By the end of the last century, it started to look like the philosophy of mind had stark choices to make. Accept the idea that intuitions give some kind of guide to possibility, and dualism seemed to be just around the corner, or so David Chalmers\(^1\) told us. Deny that intuition gives any guide to possibility, and we can dismiss easily Chalmers's "Zombie" argument and its cousins\(^2\)—but then, as Frank Jackson\(^3\) has taught us, we have no explanation, paradoxically, for the core arguments for the relevance of the a posteriori in metaphysics. The best explanation depends crucially on thought experiments—intuitions about possible cases—like twin earth.

Here is a reminder of the story so far. Intuition seems to tell us that "Zombies" are possible: physical duplicates that lack qualia.\(^4\) Assume we have qualia. If there is a possible world that duplicates this one physically and only physically, and that lacks qualia (which is

---

\(^*\)This paper has benefitted enormously from input from David Chalmers, Frank Jackson, Fred Kroon, Kristie Miller, Philip Pettit, Denis Robinson, and Laura Schroeter.


\(^4\) Throughout this paper I am using 'qualia' in the sense that is neutral as to whether they are physical states or not. Of course 'qualia' is a technical term, so the sense in which I am using it has, if we are to find a folk analysis, to be thought of as a synonym for nontechnical notions like 'sensory experience', or 'raw feel' stripped of any particular philosophical theory about such experiences. There is, of course, another terminological tradition in which the name is reserved for nonphysical states, and the debate is whether there are any.
what a Zombie world amounts to) then we have a minimal physical duplicate\(^5\) of actuality that fails to be a duplicate _simpliciter_, and so physicalism is false. The complete way things are physically would turn out not to be sufficient for qualia.

How should the physicalist respond? One way is to rule out such a case on a priori grounds. What I might call _procrustean analytical functionalism\(^6\) does this: it is just a priori on proper reflection, that mental states are what play the relevant causal functional roles, and so a thorough understanding of the meaning of mental language rules out the logical possibility of the roles being played in the absence of mental states. Aside from its implausibility, this kind of functionalism embraces the a priori presumably in part through reflection on intuitions about cases, and thus would need to give an account of why the fairly robust intuitions about the possibility of Zombies is trumped by the apparently more recherché intuitions which ground its claims for analyticity. It seems arbitrarily to privilege certain intuitions.

Probably the most prominent response is to deny that intuition is any guide to possibility at all.\(^7\) After all, the a posteriori physicalist tells us, it may have seemed intuitive that it was impossible for water to be black and tarry. But we have learned from Saul Kripke, Hilary Putnam, and others\(^8\) that this is false, since if H\(_2\)O were to be black and tarry it would be water nonetheless. Equally so, they would argue, for physicalism about the mind. Neuroscience will tell us what minds are, and thus tell us what is necessary. Any intuitions about the possibility of mental states not of this appropriate kind are in the same pre-scientific class as those we might have had about the possibility of non-H\(_2\)O "water" before the nature of water was well understood. So there are necessary truths, such as the nature of water, which since

---


they are a posteriori do not depend on intuitions. So intuition is not a guide to possibility or necessity.

But this is too simple. There is a compelling account of how the necessary a posteriori works according to which in each case the arguments for necessity depend on an intuition about possibility. In the case of water, an intuition that what on twin earth failed to share the chemical structure of the watery stuff we are in contact with would not be water, in the case of mental states the intuition that what failed to share neuro-functional properties with paradigm minds would not count as mental. So this approach could be redescribed as in fact depending on an analytic claim. In the case of qualia, it would be the claim that qualia are whatever neural state types are in fact actually causally responsible for our discriminative powers. Once we have added this additional analytic premise, we can see that it is very similar to analytical functionalism: it just has an actually clause added to make it rigidify on the actual occupants of the role. The claim that intuition is no guide to possibility turns out to depend on intuitions about possibility, and thus is either inconsistent or else open to the charge of choosing between intuitions arbitrarily. If we accept those intuitions as being a guide to the meaning of ‘qualia’, why reject the Zombie intuition?

Out of all this, there emerge a number of claims that seem right to me. First, physicalism is, I think, probably true, and some kind of substantive functionalism is my preferred brand. Mental states are, I think, whatever play certain causal roles (or if you prefer, whatever is intrinsically similar to that which actually plays the roles: for current purposes this is an in-house dispute). So the first claim will be:

(1) Substantive functionalism about qualia: qualia are whatever play certain roles.

But, I do not intend this claim to be directly analytic. So that gives us (2).

(2) The claim that ‘qualia are whatever plays certain functional roles’ is not directly analytic.

---


11 This is not uncontroversial: see Stalnaker.

12 Of course, one could tell a story about equilibration, in which the Zombie intuition is reluctantly cast aside in favor of a larger consistent set, but that underplays the vividness of it for many.
On the balance of probabilities, physicalism is true, so:

(3) Whatever plays those roles is physical, so the qualia around here are physical states.

I agree with Chalmers that the Zombie intuition is powerful, and that modal intuitions play a role in settling meaning.

(4) The intuition about possibility that underlies Chalmers's Zombie arguments has as good a claim to be taken as a prima facie guide to possibility as any other equally robust intuition. No story that tells us that Zombies are impossible can do so without explaining why this intuition is misleading.

(5) In general, intuitions (or perhaps the dispositions that the intuitions are a fallible guide to) are a guide to possibility via analyticity.

These five principles seem difficult to hold together. Agreeing that we have the Zombie intuition and that it is a guide to possibility seems to rule out (3). And (2), which says that it is not directly analytic that functionalism is true, undermines the usual reasons for holding (1).

I think these claims can in fact be reconciled. I plan initially to explain the reconciliation by raising the problem in a parallel case: the example of water that has occupied so much of the literature.

