific differences to add, subtract or identify, then we cannot correct it by adding, subtracting or identifying specific differences. Since any solution containing $x = y$ must succeed in removing the differences between x and y (see 3.2), it must do so by some other means. Hence it must be a semi-solution.

3.6. *Principle (*)*

We saw in 2.4 that, in explaining away the differences between identificanda it was possible to embark on a series of facilitatory identifications spiraling up to the higher orders. The following limit to this process is almost obvious:

(*) *If the identification of x and y is the highest ordered identification in the solution s between elements that are (hitherto) regarded as discriminable,*

7. Otherwise, if there could be distinct but indiscernible properties—call them F_1 and F_2 —then x and y could have distinct properties but not be discriminable. For suppose they differ only in that x (alone) has F_1 and y (alone) has F_2 . Since $F_1 \neq F_2$ it follows that x and y have distinct properties, and so are discernible; yet they are not discriminable since F_1 and F_2 , being themselves indiscernible, are incapable of distinguishing them. Without a qualitative difference between F_1 and F_2 , their numerical difference is insufficient to distinguish x from y .

then the discriminability of x and y is not (in any part) explained away in s by facilitatory identification.

Intuitively, if any part of the discriminability of $x = y$ were explained away by facilitatory identification, then $x = y$ would not be the highest ordered identification between discriminables after all. The facilitatory identification would be, since it identifies the differentials of x and y .⁸

3.7. Unfinishability

If we define a *finite* solution as one that contains only finitely many members, we can motivate the principal result of this paper:

(Unfinishability) *For all x, y where x and y were (hitherto) regarded as discriminable, any finite, preservative solution in which x and y are identified is a semi-solution.*

To see why, suppose (arbitrary) x and y that are represented as discriminable. And suppose some finite preservative solution s containing $x = y$. Since s is finite, it cannot contain identifications at ever higher orders. It must therefore contain some element $x^*=y^*$ (even if it is just $x=y$ itself) that is the (joint) highest ordered identification in s between discriminables. There are two cases:

1. x^* and y^* are merely regarded as being *somehow* discriminable, without any specific differential being provided. It follows by *Absence of Differentials* that any coherent solution containing $x^*=y^*$ is a semi-solution.
2. Otherwise specific differentials are provided. s must explain these differentials away if it is to count as a solution at all. Since $x^*=y^*$ is the highest-ordered identification in s of discriminables, we know by (*) that the differentials are not explained away by identification. And since s is preservative it does not explain them away using additions or subtractions. Hence s must contain an element that is neither an addition, a subtraction, nor an identification. Hence s is a semi-solution.

8. It's important, in securing this result, that those differentials would be discriminable themselves, as the discussion in 3.4 assures us they would. For otherwise the higher-order identification would not count as being between discriminables, and (*) would not follow.

3.8. A Second Argument

A second argument better displays the regressive structure of the problem. First we show:

(Regress) *For all x and y that were (hitherto) regarded as discriminable, if a full, preservative solution s contains $x = y$, then the solution also contains $x' = y'$ where x' is some property of x , y' is some property of y , and x' and y' were (hitherto) regarded as discriminable.*

To see why, suppose the antecedent: That (arbitrary) discriminables x and y are identified in a full, preservative solution s . Since s is full, it must explain away the differences by some combination of (i) addition, (ii) subtraction and (iii) identification. But since s is preservative, options (i) and (ii) are excluded. Hence the solution contains at least one element of the form $x' = y'$ for properties x' of x and y' of y , where one or both is a differential with respect to x and y . But to be a differential with respect to x and y , x' must be discriminable from every property of y , including y' (from section 3.4.) The same goes, with the appropriate changes, if y' is a differential. Either way $x' = y'$ is an identity between discriminables, hence we get the consequent of *Regress*.

Regress tells us that for every identification between discriminables in s there is another one at the higher order. It follows that if s contains at least one, it contains infinitely many. Since s was chosen as an arbitrary full, preservative solution in which discriminables are identified, it follows that all such solutions are infinite. Conversely, if a preservative solution identifying discriminables is *not* infinite then it must not be full. Thus we re-derive *Unfinishability*.

3.9. Remarks on Unfinishability

I stress the difference between the problem encoded in *Unfinishability* and the more obvious problem below:

“Of course you cannot identify elements in your ontology, consistent with Leibniz’s Law, if you think that they have different properties and you are unwilling to make further revisions.”

This problem is obvious only to the extent that you forget the option of identifying the elements’ properties. That option gives us the *prima facie* possibility of identifying elements by identifying—and hence preserving—the properties attributed to each. The discussion above, in contrast, allows for this option, but notes

that it can only be exercised at one order at the cost of a structurally identical problem at the next. This is what gives rise to *Unfinishability*.

One may fear that *Unfinishability* is too strong, because it may exclude full, finished identifications of the sort we make all the time. If so, then the arguments above overreach. I return in 5.4 to address this.

