Can the Russellian monist escape the epiphenomenalist’s paradox?

Lok-Chi Chan

**Abstract:** Russellian monism – an influential doctrine proposed by Russell (1927/1992) – is roughly the view that physics can only ever tell us about the causal, dispositional, and structural properties of physical entities and not their categorical (or intrinsic) properties, whereas our qualia are constituted by those categorical properties. In this paper, I will discuss the relation between Russellian monism and a seminal paradox facing epiphenomenalism, the paradox of phenomenal judgment: if epiphenomenalism is true – qualia are causally inefficacious – then any judgment concerning qualia, including epiphenomenalism itself, cannot be caused by qualia. For many writers, including Hawthorne (2001), Smart (2004), and Braddon-Mitchell and Jackson (2007), Russellian monism faces the same paradox as epiphenomenalism does. I will assess Chalmers’s (1996) and Seager’s (2009) defences of Russellian monism against the paradox, and will put forward a novel argument against those defences.

**Keywords:** Russellian monism; epiphenomenalism; paradox of phenomenal judgment; qualia; zombies

**1 Introduction**

Russellian monism (henceforth, RM) is an influential doctrine in the philosophy of mind. A seminal formulation of it is provided by Bertrand Russell (1927/1992), after whom it is named; and it has recently attracted and been developed by many followers (see, for example, Chalmers 1996; Stoljar 2001, 2014; Heil 2003, 2004; Seager 2009). The doctrine consists of the following three components:

(1) **The scientific categorical ignorance thesis:** Categorical properties are not knowable by natural science.[[1]](#footnote-1)

(2) **The categorical qualia thesis:** Our qualitative experiences (i.e. qualia) are (solely or partially) constituted by their categorical bases (in some manner that is not [solely] via the causal, dispositional, and structural properties that are, in turn, borne by these bases).

(3) **The common bases thesis:** Properties studied by science and our qualitative experiences share the same kind(s) of categorical basis.

Following Rae Langton (1998, pp. 120n32), John Heil (2005, p. 343), and many others, I take it that dispositions and causal powers are exactly the same things, and thus that the two terms can be used interchangeably. The aim of this paper is to address a question concerning the powers of qualia as described by RM, and to thereby assess the plausibility of RM with regard to the topic of this special issue: mental powers. The question is whether qualia possess powers that enable us to know of them. As I address this question, I will also provide a brief assessment of whether RM provides a better metaphysics of powers than that of dualism. My answers to these questions are *no*.

Many influential writers, such as David Chalmers (1996), Daniel Stoljar (2014), John Heil (2004), and William Seager (2009), consider RM to be a significant improvement over dualism (see Section 5). This is because it is supposedly a kind of monism, and thus, as Stoljar remarks, ‘avoids the inelegance of dualism’ (2014, p. 19). I have argued elsewhere that we in fact have no need for non-physicalism or mysterianism – RM included – to explain qualia at all (Chan & Latham forthcoming), but it is nonetheless of genuine philosophical interest to see whether RM is really such an improvement over dualism. In this paper, I will compare RM with a particular version of dualism: epiphenomenalism. The discussion will focus on whether the Russellian monist (henceforth, the RMist) can escape a seminal paradox facing epiphenomenalism: the *paradox of phenomenal judgment*(henceforth, PPJ).[[2]](#footnote-2) More specifically, I will assess Chalmers’s (1996) and Seager’s (2009) defences of RM against the paradox, and will put forward a novel argument against those defences.

To tell the story of PPJ, we must begin with epiphenomenalism. Over the course of history, numerous versions of epiphenomenalism have been developed. I will focus on the version most contemporary philosophers have in mind when they use the term: the one Keith Campbell (1984) calls *new epiphenomenalism*. In this view, qualia are not physical entities; and physical states cause qualitative states but not *vice versa* (see Jackson 1982; Kim 2005; Robinson 2012). The doctrine is widely believed to have a self-defeating – and thus obviously implausible – nature, which is presumably one of the main reasons for its current unpopularity.