1. WATER AND AQUEOUS ESSENCE

Suppose that things were a little different from how they are with respect to the word 'water'. Imagine we speak a slightly different language (and have slightly different thoughts) in which the following claims are true. Although H₂O is in fact water, once it was thought that there was an "aqueous essence"—a nonphysical property distinct from functional properties like fluidity, clarity, potability, and so forth—which was what made water the substance that it is. It was not that it was thought that aqueous essence was the nonphysical cause of clarity, fluidity, and so on. These properties were thought to be markers of the presence of aqueous essence if there was any. Aqueous essence was thought to be causally idle, much as parallelists about the mind think the mind is causally idle. And suppose that this thought was still present enough in people's thinking about water that they were prone to have intuitions that something could be a physical duplicate of water without being real water (in virtue of lacking the aqueous essence). I do not think the concept of water is or was like this, but my purpose here is to show how such a concept would be both coherent, and compatible with physicalism about water, understood in this different way.

To show that this variant concept is both coherent and compatible, we will have to provide some sort of analysis of 'water'. If it is analytic
that water has to possess some nonphysical aqueous essence, then, of course, if physicalism were true there would be no water. So, any such analysis would not have the coherence I am suggesting. Suppose, on the other hand, the right understanding of ‘water’ was what may actually be the correct understanding: the chemical properties that explain the functional properties of the paradigm samples of water. Then if physicalism were true, the intuition about possibility would be just mistaken, and thus the package I am trying to make consistent equally would shatter.

But what about the following analysis?

AE: If there is actual aqueous essence, then that which possesses it is water and necessarily so. Otherwise, water is whatever shares explanatory chemical features with paradigm samples.

This is what I shall call a conditional analysis. There is a sense in which the right substantial analysis depends on how things actually are. But grasp of the concept does not depend on knowing how things actually are, hence the concept could be grasped before knowing either whether there is any actual aqueous essence or even what the underlying chemical properties of the paradigms are.

How should we understand a conditional analysis like this? I propose to use the strategy of two-dimensionalism: using a two-dimensional matrix, where we consider different hypotheses about how the actual world is, and then consider what sets of counterfactual judgments about logical possibility we would make on various presuppositions about the actual world.

Usually, of course, the different ways actuality is conceived to be are variant physical hypotheses (the existence of XYZ—understood as physical but perhaps counternomic, and so on). But there is nothing to prevent us from considering ways things could be considered as actual that include the presence of strange possibilities such as nonphysical aqueous essence. After all, none of us should assign a zero credence to the claim that things are in fact this way!

I.1. Evaluation of ‘water is H₂O’. So let us evaluate some claims about water under these varying circumstances. Let us start with ‘water is

---


14 I write here in the indicative mood, which suggests an understanding of two-dimensionalism according to which one dimension is purely epistemic possibility, and the other metaphysical (Chalmers prefers such an analysis in his forthcoming “The Foundations of Two Dimensional Semantics”). Nothing here depends on this reading.
H₂O'. I shall illustrate this evaluation with a matrix consisting of rows each of which illustrates a different hypothesis about actuality (the hypothesis on the row headings). Each box in the row illustrates what to say about the expression being evaluated in a counterfactual world, the relevant features of which are described in the column heading. I shall not provide further matrices for the other expressions; this example will, I hope, be enough to help those who are in general helped by diagrams.

The headings on rows and columns ‘AE’ mean that there is some aqueous essence in the world (that is, the nonphysical substance co-present with whatever is responsible for clarity, potability, and so forth). The headings ‘H₂O’ and ‘XYZ’ stand for those physical substances that are what is potable, clear and so forth—in short, the substances that play the water role. By hypothesis about the deviant concept, if there is any AE, water is AE, not that which plays the water role.

<table>
<thead>
<tr>
<th></th>
<th>AE, H₂O</th>
<th>AE, XYZ</th>
<th>−AE, H₂O</th>
<th>−AE, XYZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>AE, H₂O</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>AE, XYZ</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>−AE, H₂O</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>−AE, XYZ</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
</tbody>
</table>

‘Water is H₂O’

In all the worlds considered as actual in which aqueous essence is present, water is not H₂O. From the perspective of *these* worlds, this is a necessary claim: all and only the counterfactual worlds that contain AE contain water, and all and only the AE across all worlds is water. So all the rows in which there is actual AE are ones where ‘water is H₂O’ is false in every box: it has a necessarily false *C-intension* at these rows.

In general, a *G-intension* of a term is a function that assigns to each world the extension of that term on the assumption that we keep the actual world fixed as whichever one is in fact actual—it is close to the ordinary use of ‘intension’. Two-dimensional diagrams illustrate the point that since this function can vary according to context, there are as many *G-intensions* as there are contexts that we can consider to be actual. The expression *G-intension* (for counterfactual intension) is due to Jackson (op. cit., pp. 48–49), and is motivated by the idea that most of the worlds in a *G-intension* are thought of as counterfactual:
the first world in a row of a table is thought to be actual, and the other worlds are from its perspective mere counterfactual possibilities. Some may be familiar with the idea from Chalmers’s expression *secondary intension*.

There are some ways things could be actually where ‘water is H$_2$O’ is false, as well as some ways things could be actually where it is true. So, in the language of two-dimensionalism, we say that ‘water is H$_2$O’ has a contingent $A$-intension.$^{15}$ The diagonal that we can draw where we consider only the worlds considered as actual and evaluate the expression in those worlds contains boxes where we assign ‘T’ to ‘water is H$_2$O’ and boxes where we assign ‘F’. This is in virtue of the first clause of the analysis. We were entitled to ignore the (physical) intrinsic nature of whatever played the water role,$^{16}$ because if there is AE then AE is water and necessarily so.

In general, an $A$-intension of some term is the function that assigns to each world the extension of that term on the assumption that that world is itself actual. Chalmers uses the expression *primary intension* for this notion.

So let us consider the worlds (considered as actual) where there is no AE. In these worlds, the conditional analysis directs us to consider what physical properties play the water role. If a world considered as actual has no AE and it has H$_2$O playing the water role, then H$_2$O is water *and necessarily so*. From the perspective of a world considered as actual without AE, any counterfactual worlds that contain no H$_2$O and do contain some mysterious nonphysical AE property do not contain water, they merely contain a weird alien property. Thus, if things are this way actually, then we should judge it to be necessary that water is H$_2$O. Similarly, if actually the water role is played by XYZ, then from that perspective the claim ‘water is necessarily XYZ’ is true.