For now, let's maintain our central course and apply these results to the problem of consciousness. I suggest that the physicalist problem of consciousness falls out as a special case of *Unfinishability*. For physicalists, I argue, in attempting to identify experiential and physical phenomena without revising our notions of either, are pursuing the kind of solution that, according to *Unfinishability*, cannot be had.

4. The Problem of Consciousness

Given a few plausible assumptions, *Unfinishability* entails the impossibility of fully identifying physical and experiential entities and kinds. To the extent that these assumptions are plausible and difficult to reject, the analysis sheds light on why the problem of consciousness is so hard.

The plan of action, in this section, is to: (1) Discuss the principles that motivate physicalists towards preservative approaches. (2) Show, using *Unfinishability*, that their approaches must therefore terminate in semi-solutions. Finally, (3) show how how-possibly questions emerge when semi-solutions are given where no full solution is available.

4.1. *T*, and its First and Third Person Parts

Let *T* be our current best picture of the world. I assume that *T* breaks down solely into a phenomenological, first person part and an objective, third person part. I take the former to include everything we have learned about consciousness through introspection, and the latter to include the whole of physical science, including everything we know about the microphysical facts and about what supervenes. I assume the physicalist wants to unify these two halves by identifying phenomena from the first person part with (supervenient) phenomena from the third person part.

There follow two restrictions on the inquiry into consciousness, *first person preservativity* and *third person preservativity*, each of which corresponds to one part of *T*.

4.2. First Person Preservativity

Goldman writes that “when a . . . concept clashes with our first-person concept of consciousness . . . we should conclude that it isn’t a consciousness concept at all” (Goldman 2002: 122). Chalmers insists that we must “take consciousness seriously” by refusing “. . . to redefine the phenomenon in need of explanation as something it is not” (Chalmers 1996: x). First person preservativity is an attempt to model this idea. Stated within the framework we have developed, it forbids us from making the mind-body problem easier by adding and subtracting properties from the first person part of *T*; i.e., by supposing that the types of phenomena we observe from the first person have properties that they seem to us to lack, or lack properties that they seem to us to have.

First person preservativity may seem rather strong. I will show how it can be weakened, without affecting the central point, in section 5.8. For now, simply note that first person preservativity follows, more or less, from commitments in the historical literature that are even stronger. I’m thinking of variants of the twin claims that conscious states are *self-intimating*—i.e., if one’s experience is *F*, it seems to one that it is *F*—and that our access to them is *incorrigible*—i.e., if an experience seems to one that it is *F*, it is. These claims are often justified by the observation that though other things can fail to appear as they really are, the *appearances themselves* cannot. To get first person preservativity, simply add that we ought not to revise our opinions about something—our own introspectible conscious lives—that we cannot be wrong about in the first place.⁹

However, you do not need to be attracted to such strong principles to be attracted to first person preservativity. For preservativity merely forbids us from adding and subtracting properties from the *types* of experiences we have, not the tokens. It does not ask us to deny that subjects in a change blindness experiment may sometimes be unaware of their alternating qualia (cf. Dennett 2005: 84); nor the possibility that dental patients may sometimes mistake a pressure for a pain (Goldman 2002: 126).¹⁰ All that it requires is that whether we can be mistaken about a particular experience on a particular day, we are not wrong in thinking that, *generally speaking*, we have pains; that there is something it is like to have them; that a sharp pain is different from a dull ache; that pains appear in our somatosensory field—i.e., we do not see or hear pains—and that, like offensive fla-

9. It is incorrigibility, not self-intimation, that guarantees we cannot be wrong. But if we assume that an experience cannot seem both *F* and not-*F*, then if experiences are self-intimating, they are incorrigible. For if it seems to one that one’s experience is *F* then it does not seem not-*F*. But, by self-intimation, it *would* if the experience were in fact not-*F*. Hence it is not. Hence it is *F*.

10. Nor need we insist that when a pain has some extrinsic relational property—e.g., *occurring on a Tuesday*—self-intimation makes it known. Though even the proponents of the strong principles would stop before making *that* claim.

vors and experiences of nausea, pains are unpleasant. Similarly, we are not wrong to say that the type *pleasure* has a typically distinct feel from the type *pain*, that the type of experience associated with red differs in value from that associated with blue, and differs in kind from that associated with a tickle.

First person preservativity is relativized, moreover, to our reasons for making a change. It does not require that properties must never be added or subtracted, *tout court*. It merely insists that they not be added or subtracted, without independent motivation, just to make the mind-body problem easier. One may allow, consistent with this restriction, for additions and subtractions to be made in the light of appropriate independent evidence.¹¹

The restriction, despite these caveats, may still be too strong for some. In the interests of an orderly discussion, though, I propose to stick with it for now. I'll return to the question of how we may weaken it, without altering the central point, in section 5.8.