The worry is as follows. If epiphenomenalism is true, qualia have *no powers to cause any of our judgments* – which are made by our physical cognitive system and which we can express to others by physical means. Accordingly, our judgments concerning qualia – or, in Chalmers’s words, *phenomenal judgments* (1996, p. 177) – cannot be responses to (and cannot be explained by) qualia. Instead, these judgments must be fully and exclusively accounted for by physical facts. As Daniel Dennett points out, our zombie replicas would have made exactly the same judgments as we do (1991, p. 403). This is self-defeating, especially when epiphenomenalism itself is a phenomenal judgment: it is a theory of the phenomenal which we develop from our simpler phenomenal judgments, such as ‘I am experiencing a red quale’ and ‘a red quale has such and such nature which resists physical explanation’.[[3]](#footnote-3) While the term ‘the paradox of phenomenal judgment’ is offed by Chalmers (p. 177), the problem has been discussed by many others. For example, David Braddon-Mitchell and Frank Jackson call it the *epiphenomenal objection to epiphenomenalism*(2007, p. 13). One further seminal point made by these two authors is: the worry inevitably leads the epiphenomenalist to convert to physicalism. For within her metaphysical framework, her reason to posit anything beyond the physical *must be* fully and exclusively accounted for by the physical, even though she may not actually know how (pp. 141-142). (Interestingly, this is in fact a part of Jackson’s own intellectual journey.) Epiphenomenalists and their sympathisers have, of course, offered some replies to PPJ (see, for example, Chalmers 1996; Robinson 2012) – some of which may apply equally to my arguments in this paper – but for our purpose here I will assume that the paradox is fatal, and that those replies are unsuccessful.

Several authors, including John Hawthorne (2001), J. J. C. Smart (2004), and Braddon-Mitchell and Jackson (2007), use the same argument against RM, simply replacing the qualitative states described by the epiphenomenalist with the qualia-realising categorical properties described by the RMist. But Chalmers (1996) and Seager (2009), in response to the worry, argue that RM is not a version of epiphenomenalism. I will explain why I do *not* object to their arguments (Section 2). Nevertheless, I will provide a novel argument to show that regardless of whether it is a version of epiphenomenalism, it is facing PPJ just as epiphenomenalism is (Sections 3 and 4). Finally, I will argue that while many writers consider it to be a significant improvement over dualism or epiphenomenalism, its relation to PPJ shows that the improvement might be overrated (Section 5).

**2 RM and epiphenomenalism**

Before I develop my argument, we shall look closely at why some believe that RM is not a version of epiphenomenalism and is thus not facing PPJ or similar problems, even though categorical properties are, supposedly, non-dispositional and non-causal much like epiphenomenal qualia are.

Chalmers, seeing PPJ as a problem to be avoided because of its counterintuitiveness, considers RM to be one of the ‘strategies for avoiding epiphenomenalism [which faces PPJ]’ (1996, p. 151) and, in addition, regards it as ‘the strategy to which [he is] most drawn’ (p. 153). In his view, there is a causal relevance between physical causal properties and their categorical bases, even though it would be, in his words, ‘a subtler kind of causal relevance than the usual kind’ (p. 154). The causal relevance is as follows. A world of (physical) causal and dispositional properties without categorical bases is ‘arguably logically impossible’ (p. 154), because it is not very reasonable to believe that these properties could be ‘empty placeholders with no properties of their own’ (p. 153). Some readers might be reminded of a view in the metaphysics of properties called *categoricalism*, according to which all causal and dispositional properties have categorical bases. Indeed, Chalmers is restating the view and an influential argument in defence of it: the *empty space argument* (as I call it). According to the argument, dispositions without categorical bases to act as their ‘materials’ and to ‘fill in’ their spatial locations are nothing more than empty spaces, and hence such dispositions are logically or metaphysically impossible (see, for example, Armstrong 1961; Blackburn 1990; Langton 1998; see also Jackson 1998 for a related point). To summarise, in Chalmers’s view categorical bases are not epiphenomenal for the reason that they are acting as necessary bases for physical causation to instantiate, and are thus not metaphysically irrelevant to it. Let us call this argument the *argument from instantiation*.