What is the take-home message? That ‘water is H$_2$O’ comes out true and necessarily so on the supposition that actually H$_2$O plays the water role and there is no AE, and false otherwise. In the language of two-dimensionalism, on this supposition, ‘water is H$_2$O’ has a necessary $C$-intension.$^{17}$

1.2. Evaluation of ‘water is the watery stuff’. Consider now the claim that water is the watery stuff (that is, what plays the water role is

\[\text{15 For Chalmers, contingent 1-intension or primary intension; for Gareth Evans and Humberstone and Davies, deeply contingent.}\]

\[\text{16 That is, fills the lakes, the river, taps, and so on.}\]

\[\text{17 The terminology gets baroque: for Chalmers, it is a necessary 2-intension or secondary intension on this supposition. On some understandings, this corresponds to what Humberstone and Davies might state by “water is H$_2$O” is superficially necessary.”}\]
water). This comes out as false when we consider worlds considered as actual that contain AE—for AE plays no causal role. And it comes out as necessarily false from the perspective of these strange worlds, for worlds considered as counterfactual are judged by the same standard as the world taken to be actual: all and only the causally idle AE is water.

What, though, of the worlds in which naturalism is true? Well, if \( \text{H}_2\text{O} \) plays the water role actually, then all and only the worlds where \( \text{H}_2\text{O} \) plays the water role are worlds where water plays the water role (a world considered as counterfactual from this perspective where XYZ plays the water role is a world where the water role is played by something other than water). But from the perspective of worlds considered as actual where XYZ plays the water role, all the worlds where \( \text{H}_2\text{O} \) plays the water role are worlds where something other than water plays the water role. So ‘water plays the water role’ is not necessary from any perspective, for there is always a world considered as counterfactual where it is false.

Sometimes at this point in a discussion of two-dimensionalism about water that does not consider nonnatural worlds considered as actual, the claim is made that, although never necessary, we can see from the analysis that ‘water is the watery stuff’ is true a priori. This is because, however the actual world is, if something plays the water role, the analysis says that it is water. So we know that—surprise, surprise—water plays the water role (if there is any). This corresponds to ‘water plays the water role’ being true down the diagonal of a two-dimensional diagram.\(^{18}\) This will not be true on our conditional analysis, for in the nonnatural worlds that contain AE, water is not what plays the water role, even if \( \text{H}_2\text{O} \) does. But there is a strange kind of conditional thing that we can know a priori on this model: if there is no AE actually, then water plays the water role. The apparent a priority of ‘water plays the water role’ does, however, correspond to its being true in all the naturalistic worlds where the same world is treated as world of index and of context. Thus, for this imaginary variant concept of water I am using to illustrate the idea, ‘water plays the water role’ has a contingent A-intension, though it has a restrictedly necessary A-intension whenever the actual world has no aqueous essence.

\(^{18}\) And thus has a necessary A-intension or 1-intension.
there is no world considered as actual in which we could deny it, even though it is not necessary. Consider the world as actual in which H₂O plays the water role. In that world, water plays the water role, so the disjunction is true. But consider a counterfactual world with no AE and in which XYZ plays the water role: neither disjunct is true. So the disjunction is not necessary, even though a priori. The disjunction is the equivalent of the contingent a priori claim ‘water plays the water role’ on a more standard understanding of ‘water’. On this toy analysis, ‘water is AE or it is whatever plays the water role’ has a necessary A-intension.

What all this shows is that a concept of water could coherently be one which prioritized the nonphysical, consistently with physicalism about water being both true and necessarily true. Nonetheless the concept’s nature explains an intuition that there could be something physically like water but which fails to be water. This is because we ought to grant some credence to the thought that the actual world contains aqueous essence; and if it does, then the very same analysis of water would tell us that there can be physical duplicates of water that fail to be water. Of course, given how we think things actually are—where we place most of our credence—the analysis tells us to use ‘water’ in a way that makes it impossible for a physical duplicate to fail to be water. But important though intuitions are in buttressing arguments for possibilities, we should not expect them to do anything quite so fine-grained as distinguishing between a genuine possibility that a physical duplicate of water could fail to be water, and a credence in things being actually such that, with the same concept of water, the content of claims about water is different enough to allow that physical nature does not settle what counts as water.

I propose to do something similar in the case of qualia. I shall develop a conditional analysis of ‘quale’ which explains the intuition that Zombies are possible, by showing how the very same concept of quale gives rise to a different content should physicalism not be in fact true in the actual world.

19 See Chalmers, “The Foundations of Two Dimensional Semantics.”
20 I use ‘necessary’ to correspond to the necessary Gintension under some supposition about actuality, what Evans called “superficial necessity.” This terminology confuses some these days, many naturalists are inclined to think anything a priori is superficial, and necessities like ‘water is H₂O’ are deep.
II. A CONDITIONAL ANALYSIS OF ’QUALE’

An important mistake, which has bedeviled the philosophy of mind, is to try to use conceptual matters to determine whether physicalism is true or not. I take that to be an empirical matter or else one of substantial metaphysics. I do not plan to directly argue for this claim here, instead I shall recommend a procedure that reverses the direction of inquiry, and hope the benefits this reversal brings are argument enough.

Instead of asking how our concept settles the truth or otherwise of physicalism, we should ask what our concept tells us about the mind, should there actually be nonphysical states (say, ones whose intrinsic nature is grasped in experience), and what it tells us about the mind should physicalism be true.

Suppose then that, contra my expectations, it turns out that there are special properties left out of the physical picture of the world, whose intrinsic nature (and perhaps no other intrinsic natures) are grasped in experience. I shall call such states henceforth the spooky states. If there are spooky states—and they are the things whose nature I grasp when I experience a quale of red—then those things are indeed the qualia of redness. If there were such states, I think we would be inclined to take it that they are necessary for having qualia. That this is somehow conceptually important for the nature of qualia is evidenced in the tendency for people to default to a kind of dualism in the absence of arguments for physicalism.