4.3. Third Person Preservativity

Ney (2008) writes that “physicalism is best taken . . . as nothing more than an attitude, a commitment to form one’s ontology according to whatever the physics of the day says exists”. This attitude is expressed, e.g., by (Lycan 2003: 13), when he suggests that the philosopher of mind ought not to second guess the ontology provided us by physics. In this spirit, the third person preservativist holds that if empirical science says a property is there, the philosopher of mind has no business denying it; and if it finds no property, the philosopher of mind has no business insisting upon it. In the framework we have developed here, this amounts to a prohibition against adding or subtracting properties within the third person part of *T*.

The most obvious operation forbidden by third person preservativity is the addition of qualia. We may not identify conscious states with brain states by supposing, in the manner of the property dualist, that qualia are new, ontically additional properties of the latter. The prohibition, despite this, *does* allow us to ascribe qualia to brain states, but only by identifying each quale with one of the brain state’s third person properties. In our framework, that means making a facilitatory identification, not an addition.

Some may think that third person preservativity is too strong. For the moment, it works as a reasonable idealization. I consider plausible ways in which it could be weakened in section 5.8.

11. E.g., we may get evidence from clinical cases of people who swear they have pains (even agonizing pains) that do not bother them at all (see Damasio 2000: 74–75.) This may cause us to revise the idea that pains are essentially horrible. With thanks to Chris Hill.

4.4. Deriving a Problem of Consciousness

We might now suppose the following, where a ‘first person x ’ and a ‘third person y ’ are, respectively, elements from the first and third person parts of T :

(**) *A solution to the problem of identifying some first person x with some third person y is both first and third person preservative if and only if it is preservative tout court.*

(**) seems pretty obvious. If a solution is preservative (*tout court*) then it contains no additions or subtractions. *A fortiori*, it can contain none from the first and third person parts of T , so it must be first and third person preservative. If, conversely, a solution does not add or subtract from the first or third person parts of T then since (we are assuming) there are no other parts, it does not add or subtract from T at all. In that case, the solution is preservative, *tout court*, with respect to T .

The principle (**) permits us to replace ‘preservative (*tout court*)’ with ‘first and third person preservative’ in *Unfinishability* to get:

(H1) *For any (hitherto discriminable) first person x and third person y , any finite, first and third person preservative solution in which x and y are identified is a semi-solution.*

Assuming no infinite solutions (I return to this in section 6), $H1$ tells us that if a solution between hitherto discriminable first and third person entities is preservative, it is not full. I.e., that when it comes to identifications between such things, there is no way to fully explain away their differences.

But this is where things get interesting. For if there is no way to fully explain away the differences, then the physicalist won’t be able to correctly conceive of one. The best she can do is to offer a semi-solution. Yet in offering a semi-solution where she can think of no full solutions, the physicalist (a) claims that though the identificanda appear to differ, there is some coherent way to explain away the differences, yet (b) can conceive of no *specific* way of explaining away the differences (for if she could then she would have at least one full solution to offer). The result is a how-possibly question: *How could it possibly be the case that x and y are identical when there is no conceivable way to explain away their differences?*

For this reason, $H1$ gives us:

(Hardness) *For any (hitherto discriminable) first person x and third person y , any finite, first and third person preservative solution in which x and y are identified will be beset with how-possibly questions.*

Note that though *Hardness* was developed, by and large, from first principles, it predicts, to a reasonable approximation, the lay of the land in the investigation of consciousness. On the one hand, we have the contemporary physicalist who, wishing to take consciousness seriously, eschews the option, in Chalmers' (1996, x) words, of "redefin[ing] [consciousness] . . . as something it is not"; but who, in respecting the authority of science, also refuses to revise the world that it describes. To the extent that this makes the physicalist into a first and third person preservative, *Hardness* predicts, any solutions she offers will give rise to how-possibly questions.

On the other hand, *Hardness* tells us, the mainstream physicalist's problem is in searching for a *preservative* solution. If she just gives up on *that*, these problems go away. So it's not surprising, from this perspective, that we find theorists who have taken exactly that option. The property dualist, for example, succeeds in identifying mental states with physical states by *adding* the mysterious qualia as new properties, over and above those supplied by empirical science. The qualia eliminativist, meanwhile, *subtracts* the qualia from the mental state, removing the reason for regarding the two as distinct.

None of this is to say that, in being the approach that leads to the regress, the physicalist approach is incorrect. For it may in fact be that $a = b$, a 's property a' is identical to b 's property b' , their properties a'' and b'' are, in turn, identical, and so on to infinity. If so, then the *right* solution would be an infinite one that captures this, not one that (say) adds a' to b as a property over and above b' , nor one that (say) subtracts a' from a . But the question of whether an infinite solution is correct is distinct from the question of whether we can provide it.

5. Objections and Discussion

5.1. No Apparent Regress

One might object that the problem outlined here involves a regress, but one experiences no obvious regress when thinking about consciousness. Hence the problem outlined here, whatever else we might say about it, is not the problem of consciousness.