Seager (2009) does not address PPJ specifically, but he has addressed a very similar and relevant concern: provided that the physical seems to be causally complete and closed, are qualia (as described by RM) epiphenomenal and how can they causally contribute to our physical world? Even though PPJ is not specifically mentioned in this problem, I suppose that the former is an indispensable part of the latter, and that any successful response to the latter must be capable of accounting for the former. The reason is simple: our knowledge of qualia is a part of our physical world! The response Seager offers to the causal closure problem is very similar to that of Chalmers. It relies on a view in the metaphysics of properties according to which the nature of a dispositional property is essentially determined by the nature of its categorical basis (see, for example, Oderberg 2009).[[4]](#footnote-4) Seager writes:

The fundamental dispositional properties of matter are just a reflection of its mentalistic intrinsic nature; calling them physical with the implication that they are entirely non-mental comes close to begging the question. (2009, pp. 217-218)

Seager is a panpsychist who accepts both RM and panpsychism, but I do not think that his argument is dependent on panpsychism. In my understanding, his view can be paraphrased as follows. The nature of a dispositional property is or determined by the nature of its categorical basis. Hence, the two kinds of properties have an equally intimate relation and are not fundamentally distinct.[[5]](#footnote-5) Therefore, physical properties and mental properties, which are identical to or supervene on the two kinds of properties respectively, are likewise not fundamentally distinct like the case of epiphenomenalism. Let us call this argument the *argument from intimacy*.

The strategies of the argument from instantiation and the argument from intimacy are the same: to say that there is some kind of dependent metaphysical relation between dispositional and categorical properties, and in particular, that they are categorical properties on which dispositional properties are dependent. I do not find these arguments to be objectionable, at least if their metaphysical assumptions – namely, the metaphysical theories of categorical properties mentioned above – are true. This reason is as follows. In my view, it is true that *non-*epiphenomenalism can be defined in a broad way, according to which the kind of property of concern – such as mental or categorical properties – has some *metaphysical and/or explanatory relevance* to physical properties and their causal powers, rather than the more restricted way that the kind of property of concern *must itself be or consist of* some powers to cause effectson physical properties. With this in mind, categoricalism and RM, as the two authors argue, indeed should not be categorised as versions of epiphenomenalism. Qualia as described by RM may, to a certain extent, be said as possessing powers. But I believe this is merely a matter of definition, which I do not consider to be the crucial matter of concern. What matters is not whether RM matches with the definition of epiphenomenalism, but whether it is facing PPJ like epiphenomenalism is. Chalmers and Seager seem to have an implicit assumption that a ‘no’ to the former question can lead to a ‘no’ to the latter question. I disagree. By providing a novel argument, I will spell out the mistake in that implicit assumption and show that the answer to the latter question is still ‘yes’.

**3 The zombie bases argument**

I call my argument the *zombie bases argument*. It is, much like Chalmers’s (1996) zombie argument, an argument from conceivability.[[6]](#footnote-6) However, unlike the zombie argument, which invites us to conceive of philosophical zombies – i.e. perfect physical duplicates of us who lack qualia – the concept of physical duplicates is not involved in my argument. What I invite the reader to conceive of are, instead, (1) a kind of categorical property that is without qualia-realising nature, i.e. is not qualitative and cannot constitute anything qualitative (in the way the RMist understands it) and (2) a possible world containing these properties.

Sentient beings that have these categorical properties as their bases are essentially lacking qualia, much like Chalmersian zombies are. Let us call these categorical properties *zombie bases*, as opposed to the kind described by RM, which we may call *qualitative bases*. The zombie bases argument invites us to conceive of the following possible world: a possible world *z* which is structurally and dynamically identical to our actual world, in the sense that it shares all the causal, dispositional, and structural properties; however, all of these properties are borne by zombie bases instead of their actual ones – which we are assuming to be qualitative bases.[[7]](#footnote-7) With this in mind, our counterparts in world *z* would all have made exactly the same phenomenal judgments (and, of course, judgments regarding all other matters) as we do. For instance, there would be an author called Chalmers who wrote a book called *The Conscious Mind*, and an author called Chan who wrote an article called ‘Can the Russellian monist escape the epiphenomenalist’s paradox?’. Nonetheless, they still do not have qualia because they are borne by zombie bases.