But what if it turns out that there are no spooky states? That is pretty much the consensus in the natural sciences, and yet this alone does not seem to lead to the abandonment of the idea that we have qualitative experience; it rather leads to attempts to account for qualitative experience physically. I think there is consensus among many materialist philosophers of mind that the states which play complex functional roles which underlie our discriminatory capacities and sensory talk are in fact our qualia. There is less agreement on what it takes to have qualia for beings unlike us. Analytical functionalists (and those who hold successor views like mine) think that what it takes is to have those roles occupied, and that in each case the qualia

---

22 This does not rule out the possibility that some versions of dualism, if false, are a priori false—because there is some hidden incoherence. But even in this case what I can consider can be reconstructed with worlds considered as actually understood as a kind of epistemic possibility—in the same way as there is some sense in which it is epistemically possible that Fermat’s Last Theorem is not a theorem (maybe there is a flaw in the proof).

23 Lewis, “Ramseyan Humility” (forthcoming).
are the occupiers of the roles.\textsuperscript{24} Others—who hold the view that Jackson and I have called \textit{empirical functionalism}\textsuperscript{25)—think that what it takes is to have states similar in some way to the states that occupy the roles in us (the debate being as to the level and degree of similarity: physical, neural, psychofunctional, computational, and so on). But what if, counterfactually, we consider worlds that have spooky states? On the assumption that actually there are none of these spooky states, strange nonphysical states in other possible worlds do not get to count as qualia unless they play the functional roles. For strange nonphysical states that fail to play the functional role would count as what I mean by ‘quale’ only if they are what I am \textit{actually} directly acquainted with in experience.

So here, then, is an analysis:

If there are spooky states then

- In the actual world the qualia are the spooky states; and all and only the qualia in counterfactual worlds are the spooky states

Else

- In the actual world the qualia are the states that play the functional roles; in other worlds, qualia (if any) are the states that play the roles \textit{in that world}.

Anyone who is a friend of empirical functionalism should feel free to substitute the following for the second bullet point: the issue of the rightness of that analysis has been argued elsewhere\textit{(ibid.)} and, again, nothing in this paper depends on it.

- In the actual world the qualia are the states that play the functional roles; in other worlds the qualia, if any, are the states that play the roles in the actual world.

\textbf{III. SOME DEFENSE AND SOME CLARIFICATION}

The analysis we have provided so far uses possible worlds \textit{simpliciter}. It does not refer to locations in a world. Unfortunately, this is too much of a simplification to be of use except as preliminary exposition. We have not said where in a world the nonphysical states occur in order to count as satisfying the antecedent.

In addition, once we specify this, we will need both to consider

\textsuperscript{24} Though some say that they are the second-order state of the role’s being occupied.

worlds considered as counterfactual with the spooky states located at various points, and to consider worlds considered as actual with spooky states located at various points.

Let us begin with the counterfactual claim. What if it were the case, contrary to fact, that George Bush were the only person in the world who had the appropriate spooky states? Let us start by evaluating this claim on the assumption that in the actual world there are no spooky states anywhere. In this case, I claim that the right thing to say is that in the counterfactual world, there is a strange nonphysical property that is not the property of being a quale. It is just a world where Bush is stranger than he is actually. But conceived of as actual, the world where Bush has spooky states is one where he has qualia. So the special conceptual priority of the nonphysical properties is not one that extends to counterfactual worlds. It is only if actually there are spooky properties as well as functional or physical ones that the concept of quale latches onto the spooky ones.

But what to say of a world conceived of as actual where Bush has the spooky states but I do not? It would be disturbing if this counted as a case where I discover that I lack qualia. This would mean that in principle I could make the discovery that I lacked qualia (and thus perhaps could legitimately be tortured) by discovering nothing about myself, but rather by discovering strange metaphysical truths about Bush.

On the other hand, the analysis needs to give the answer that a world conceived of as actual where I possess the spooky property is one where it is true that qualia are spooky, and necessarily so (in the sense of necessary Gintension).

How to pull off this trick? One possibility is to do it all with averages. 'There are spooky states' means something like "it is usual in the paradigm community for there to be properties that obey the identification hypothesis" or something like that. This however leaves one open to various kinds of implausible contextualism. In particular, if actually most people have the property, and I do not, then I discover I lack qualia.

The right solution is to accept that the concept of qualia has a hidden indexical structure. My use of the term is governed by a presupposition that I cannot fail to have qualia if it seems to me that I do, even though I could counterfactually lack qualia. (Relative to an actual world in which I have the spooky property I could lack

---

qualia by lacking the spooky property. Relative to actual worlds where there is no spooky property I could lack qualia by being, say, a Peacockian Martian Marionette.\(^{27}\)

So, in my mouth the term ‘quale’ functions thus: if there is a nonphysical property whose intrinsic nature I am acquainted with in experience, then that is my quale, and it is a necessary condition for anything to be a quale that it be a property like that. If not, then my quale is whatever state plays a certain causal functional role, and this is sufficient for anyone to have a quale.

The egocentricity captures, I think, what is acceptable about the problem of other minds. I can think of someone else that he probably does have qualia. But on the somewhat mysterious assumption about the nature of qualia that would be right in a world conceived of as actual in which there are spooky states, it is an open question for me. Of course, assuming physicalism is actually true, the functional analysis deflates the problem of other minds, for I have reasonable epistemic access to others being in states which play the roles in them. Therefore, whether there is a problem of other minds will depend on whether (actually) there is a spooky state with which I am acquainted. So the worlds conceived of as actual need to be centered worlds, and the claim that ‘there are spooky states’ needs to be evaluated at their centers.

There is still one final comment to make on these matters: of course if the way things actually are is one in which Bush and no-one else has the spooky states, then he, Bush, can claim and claim truly that the rest of us are all Zombies. But of course the content of ‘Zombie’ in his mouth is a little different from the content of Zombie in mine, when I claim in my idiolect that I am not a Zombie (even though both our terms function similarly). This is something of a cost, but I think that it does reflect what remains of the significance of the problem of other minds. The problem arises exactly because there is something idiolectic about all of our uses of terms like ‘quale’. And the problem fades on plausible assumptions about how things are that ensure that the idiolectic differences come to nothing.