In answer, let us see how the regress might arise in the context of the problem of consciousness. Imagine a researcher tries to identify (say) pain with (say) lamina I neurons stimulating parieto-insular cortex. The problem is that they seem different, and they seem different because the pain has noxious properties that seem discriminable from any properties of the brain event. If he assures us merely that these noxious properties are identical to some unspecified properties of the brain event, he merely gives us a semi-solution, as *Hardness* predicts. If, on the other

hand, he plows ahead and identifies some noxious property (N) with (say) (L) the lamina I stimulation's tendency to cause activity in the anterior cingulate, the new problem is that N and L seem discriminable. If he is able to clearly articulate *how* they seem discriminable—i.e., by describing the properties that seem to distinguish them—he must now cross-identify *these* properties in turn. But these too will seem discriminable (or else they would never have distinguished N from L). Thus he has a new problem structurally identical to the first. Wherever the researcher stops, pending further empirical research, or wherever he can no longer articulate the differences, his legacy to us is a semi-solution.

This, of course, does not establish an infinite regress. But it does defend against the objection that there clearly isn't one.

5.2. *Mental and Physical States are Not Represented as Discriminable*

A distinction I have ignored is between representing x and y as discriminable and *not* representing them as *indiscriminable*. One's prior view may represent x as having x' but remain silent on whether or not y has it, neither representing it one way nor the other. Since this is possible, one might suppose that our mental states and brain states, rather than being represented, according to introspection, science, etc., as discriminable, are merely not represented as indiscriminable, and that the principles *Unfinishability* and *Hardness*, being ostensibly about entities that are positively represented as discriminable, therefore fail to apply.

In the end, though, this does not significantly alter the problem. For the purposes of identifying x and y , all that matters is that x is represented as having a property x' that distinguishes it from y . It does not matter whether y is represented as lacking it, or merely not represented as having it; either way, one cannot rationally identify x and y while ascribing x' to x alone. This means that, as before, x' must either be added to y , subtracted from x , or held identical with some known property y' of y . But when options of the first two kinds are prohibited, the third leads to regress.¹²

5.3. *Discriminability and Contrary Properties?*

Objection: "When x and y are regarded as discriminable, they are either correctly or incorrectly represented as having contrary properties, which we can call x' and y' . If they are correctly represented as contraries, there is no way to correctly identify x and y without subtracting at least one of them from its bearer. If, however,

12. The regress proceeds much as before, except it goes from an x and a y that are not represented as indiscriminable to an x' and y' that are not represented as indiscriminable, via the observation that if x' had been represented as indiscriminable from y' , it would never have distinguished x from y .

they are incorrectly represented as contraries, you can continue to assign x' to x and y' to y by merely revising your belief that they are incompatible. But it is unclear, in that case, that there is anything problematic here at all.”¹³

Suppose, taking the second horn, that they are not genuine contraries. Then, I agree, an option is to revise your belief to that effect. However, it's a mistake to think that now your work is done. For if that is the only change, you still represent x , but not y , as having x' and y , but not x , as having y' . Sure, you no longer regard them as *contrary* properties, but you still ascribe them unevenly. Since, as stressed in the prior section, one cannot rationally identify x and y while ascribing x' to x alone (ditto, etc., for y') you must either add x' to y , subtract it from x , or identify it with some property of y . But that just brings you to the beginning of the problem.¹⁴

5.4. Counterexamples to Unfinishability? (I)

Unfinishability is a general restriction on identification, not pertaining specifically to the mind-body problem. One might worry that, in being so general, it is likely to predict problems, not just for mind-body identities, but wherever we try to make an identification.¹⁵

Does *Unfinishability*, for example, predict problems with identifying Hesperus and Phosphorus? No. The (correct) solution in 2.1 contained additions, and hence was not preservative. *Unfinishability*, however, only prohibits finite, full, *preservative* solutions.

Similarly, *Unfinishability* predicts no problems in identifying water (W) with H₂O (or heat with molecular motion, or light with visible electromagnetic radiation.) To see why not, suppose that the chief obstacle to identification is that H₂O instantiates the property (M) *has a micro-particulate structure*, whereas water does not. Then an appropriate solution would be $[W + M, W = H_2O]$, i.e., we simply add M to our notion of water. But this is not preservative, and hence no counterexample to *Unfinishability*.

What's the difference with consciousness? Simply that, with consciousness, we are reluctant to deploy regress-stopping moves like addition and subtraction, since it requires us to naysay either the scientist's account of what there is or the individual's account of what it's like for her. In eschewing additions and subtractions, however, we must pursue a strategy of pure identification (or settle for a semi-solution). But the strategy of pure identification requires an infinity of operations.