If world *z* is really possible, whether our categorical bases are qualitative bases or zombie bases makes no difference to our phenomenal judgments. Those judgments must thus be fully accounted for by non-categorical properties and the dependence of these properties on *any* kind of categorical basis, which world *z* contains. Hence, PPJ persists. Further, zombie bases in world *z* can do perfect well what Chalmers and Seager think our categorical bases do. Namely, the metaphysical and/or explanatory relevance between our non-categorical properties (such as causal, dispositional, and structural properties) and their categorical bases in world *z* is exactly the same as our actual world. Hence, such relevance is insufficient to account for our phenomenal judgments about qualia.

The zombie bases argument, if correct, shows that there must be at least one mistake or shortcoming in Chalmers’s and Seager’s arguments; but it does not straightforwardly tell us what the mistake or shortcoming really is. Nevertheless, it gives us a very good hint about how to locate the mistake, which I believe to be the confusion between the two questions below:

**The structural question:** Can qualitative bases contribute to a systematic explanation of the fact that our brain system can output information about our qualia?

**The tracking question:** Can our brain system output information about our qualia –

which we are assuming to be constituted by our categorical bases – by a process that *tracks* them?[[8]](#footnote-8)

The dependence of dispositions on their categorical bases, as Chalmers and Seager point out, contributes to our answer to the structural question, in exactly the same way that it contributes to our inquiries into the structure of all physical objects, functions, and states of affairs – which include a chair, my Microsoft Word, my view that RM is false, the eliminativist’s view that there are no qualia, and so forth. This does, indeed, distinguish RM from epiphenomenalism. Unfortunately, what the zombie bases argument shows us is that this alone cannot be a *tracking relation* that can help us to answer the tracking question.

What is tracking? The standard answer is: when an information-generating process tracks, it can output information that adequately corresponds to and represents the facts of concern. This answer is, of course, very abstract and imprecise, since what actually constitutes a tracking relation remains an open question in the literature. A seminal conception offered by Robert Nozick concerns *counterfactual sensitivity*: a tracking process should guarantee that if p were true, then S would believe that p, and that if p were not true, then S would not believe p (1981, p. 172).[[9]](#footnote-9) While this conception faces a severe problem when applied to necessary facts such as mathematical truths – for they cannot possibly be counterfactually false – it works pretty well with contingent facts. Under this conception, the zombie bases argument clearly demonstrates that RM faces the problem of tracking qualia, for our phenomenal judgments are not sensitive to the difference between qualitative bases and zombie bases: if our bases were zombie bases, the RMist would nevertheless believe in her view!

However, alternative conceptions of tracking are possible. We can also follow Gilbert Harman (1986, p. 63) and Richard Joyce (2016, p. 148) and ask for some precise nomological or epistemological account(s) of *how* p could help explain S’s belief of p: we can say that the information-generating process tracks p if a reasonable account is true. Under this conception, the measure of reasonableness of an account may be a pluralistic matter: it involves the adequacy of observation methods, deduction, induction, abduction, scientific modelling, our causal connections with the fact concerned, our evolutionary capacities for dealing with the fact concerned, the designs of our instruments, and so on and so forth. A lot of these concepts are fairly primitive, and hence a single, unified measure of their efficacies is not obviously possible. Despite the possible lack of such a single, unified measure or standard, we can nonetheless say that RM can hardly offer any precise nomological or epistemological account of the kind Harman and Joyce ask for. For the nature of the doctrine in effect eliminates all serious possibilities of such an account with a vengeance. On the one hand, the doctrine, by definition, does not allow any kind of tracking via causal, dispositional, or structural means. On the other hand, Chalmers’s argument from instantiation and Seager’s argument from intimacy have successfully informed us of some possible metaphysical and/or explanatory relevance of qualitative bases; and explanatory relevance might, indeed, be counted as a tracking relation. However, the zombie bases argument shows that what they miss out: what we actually track by these kinds of relevance – just as the works of David Armstrong, Simon Blackburn, Rae Langton, Frank Jackson, and others tell us – is merely the fact that we are borne by *some* categorical bases; we still cannot track the *qualitative nature* of those bases (if there is any).