We can now modify the analysis with which we began by relativizing in the following way:

For some speaker, S (who is the center of a centered world):

If there are spooky states at the center of the actual world then

\(^{27}\) Christopher Peacocke, *Sense and Content* (New York: Oxford, 1983), see pp. 204–05.
In the actual world all qualia are the spooky states; and all and only the qualia in counterfactual worlds are the spooky states.

Else

In the actual world the qualia are the states that play the functional roles; in other worlds qualia (if any) are the states that play the roles in that world.

IV. DOING SOME TWO-DIMENSIONAL MATRICES WITHOUT TABLES

Analogously with the case of water, I will now evaluate a few propositions associated with qualia to see how they behave in two-dimensional contexts.

IV. 1. Evaluation of ‘qualia are NSI states’. I shall start with the evaluation of a physicalist claim about the nature of qualitative states. Suppose that some physical state or states are found to underlie the typical functional roles (in recognition and manipulation and so on) of qualia. Call these the physical correlates of qualia. On some physicalist doctrines, these count as necessary conditions for qualia. Let us use the label ‘NS1’ for some neural state that correlates with qualia. On these sorts of physicalisms ‘qualia are NS1 states’ would be true necessary a posteriori. This empirical functionalist reading is not something with which I agree; however, nothing here hangs on this as I said above.

Under what conditions is it true that qualia are NS1 states given our analysis? If in the actual world there are relevantly placed spooky states, then qualia are never NS1 states. So in a two-dimensional matrix ‘qualia are NS1 states’ has a necessarily false Gintension in all the rows where actually there are spooky states. In the worlds considered as actual where there are no spooky states, it is irrelevant whether or not qualia are NS1 states in the actual world: qualia are NS1 states in every world where NS1 states play the functional role regardless of what play the roles actually. So ‘qualia are NS1 states’ is true down a part of a column in which we look at worlds where qualia play the NS1 role: that part of the column in which the designated or actual worlds are non-spooky.

Still, there is no row in which it is always true, and nor is it true

28 Braddon-Mitchell and Jackson, “The Divide and Conquer Path to Analytic Functionalism.”
29 Of course this is an absurd simplifying assumption: every quale type will correlate with a different physical type, and, for all we know, every token as well.
30 This is the contrast with empirical functionalism, where NS1 states are qualia if and only if they actually are.
down the so-called diagonal: there are worlds considered as actual in which it is false (any world where either NS1 states do not play the role, or in any world at all when the actual world has spooky states). Thus this claim is what Gareth Evans\textsuperscript{31} called deeply contingent—indeed it is contingent simpliciter.

\textbf{IV.2. Evaluation of ‘qualia are whatever play the functional roles’}. For what I have called procrustean analytical functionalism, the claim that ‘qualia are whatever play the functional role’ is analytic. It is true right down the diagonal—in every world considered as actual. It is also necessary. Even empirical functionalism gives us an $A$-intension in which it is necessary (though then it will not have necessary $C$-intensions because in every row there will be worlds where what plays the role is not what plays the role in the actual world). On the conditional analysis we are considering, however, ‘qualia are whatever play the functional roles’ will come out as having neither a necessary $A$-intension nor in general necessary $C$-intensions. For if in the actual world there are spooky states, then at any world where there are spooky states that do not play the causal functional roles, the states playing the roles are not qualia. Thus, the claim that ‘qualia are whatever play the functional roles’ has a contingent $A$-intension. On the (controversial) assumption that necessary $A$-intension amounts to analyticity, we can see that on this assumption the constitutive claim of procrustean analytical functionalism comes out false. It is not analytic that qualia are whatever play the functional role.

And yet the conditional guarantees something like analytical functionalism: the part of a two-dimensional matrix that leaves out all the rows where the actual world has spooky stuff is the same as the matrix for analytical functionalism. On the assumption that the actual world is physicalistic, we are required to regard whatever plays the role (actually and counterfactually) as qualia. But what kind of claim is this? Remember that I have rejected bare necessities. There needs to be some analytic claim that grounds everything. There needs to be some proposition with a necessary $A$-intension that will do the job for my conditional analytical functionalism that the claim that ‘qualia are whatever it is that plays the role’ does for procrustean analytical functionalism.

\textbf{IV.3. Evaluation of ‘there are spooky states and they (and only they) are the qualia, or there are no spooky states and the qualia are the states that play the functional role’}. This is the claim that does the job. No matter how things are actually, I know that I am either in direct contact with the

essential properties of spooky stuff in my phenomenal experience, or I am not. If I am, I know that they are the qualia (so the conditional that if there are spooky states, then ‘they are qualia’ is true at every world down the diagonal) and I know that if there are no spooky states, then the states that play the role are qualia (so the claim that if there are no spooky states then qualia are the states that play the role is true down the diagonal). All of this is part of a priori analysis. I hope that this article will do part of that job—the part of persuading the reader that if there are such spooky states, then they are qualia, and elsewhere I have a priori arguments for analytical functionalism that can be modified to support the more limited claim that if there are no spooky states then qualia are the states that play the role.32

V. WHY THIS EXPLAINS THE ZOMBIE INTUITION

The point of this somewhat complicated analytical framework is not just that it is independently correct. For our current purposes, the benefit is that it gives an account of our concept of qualia that allows for the truth of physicalism while explaining the Zombie intuition: the intuition that there could be a physical duplicate that lacked qualia.33 Why is this so? Well, technically the answer is clear. The claim that ‘there is a physical duplicate of us with no qualia’ has a contingent A-intension on this analysis. For there is a world considered as actual where the spooky property exists (say I am an ectoplasmic being with no causal connection to the world, and am directly in contact with the intrinsic nature of experience). In this world, if there are any physical beings floating around that answer to the physical descriptions we give of ourselves,34 and are unconnected to the spooky properties, they lack qualia. And there are ways things could be actually in which although we are physical beings and possess qualia (because of psychophysical laws that connect the physical to the nonphysical) it would be true, if things are actually that way, that this would be logically contingent, since if that world were actual then there would be counterfactual worlds in which the laws are missing. It would be true to say that if we were not connected to the nonphysical in that way, we would lack qualia. So there is a way things could actually be,

33 Strictly speaking, this is an absent qualia case. Chalmers thinks of Zombies more generally as lacking consciousness: more on this later.
34 There is a complexity here: epistemically, if I am actually ectoplasmic, what do we say about the physical? Plausibly on an indicative reading of the conditional, the physical should be as it seems to be, and the ectoplasmic difference as small as possible reconciled with my knowledge.
according to which it is true that I do not lack qualia, but it is possible that my physical duplicate would (say, in worlds where any law-like connection between the physical and the nonphysical is broken).