13. Jonathan Simon made this objection to an earlier version of the paper.

14. Specifically, to the version of the problem discussed in section 5.2.

15. Matthias Scheutz and Alejandro Perez Carballo, among others, have expressed similar worries to me.

5.5. *Why Aren't Other Semi-Solutions Puzzling?*

It may be that, in identifying water with H_2O , I do not have much idea what gets added, what gets subtracted, etc. I just trust that the differences get explained away *somehow*, and leave the details to the chemists. But though I have to give a semi-solution I do not end up *mystified* about the identity of water and H_2O . So what's the difference?¹⁶

The answer is that I can finitely discharge any quantificational placeholder in the water/ H_2O semi-solution. E.g., when I leave a placeholder that covers for water's lacking M , I anticipate that, later, I may discharge it by either (i) subtracting M from H_2O or (ii) adding M to water. And there are other options.¹⁷ The fact that I do not yet know which option is right means I have how-*actually* questions to answer.

When addition and subtraction are prohibited, in contrast, attempts to finitely discharge a placeholder result in the reimposition of another placeholder. That's because, given the prohibition, a placeholder can only be discharged by identifying discriminable properties. Yet those properties, in being discriminable themselves, must also have their differences explained away, which, given the prohibition, requires that *their* discriminable properties be identified, and so on. Sooner or later we must abandon the regress. But that means leaving a placeholder in lieu of the next identification.

In leaving a placeholder I promise that this can all be worked out *somehow*. But I can conceive of no finite way in which it could. So where, with water and H_2O , I had how-actually questions to answer, with consciousness and physical states, I have how-possibly ones. This is why, according to the present account, the latter problem leaves me mystified.

5.6. *Counterexamples to Unfinishability? (II)*

Objection I: "Suppose I wish to identify Cicero (C) with Tully (T), but I represent them to have no specific properties (this is unlikely, but it does not matter.) Would the finite, full, preservative solution [$C = T$] work? If so, isn't this a counterexample to *Unfinishability*?"

No. Consider both cases: Either I do not regard Cicero and Tully as discriminable, or I regard them as having (unspecified) differences. In the first case, since *Unfinishability* only prohibits finite, full, preservative solutions *for things hitherto regarded as discriminable*, [$C = T$] is no counterexample. In the second case, [$C = T$] does not address the (unspecified) differences between C and T . Hence it is not

16. I thank an anonymous reviewer for this objection.

17. E.g., identifying M with some property P of the water then explaining away M and P 's differentials using addition or subtraction.

truly a solution (cf. 3.2) and hence not the kind of solution that *Unfinishability* prohibits.¹⁸

Objection II: “Suppose I represent Cicero and Tully to have, respectively, C' and T' but no other properties, and I do not represent any specific differences between C' and T' . Then isn't $[C' = T', C = T]$ a full, finite, preservative solution, and hence a counterexample?”

No, since we get essentially the same dilemma. Either I do not represent C' and T' as discriminable, or I represent unspecified differences. In the latter case $[C' = T', C = T]$ fails even to be a solution, since it does not resolve the (unspecified) differences. In the former, C' and T' , being themselves indiscriminable, cannot act as differentials for C and T . Which means that, lacking other differentials, C and T turn out not to be discriminable either.

Note the way in which the indiscriminability of C' and T' cascades down. This happens no matter how long the list, and no matter how much it branches. If, for example, C' has C_1'' and C_2'' whereas T' has T_1'' and T_2'' , then the list $[C_1'' = T_1'', C_2'' = T_2'', C' = T', C = T]$ can be a genuine solution only if C_1'' is not discriminable from T_1'' and C_2'' is not discriminable from T_2'' . But in that case (*ceteris paribus*) nothing distinguishes C' from T' and so nothing distinguishes C from T . Hence this is not a counterexample to *Unfinishability* either, since it does not contain an identification between discriminables. The same goes for any putative counterexample of this sort.¹⁹

5.7. An Appearance/Reality Distinction for Conscious States

Objection: “Why not deny first person preservativity? The idea that conscious introspection is infallible is a legacy of bad folk psychology. In fact, there is a coherent appearance/reality distinction for consciousness just as for anything else, and hence a coherent sense in which we may be wrong about our own conscious states.”²⁰

I make four points in response: (1) The present paper is principally diagnostic.

18. Recall from section 3.5 that, when dealing with unspecified differences, a quantificational placeholder is required. But then we have a semi-solution, which cannot be a counterexample to *Unfinishability*.

19. More rigorously, for any finite, genuine solution containing only identifications: *Base:* The highest order identifications must be between hitherto indiscriminables or else the solution would not be genuine. *Inductive:* If all identifications at higher orders are between hitherto indiscriminables then $x = y$ is between hitherto indiscriminables. (Because x and y cannot have had their prior differentials identified or revised away, since the solution contains only identifications and all higher order identifications are between hitherto indiscriminables (and indiscriminables cannot be differentials). Hence if x and y are indiscriminable now, they must have been indiscriminable all along.)

20. I thank Chris Hill for an extended objection of this sort that I do not have room to do full justice to here. Jonathan Simon makes a related objection.

Infallibilist intuitions about phenomenal consciousness have long informed the philosophy of mind, and still exert a strong pull. I want to illuminate the connections between these commitments, however vaguely held, and the problem of consciousness. I do not need to subscribe to them. (2) Having said that, rejecting first person preservativity is a tricky business. Firstly, there is the enduring question of how one can possibly be wrong about what it's like for oneself. Secondly, in allowing for mistaken representations of our own consciousness, we must be careful not to generate other problems. (To see a big one, ask yourself what it must be like to *really* have the conscious sensation of green but to misrepresent oneself, in the relevant way, as having the conscious sensation of red? See Dennett 1991: 132 and, especially, Neander 1998) (3) Thirdly, deviations from first person preservativity ought to be independently motivated, by which I mean that we should not abandon preservativity *in order to* avoid the regress since the regressive solution may be correct. Mental states may be identical with brain states, their properties identical to neurophysical properties, their properties identical in turn, etc. (4) And even if we do reject first (or third) person preservativity, that may not be enough. For it also matters *how* it is rejected. I argue next that, for most physicalists, it is unlikely to be rejected in the right way, and that, if it is not, the regressive problems will continue to arise.

5.8. Weakening Anti-Preservativity

My aim here is not to defend preservativity, but to use it to account for why how-possibly questions arise for physicalist proposals. The preservativity described in section 4, however, is stronger than what some physicalists may be willing to endorse. What explains the existence of a puzzling mind-body problem for *those* physicalists? I'll answer by showing how weaker variants may lead to the same problems.

How might preservatism be too strong? Well, a reasonable physicalist, notwithstanding her reverence for the ontology provided by empirical science, may permit the *occasional* adjustment to that ontology if it makes the mind-body problem easier. And though she may generally resist e.g., denying that painful states are generally unpleasant, or that red and blue sensations are qualitatively different, she may allow herself to occasionally modify the image of consciousness provided by introspection. One further way in which preservativity seems too strong, moreover, is in forbidding the addition of third person properties to mental states, e.g., the addition of *involving neural oscillations in range 7.5–12.5 Hz* to the conscious state *feeling relaxed*. Couldn't one justifiably add such properties to mental states while still taking consciousness seriously?²¹

21. David Chalmers pushes the last of these points in discussion.

I'm inclined to say yes. But even if physicalists permit such revisions, I would still expect the problem to arise. For though they relax their preservativity here and there, physicalists generally do not relax it enough, nor in the right ways, to banish the problem described in this paper. (Though I emphasize, recalling the end of section 4.4, that they may be right not to.)

The reason is that the typical mainstream physicalist is a realist about phenomenal consciousness.²² She wants to explain it, not explain it away. Hence, whatever other violations of preservativity she is willing to tolerate, she does not abide the wholesale elimination of qualia. But that probably means that after she has made all the additions and subtractions she is warranted and willing to make, mental states are still ascribed properties—qualia—that appear to distinguish them from physical states.²³ This is sufficient to recover the problem. For once she has made all the revisions she is willing to, the physicalist is effectively a preservativist *from now on*. Which means that since her revisions did not remove the source of the discriminability, she is at the beginning of the problem, searching for a full, finite, preservative identification of things that seem discriminable. If the broad point of this paper is right, her search is forlorn. There is no conceivable solution that meets her requirements.

6. Summary Remarks

My aim has been to elucidate an unappreciated logical tangle within the attempt to identify mental and physical (or functional) phenomena. The problem has no ontological upshot. On the assumption that it is possible for a single element to be misrepresented as two distinct discriminables, the problem will impede attempts to identify x with y even when, in fact, $x = y$.

Is the problem *psychological*? Perhaps in a limited sense. It's true after all that, in order for the problem to arise, one needs to have a distinct concepts for the experiential state (e.g., pain) on the one hand and the physical state (e.g., c-fibers

22. For exceptions, see (Dennett 1993), (Blackmore 2002), (Rey 2007). But I do not inherit the burden of explaining *their* puzzlement. For in denying the reality of phenomenal consciousness, they do not encounter the hard problem of identifying it with something physical.

23. This is not deductively guaranteed, however, since it is possible in principle to add and subtract properties, not to the mental and physical states themselves, but to higher order kinds, in a way that makes the qualia indiscriminable from properties of the brain states. For example, by denying those properties of the qualia that make them seem phenomenal (or primitive, see Pereboom 2011: 17) or by adding new, nonphysical properties, not to the brain states, but to the properties of the brain states, creating a kind of second-order property dualism. Though these are potential options, I fear they present themselves as higher-order variants of, respectively, failing to take consciousness seriously and adding to the ontology provided by science. As such, they may seem as antithetical to physicalist realism as the more familiar options at the lower orders.

firing) on the other. But no special kind of phenomenal concept is required. Concepts of any sort will give rise to the problem if they fail to represent their objects as indiscriminable.