The physicalist says, however, that things are not actually this way. The way things are is a way that makes it false to say that a physical duplicate of mine could lack qualia. On physicalist assumptions about actuality, ‘Zombies cannot exist’ has a necessary G-intension. The contingent A-intension explains the intuition that it is possible; the necessary G-intension explains the sense in which it is not.

Another way of putting this has to do with the fine-grainedness of modal intuitions. While I think they are important, it is not clear that they can distinguish two cases. An intuition is a kind of picture that makes sense: I picture a world with physical duplicates that lack qualia, and it does not seem to be incoherent. But of what is it a picture? I do not think that intuition gets this clear. Is it a picture of what is possible, given how things are; or does it correspond to the fact that as a physicalist, but hopefully not an arrogant one, I do not give zero credence to physicalism being false? And if I do not give a zero credence to this, I cannot give a zero credence to the expression in my mouth ‘Zombies are possible’ expressing a truth—even though it does not do so given how things are, and cannot do so. Intuition does not distinguish between a non-zero credence in things being such that ‘there could be physical duplicates that are Zombies’ might be rightly thought true, and a direct intuition that it does express a truth given how things actually are.

This begins to explain why there is such a dispute in the philosophy of mind about whether Zombies are possible. I think it also explains more fine-grained features of the dispute. It is no accident that it tends to be physicalists who lack the intuition that Zombies are possible, and dualists who have it. The uncharitable explanation is that each has intuitions designed just to shore up their view. But a two-dimensional understanding of the concept gives a better explanation. A committed physicalist has a firm view about what world is actual. For her, the relevant possibilities are about what is possible given how things are, and if things are the way physicalism says, then Zombies are impossible (given the necessary C-intension). The committed dualist thinks there are actual spooky properties, and that they are our qualia. She thinks she lives in a world where ‘Zombies exist’ has a contingent G-intension. So, the possibility of Zombies is something to be defended strongly. Thus, on the second understanding of the concept, we should expect anyone with firm intuitions about how things actually are—about whether physicalism is actually true or false—to have strong intuitions in different ways about the possibility of Zombies: for they differ in
the G-intension that they assign to the sentence. And, of course, someone who is undecided about the truth of physicalism will tend also to defend the claim that Zombies are possible, but for a different reason. If someone is undecided about the truth of physicalism, then she entertains alternative hypotheses about how things actually are, and thus will tend to concentrate on the A-intension. Since the A-intension is contingent, it will seem as though ‘Zombies are possible’ is true simpliciter, and so without the aid of the clarification provided here, it will seem like an argument for dualism. The upshot is: some intuitions about what is actual are driving some intuitions about what is possible. Thus it is a kind of circularity to argue from those intuitions about what is possible to claims about what is actual.

V.1. An objection: “this just says we might have meant something else.”

Here is an objection that has been pressed on me by many. It runs thus: what you have explained is not why we have an intuition that Zombies are possible. You have only shown that one might be a physicalist and think that we might have meant something else by Zombie such that it would have been right to say that Zombies are possible. And, of course, we always knew that in the world where ‘Zombie’ means “Zebra,” it would be right to say ‘Zombies are possible’. And epistemically ‘Zombie’ might (though it is very unlikely) mean “Zebra”: after all, I have confused other ‘Z’ words before!

What this amounts to is taking the conditional analysis I have given of ‘qualia’ and reading it as a kind of metasemantic account of what the semantics of ‘qualia’, ‘Zombie’, and so on would be in various circumstances.

There are reasons in general why I think A-intensions deserve to be regarded as giving the meanings of terms. After all, we are often ignorant of how the actual world is (consider our state in respect of ‘water’ in the fifteenth century) and yet manage to communicate something, and be—in some sense of competent—competent with the term.35 I take it that if this conditional is right about what we would say and how we would behave on different suppositions about what is actual, it goes some of the way to explicating the competence we have when we are ignorant of exactly how things are, and what we communicate when we do communicate in such cases. While ignorant of the nature of water, when I ask for the water to be passed, it comes my way, even if my interlocutor is equally ignorant. Supposing that ‘water is the actual watery stuff’ exhausts their grasp of the term, explains the otherwise mysterious arrival of the water. Supposing that

35 Jackson, “Why We Need A-intensions” (forthcoming).
the semantics are exhausted by water's connection to \( \text{H}_2\text{O} \) leaves that utterly mysterious. Now, it might be that there are historical contingencies. Perhaps if the Saxons had never invaded Britain, 'water' would have meant "foul foreign drink." But this is genuinely a mere metasemantic claim, for my use of the term does not depend on my opinions about British history. And similarly, the kinds of historical influences that could have made 'Zombie' mean "Zebra" are ones that do not regulate my use.

V.2. A further objection: the wrong conceivability has been explained. Chalmers has pressed a further objection. I have explained why it is conceivable that it is possible that there could be a world that has physical states of the required functional kind and no spooky states, and yet lacks qualia. But Chalmers thinks explaining this conceivability does only half the job. He thinks that what is equally important is that it is directly conceivable (and thus epistemically possible) that there are physical states of the right functional kind, no spooky states, and no qualia. In other words, for all that we know a priori, things might be such that there are functional states and no qualia.

This is, however, exactly what I deny. The conclusion in the section above is that the epistemic gap between physical states and qualia is a shadow of the conceivability of the possibility. It is not directly conceivable that there be functional states in the world, no spooky states, and no qualia, in part because on my view it is inconceivable that physical states could actually fail to be qualia if there are no spooky states. It seems to me that this is not a technical by-product of my account, but accounts for widely shared intuitions.