It's true that as long as one does *not* represent the identificanda as discriminable, one will not run into the problem. In this much at least, the problem depends on one's having concepts with the right *content*, and content depends on psychological state. However, the *conditional* claim—that *if* one represents *x* and *y* as discriminable, *then* one cannot have a full, finite, preservative solution in which they are identified—has little to do with one's contingent psychology. It will hold of any thinker rational enough to reason well and in observance of Leibniz's Law.

Because certain problems admit of no finite, full, preservative solution, anyone unable to provide an infinite solution and determined to provide a preservative one will find they can only conceive of semi-solutions. But semi-solutions are promissory notes, and when you offer a promissory note with no idea of how to make good, how-possibly questions are a natural upshot. Again this seems to be a pretty general truth about rational beings, and nothing particularly to do with our contingent human psychology.

I stress before finishing that I do not mean to draw a strong mysterian moral about the impossibility of solving the problem of consciousness. For one thing, though *Unfinishability* was motivated by plausible first principles, *Hardness* required further assumptions. For example, it was assumed that our best theory can be partitioned into a first person and a third person part. But this may ultimately be unseated by the development of a "third conception" that acts as a bridge between the two (see Nagel 1998; Stoljar 2001: 273–275). Moreover, though preservativists (or those who are de facto preservativists, after making all the revisions they are willing to in the sense of section 5.8) are unwilling to make revisions to our first and third person notions for the purposes of making the problem of consciousness easier, they may be willing to make them for other reasons. Data from physics may radically alter our image of the physical, and data from psychology our image of the mind, in ways we cannot currently foresee. Perhaps some such change may render psychophysical identifications tractable.

A different way in which the problem might be solved is through the provision of an infinite solution. I know of none, however, and am tempted to think that, for finite creatures like us, there cannot be any. But even if we *can* produce infinite solutions, and it is simply a contingent matter that we have not, the analysis nonetheless predicts how-possibly questions. For whenever you give a semi-solution but cannot *right now* conceive of a infinite full solution, you get how-possibly questions for the interim.

Now and again, though, someone will suggest to me that it is easy to come up with infinite solutions. We need only declare the identity of some mental *x* with some physical *y* and let the rest take care of itself, since the very identity of *x* and

y entails the corresponding identity of their properties, of the properties of their properties, and the properties of those, and so on, to infinity. Hence, with a single stroke, we complete the required infinity.²⁴ This approach, alas, does not work. For though the identification entails the identity of all the properties, no specific mapping of properties is entailed. All we get ‘for free’ is a semi-solution—a quantificational assurance that the properties at each order pair-off *somehow*. But as we saw, when we have a semi-solution with no conceivable full solution, we just end up with how-possibly questions.

Acknowledgements

I thank Bert Baumgaertner, David Chalmers, Kevan Edwards, Joel Friedman, Brie Gertler, Cody Gilmore, Robbie Hirsch, Janet Levin, Jimmy Licon, Brian McLaughlin, Angel Pinillos, David Pitt, Jeremy Pober, Georges Rey, Matthias Scheutz, Adam Sennet, Ted Shear, Paul Teller, participants at the 2012 and 2014 Pacific APA and at the inaugural Northern California Consciousness Meeting, and to anonymous referees, for helpful advice, criticism and feedback. With special thanks to Alejandro Perez Carballo, Chris Hill and Jonathan Simon for extended comments on earlier drafts of this and related work.

References

- Aydede, Murat, and Guven Guzeldere (2005). Cognitive Architecture, Concepts, and Introspection: An Information-Theoretic Solution to the Problem of Phenomenal Consciousness. *Nous*, 39, 197–255. <http://dx.doi.org/10.1111/j.0029-4624.2005.00500.x>
- Balog, Katalin (2012). Acquaintance and the Mind-Body Problem. In Christopher Hill and Simone Gozzano (Eds.), *New Perspectives on Type Identity: The Mental and the Physical* (16–42). Cambridge University Press.
- Blackmore, Susan (2002). There is No Stream of Consciousness. *Journal of Consciousness Studies*, 9 (5–6), 17–28.
- Block, Ned (2007). Max Black’s Objection to Mind - Body Identity. In Torin Alter and Sven

24. Matthias Scheutz, for one, suggested this to me. Alejandro Perez Carballo makes the related suggestion that though we cannot make an infinity of explicit changes to our beliefs, it may be enough that we are disposed to. But though it’s plausible that we would be disposed, in making a single lower-order identification, to infer each of its infinitely many higher-order consequences, there are no consequences here of the relevant sort. Consider the imaginary researcher in section 5.1 who moves from identifying pain with lamina I activity, at the first order, to identifying the unpleasantness of the pain with the activity’s effect on anterior cingulate, at the second. However he justifies the first of these identifications, the second does not deductively follow. He remains free to identify the unpleasantness with some other feature of the lamina I activity. So there is no reason to think that, having made the first identification, he would be naturally disposed to favor the second. The same goes, *mutatis mutandis*, at every step of the regress. Cf (Molyneux 2012: 293–294).