V.3. The argument from mere causal shift of meaning. There is another complaint of the same ilk. Presumably it is possible that there is some strange empirical fluke which ensures that the discovery that physicalism is false, say, would make me shift meaning so as to adopt a new concept 'qualia' and thus 'Zombie', such that it was right to say 'Zombies are possible'. Perhaps this has happened in history. Perhaps 'water' really does mean \( \text{H}_2\text{O} \) now, where it did not before. The discovery that water is \( \text{H}_2\text{O} \) has changed our concept due to the high salience of chemical nature. The objector goes on to say that we cannot tell the difference between the concept being the conditional one, and a metasemantic conditional changing the concept. Do we have a conditional concept, or a conditional about concepts?

But there is a difference, I think. There may well be all kinds of empirical discoveries that could change our conceptual palette. Sometimes we do change concept, and retain orthography, and maybe the concept 'water' is not the same as the concept that was around before the periodic table. But there is a test, at least in principle. Has
history changed our concept, or are we still under the governance of
the $A$-intension given by ‘water is the actual watery stuff’? If the former,
then ‘water’ just means $H_2O$, the concept has changed, and if we
discover that actually there is no $H_2O$, we will have discovered that
there is no water, and that is the claim we will make. I will mention
only the beginnings of an account of governance here, but the idea
will be that governance holds if after having changed, we are still
disposed to change back if we revise back our empirical opinion. If
the change would not be revised back should our empirical opinions
change back, then governance does not hold. If we are still under
the governance of the old $A$-intension, then discovering that there is
no $H_2O$, but that (surprisingly!) the watery stuff is $XYZ$ will make us
claim that we were wrong about water being $H_2O$—in fact, it is $XYZ$.
Further changes of view about the actual world—back to belief that
the watery stuff is $H_2O$ or on to other candidates—will always result
in the claim that water is whatever we take to be the (actual) watery
stuff. The difference between a metasemantic claim about what would
change a meaning or replace a concept, and a conditional semantic
analysis is just this kind of continued governance. And the fact that
I take it that we (most of us) hold fixed that we have phenomenal
experiences—if I become persuaded of dualism I think they are dualis-
tic, if I lose the faith I decide they are physical, if I regain the faith
I think they are dualistic, and so on—is what supports the idea that
the conceptual structure is conditional. It is not that were I to become
persuaded of dualism that I would change my concept, for I remain
sensitive to the claim about the actual world, and would revise back
if I became a physicalist. Of course, one might have a different concept
from most of us, which was merely metasemantically conditional.
Someone who was a physicalist about qualia, but on the discovery
that dualism was true became a dualist about qualia, and then, on
returning to his view that there were no spooky states, became not a
physicalist about qualia again, but rather an eliminativist, would per-
haps have such a concept.

VI. WHAT KIND OF ANALYSIS?
The benefit of the story so far is that it accounts for the intuition
about the possibility of Zombies, which a straightforward analytical
functionalism as a theory of the meaning of qualia language could not.

The difficulty is that the story is still intended to be a kind of
analysis: the conditional is in some sense a priori and analytic. But it
might not sound like it is something that is a plausible story in pure
analysis. There are two methods that we have come to expect might
deliver meanings. The first is empirical inquiry—the way discoveries
about the actual nature of water guide us to its \((C)\)-intension. The
second is introspection—the way we figure out that bachelors are
unmarried by consulting our own understanding, or the understand-
ing of others if we suspect that our own is deficient.

The conditional analysis I have proposed is not supposed to be a
posteriori. We do not look at the nature of actual qualia to find that
the conditional analysis is true. It is, after all, an analysis.

But if it is right, neither is it based (solely) on self-evident introspec-
tion. So how can it be analytic? What might be the basis for the claim
that something like this analysis is true?

The answer is that it describes how agents seem actually to behave.
Many who in fact believe that phenomenal experiences give acquain-
tance with the essential nature of nonphysical properties or substances
believe that phenomenal experience is essentially this. But in fact few
people who once believed in such properties or substances, but be-
came converts to physicalism, are in fact error theorists about phenom-
enal experience—rather, they flocked to naturalistic substitutes. Per-
haps some of those who now believe in the dualistic properties would
come to be error theorists, but it seems unlikely because the notion
of phenomenal experience plays an indispensable role in our everyday
communication about how the world is.

So what do we say to those who claim that not only do they believe
in nonphysical properties, but they regard them as necessary for
qualia? When they do the thought experiments, they insist, it is clear
that spookless worlds are Zombie worlds. The answer is that the sort
of evidence that underwrites these higher-order analyses is not only
thought experiments; or at least not thought experiments that can
be performed without particular experience. Thought experiments
only tell us what we \emph{think} we would say under certain circumstances.
Some people might in fact believe in the nonphysical properties, and
believe them to be essential to phenomenal experience. They might
also think that they would become error theorists about phenomenal
experience should they come to think, on metaphysical grounds, that
there are no nonphysical properties. This latter thought could be
one that they have access to via thought experiment: a kind of self-
simulation. They imagine that they have become persuaded of the
metaphysical error of dualism, and then wonder if this means they
would persist in ascribing the intension they currently ascribe to
“phenomenal properties,” and thus they become error theorists. They
do the thought experiment (roughly, the indicative version of the
counterfactual thought experiment) and conclude that this would be
the discovery that there would be nothing worth calling ‘qualia’.

But they might be mistaken about this. This is why, at the most
abstract level, thought experiments are an only fallible guide to analysis. The evidence, insofar as there is any, is that people are prone to change their views on the Gintension, rather than conclude that there was never anything that they were talking about when they used the word 'qualia'.

Why should this, if it is true, be victory to me? Why should I prioritize, in establishing the intension of a term for an agent, what he would in fact say in certain circumstances over the analysis he explicitly holds which is based on his best thought experiments concerning what he would say?

Something like a principle of deference to better-informed future time slices is what is at work here. If some future temporal stage of yourself were actually to find out that a certain metaphysics is true, and if there were no conspicuous loss of rationality attendant on this discovery, then your future self is better informed, that is, possessed of details of how things are that are elided in mere thought experiments. Why would it be reasonable to choose to go with opinions based on your pale attempts to self-simulate, rather than admit that while your self-simulations are the best guess you can give, what is right of you is determined by what you really would do when better informed. It is considerations like this that, I think, move those who think that when confronted by intelligent-seeming androids and so on, those in the thrall of John Searle would largely change their minds about the androids' unintelligence—and thus the term 'intelligent' in their mouths does not actually mean anything different from the term in the mouths of the AI community.