- Walter (Eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism* (249–306). Oxford University Press.
- Carruthers, Peter (2004). Phenomenal Concepts and Higher-Order Experiences. *Philosophy and Phenomenological Research*, 68, 316–36. <http://dx.doi.org/10.1111/j.1933-1592.2004.tb00343.x>
- Chalmers, David (1996). *The conscious mind*. Oxford University Press.
- Damasio, Antonio (2000). *The Feeling of What Happens*. Vintage, Random House.
- Dennett, Daniel. C. (1991). *Consciousness Explained*. Penguin.
- Dennett, Daniel. C. (1993). *Quining qualia*. The MIT Press.
- Dennett, Daniel. C. (2005). *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*. The MIT Press.
- Goldman, Alvin (2002). *Pathways to Knowledge, Public and Private*. Oxford University Press. <http://dx.doi.org/10.1093/0195138791.001.0001>
- Hill, Christopher (1997). Imaginability, Conceivability, Possibility, and the Mind-Body Problem. *Philosophical Studies*, 87, 61–85. <http://dx.doi.org/10.1023/A:1017911200883>
- Hill, Christopher. S. (1984). In Defense of Type Materialism. In *Meaning, Mind and Knowledge*. Oxford University Press. <http://dx.doi.org/10.1007/BF00869337>
- Ismael, Jenann (2007). *The Situated Self*. Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780195174366.001.0001>
- Levin, Janet (2007). What is a Phenomenal Concept? In Torin Alter and Sven Walter (Eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism* (87–110). Oxford University Press.
- Levine, Joseph (2004). *Purple Haze: The Puzzle of Consciousness*. Oxford University Press.
- Loar, B. (1990). Phenomenal States. *Philosophical Perspectives*, 4, 81–108. <http://dx.doi.org/10.2307/2214188>
- Lycan, William. G. (2003). Chomsky on the Mind-Body Problem. In Louise M. Antony and Norbert Hornstein (Eds.), *Chomsky and His Critics*. (11–28). Blackwell Publishing.
- Macdonald, Cynthia (2004). Mary Meets Molyneux: The Explanatory Gap and the Individuation of Phenomenal Concepts. *Nous*, 38 (3), 503–524.
- Molyneux, Bernard (2011). On the Infinitely Hard Problem of Consciousness. *Australasian Journal of Philosophy*, 89 (2), 211–228. <http://dx.doi.org/10.1080/00048401003745985>
- Molyneux, Bernard (2012). How the Problem of Consciousness could Emerge in Robots. *Minds and Machines*, 22 (4), 277–297. <http://dx.doi.org/10.1007/s11023-012-9285-z>
- Nagel, Thomas (1998). Conceiving the Impossible and the Mind-Body Problem. *Philosophy*, 73 (285), 337–352. <http://dx.doi.org/10.1017/S0031819198000035>
- Neander, Karen (1998). The Division of Phenomenal Labor: A Problem for Representational Theories of Consciousness. *Nous*, 32 (S12), 411–434. <http://dx.doi.org/10.1111/0029-4624.32.s12.18>
- Ney, Alyssa (2008). Physicalism as an Attitude. *Philosophical Studies*, 138, 1–15. <http://dx.doi.org/10.1007/s11098-006-0006-4>
- O’Dea, John. (2002). The Indexical Nature of Sensory Concepts. *Philosophical Papers*, 31 (2), 169–181. <http://dx.doi.org/10.1080/05568640209485100>
- Papineau, David (2002). *Thinking About Consciousness*. Oxford University Press.
- Pereboom, Derk (2011). *Consciousness and the Prospects of Physicalism*. Oxford University Press.
- Perry, John (2003). *Knowledge, Possibility, and Consciousness*. MIT Press.

- Rey, Georges (2007). Phenomenal Content and the Richness and Determinacy of Colour Experience. *Journal of Consciousness Studies*, 14 (9-10), 112-131.
- Stoljar, Daniel (2001). Two Conceptions of the Physical. *Philosophical And Phenomenological Research*, 62 (2), 253-281. <http://dx.doi.org/10.1111/j.1933-1592.2001.tb00056.x>
- Stoljar, Daniel (2005). Physicalism and Phenomenal Concepts. *Mind and Language*, 20, 469-494. <http://dx.doi.org/10.1111/j.0268-1064.2005.00296.x>
- Sturgeon, Scott (1994). The Epistemic Basis of Subjectivity. *Journal of Philosophy*, 91, 221-35. <http://dx.doi.org/10.2307/2940751>
- Tye, Michael (2003). A Theory of Phenomenal Concepts. *Philosophy*, 53, 91-106. <http://dx.doi.org/10.1017/CBO9780511550294.006>