One last possibility remains. Perhaps an agent might not just have false beliefs about how he would behave in these circumstances. He might not care how he behaves or uses language in conditions of superior information. He might explicitly announce that the terms he uses definitely have the intentions he takes them to have, making it explicit that he is not bound by how his better-informed, and by some lights, reasonable future time slices might think or behave. Call this view analytical Muggletonianism. The Muggeltonians were a seventeenth-century English religious sect who believed that reason led to evil and error. Analytical Muggletonians do not defer to their

future better-informed or more reasonable time slices. For them, current thought experiments and so forth are the whole of what determines the intentions of their terms, and they regard their future time slices, insofar as they disagree with current opinion, as mistaken—or as having radically changed the subject without realizing it.

Could there be analytical Muggletonians? I am happy to accept that there could be. Others may make some constitutive claim that the Muggletonian, for all his self-binding stipulation, is just mistaken. I see no reason to deny that we can and do define things stipulatively. But to the extent that analytical Muggletonianism is mere stipulation it seems hardly a move that will help resist my analysis. For anyone who is not a Muggletonian and discusses qualia wants to be using the term that the rest of us use when ‘qualia’ proceeds from our mouths or keyboards. And to the extent that most of us find normative reason to defer to our better-informed future selves (or our better-informed colleagues in the present), we are using the term differently from the Muggletonian.

All of this may appear to paint a slightly different picture of the analytic than is traditional. The analytic truths supervene on psychology: that much is preserved. But they do not supervene on the immediately introspectible part of psychology, or even on rational progress from the introspectible parts of psychology. For no amount of armchair speculation can deliver the judgments that would in fact be delivered in real circumstances. My acceptance of the possibility of analytic Muggletonianism means, however, that there is still an introspectible component that governs meaning: it is the deference that most of us do have to future better-informed opinion. We defer to our own better-informed counterparts in much the same way as, on other matters, we defer to other, better-informed, individuals. It is in virtue of this that what we mean by ‘qualia’ is still, however indirectly, in our own control.

You do not get to be an analytical Muggletonian just because there are some better-informed future time slices to whom you do not defer, for it could be perfectly reasonable to be wary of the semantic intuitions of some future time slices of yourself overwhelmed by too much information: hence the ‘by some lights reasonable’ clause.

So it should be clear that nothing in what I say about the two-dimensional framework assumes that the framework itself removes the need for an account of what an analysis is. I agree with Block and Stalnaker (op. cit.) that two-dimensionalism only formalizes the conceptual component of meaning, it does not provide it. The claim that the conditional analysis of qualia I give here is correct is supposed to be independently justified by arguments like the ones in this section. The role of the two-dimensional apparatus in this paper has been to show how such an analysis undercuts the arguments from conceivability.
VII. SPECULATIVE METAPHYSICS

How do we determine the truth or falsity of physicalism? For this is something that we need to know if we are to determine the Gintension claims involving terms like ‘qualia’, or to establish the (superficial in Evans’s sense)39 possibility or otherwise of Zombies. Analysis is of no help with this task. Just as the analytical functionalist has to appeal to other considerations to determine whether the states that play the causal roles are physical or not, the kind of conditional functionalism I advocate here requires the metaphysics of the mind to be settled by some other method than analysis. Analysis comes after speculative metaphysics. There is an interesting question, perhaps outside the scope of this article, as to what kind of endeavor that is. For it does not seem to be the kind of thing we can settle a priori, and yet it seems to be beyond empirical confirmation as well. Still, this is not a problem that is local to philosophy of mind. Any robust realism allows that there are questions like this. Are there spatiotemporally and causally isolated parts of actuality? Experience will not settle it—since if there are, they are causally isolated. And yet surely it is not a priori either. Perhaps it is a posteriori for some ideal agents. The challenge for philosophers is why we think there are even plausibility arguments either way on such questions.

One final worry in this direction: What if I do think that the truth or otherwise of dualism is an a priori one? Suppose that my suspicion is that the idea of a property whose intrinsic nature is available to experience is somehow incoherent (and on alternate Tuesdays this seems right to me). Can I still hold an analysis of qualia which is conditional and that has claims about the truth or otherwise of physicalism in its antecedent?

I think so, but it will get messy. There is some good sense of epistemic possibility on which when I worry about the coherence of the spooky state, it is epistemically possible either way. But of course we cannot then regard the worlds considered as actual as being possible worlds simpliciter. For on this assumption, there may be no possible world with the spooky states. Thus, we will have to regard the rows in a two-dimensional matrix other than the row containing the actual world, as containing different sorts of items—not logically possible worlds, but epistemic possibilities of a certain kind.

VII.1. Qualia and consciousness. One final clarification: so far I have talked about qualia throughout. But the argument for dualism from the possibility of Zombies is often thought of as depending on issues

39 See Davies and Humberstone.
about consciousness. Elsewhere,\(^{40}\) I have argued that we get consciousness for free when we mix qualia with other mental states where the issue of physicalism is more easily settled. Anyone who accepts that can go away happy. Anyone who does not has an ‘elementary exercise for the reader’ to perform: she must reconstruct the story so far in terms of consciousness.

VIII. CONCLUSION

So a powerful argument for dualism is defused. The Zombie intuition can be taken as a guide to meaning, and yet not with the result that we must abandon physicalism. This is because the intuition is the product of a complex conditional analysis which I claim is true of ‘quale’, and when we see what part of the meaning of the term gives rise to the intuition, we see that it is metaphysically idle. Of course, this does not settle the issue of physicalism the other way. Speculative metaphysics lives. But many of us as are convinced as we ever were of the plausibility of physicalism. And on the conditional analysis that I offer, if we are right about this metaphysical claim, then a substantive functionalism is true about qualia: they just are states that realize certain functional roles.

DAVID BRADDON-MITCHELL

University of Sydney

\(^{40}\) Braddon-Mitchell and Jackson, \textit{op. cit.}, p. 135ff.