To conclude, while qualia by the lights of RM possess powers, the zombie bases argument reveals that they and their powers are not connected in a way that enables us to track them – that is, to adequately generate information of them. Chalmers and Seager either have the confusion that the metaphysical and/or explanatory relevance they describe is alone sufficient to explain informational tracking, or have missed the point that what is in need of explanation is informational tracking. Either way, their arguments cannot offer RM any help in solving PPJ.

**4 Further considerations and possible objections**

I will discuss four possible objections, and by doing so further develop the argument I have presented so far.

4.1. The objection from metaphysical necessity

According to some theories of categorical properties, there is a metaphysically necessary relation between dispositional properties and their categorical bases, which do not allow a dispositional property to be multiply realised by different categorical bases (see, for example, Heil 2005). In this case, world *z* and zombie bases bearing our dispositional properties would be metaphysically impossible. For example, in response to Chalmers’s zombie argument, Heil remarks, ‘Agents or systems possessing identical powers must be qualitatively identical as well’ (2003, p. 247).

My reply

It is difficult, at least for me, to see why our dispositional properties cannot be borne by different categorical bases. Jackson (1998, pp. 23-24) has, for example, described a possibility that, even in our actual world, a nomological role posited by our scientific theories might in fact be played by two distinct categorical bases. This seems to be fairly appealing. After all, it is difficult to see why it is impossible for two different things to display the same pattern of power and behaviour. Heil famously holds the view that dispositional properties are *identical with* their categorical bases because the latter are the ‘qualities’ of the former.[[10]](#footnote-10) But even he somehow describes dispositional properties and their ‘qualities’ as if they are taking turns to manifest in different situations – such as scientific measurements and our qualitative experiences (see Heil 2003). This seems to imply that those dispositional properties and qualities could be combined in some different ways – even though he himself may not agree.[[11]](#footnote-11)  With this in mind, the metaphysical necessity view is not easy to justify; not even many RMists would find it attractive. For example, Chalmers in fact makes it clear that zombie bases are possible (1996, p. 154). He simply does not discuss them in detail, because he does not consider the possibility that they may lead to a serious objection to RM or his argument from instantiation.

Perhaps there can be a version of the objection from metaphysical necessity that is less radical than Heil’s and is thus appealing to more people. The idea is that there is a metaphysically necessary relation between *the* *possession of some qualia-realising nature* and *the* *possession of some disposition-realising nature* by a categorical property. Hence, whether or not dispositional properties are realisable by multiple categorical bases – such as those that might lead to, say, an inverted spectrum – they are not realisable by zombie bases. In this case, the metaphysically necessary relation is either (1) brute and inexplicable or (2) explainable by some further fact. I will show that both are problematic.

(1) In the former case, it is certainly true that a brute and inexplicable metaphysically necessary relation easily blocks the metaphysical possibility of world *z* and zombie bases bearing our dispositional properties. However, first of all, the proposal that brute and inexplicable metaphysically necessary relations exist is itself puzzling and controversial. This is because it is in fact an appeal to mysticism: it is an easy – rather too easy – response to so many philosophical arguments, while the nature of any such relation by definition cannot ever be precisely described. The worry is that the existence of such mysterious things is simply a made-up posit for an *ad hoc* purpose. Secondly, and more importantly, it seems that the absence of such a metaphysically necessary relation is nonetheless conceivable, though metaphysically impossible. With this in mind, regardless of whether world *z* and zombie bases bearing our dispositional properties are metaphysically possible, they are likewise conceivable. In this case, conceivable world *z* can still be considered atheoretical model. Call it model *z1*. Now the zombie bases argument works again. Compare the model *z1* with the theoretical description of the actual world by the RMist *prior to* the addition of the brute and inexplicable metaphysically necessary relation. In terms of their content, they are not only structurally and dynamically alike, but also are not different in their descriptions of the epistemic structures of their habitants. Hence, model *z1* clearly shows us that the RMist, *prior to* the addition of the metaphysically necessary relation, has not provided any idea of tracking relation that can solve PPJ. No doubt, now the RMist could add in the brute and inexplicable metaphysically necessary relation to refute model *z1*, but the real problem is: can this mysterious relation really serve as an *informational* tracking relation? It is very difficult to see how it could be so. For me, this is in fact fairly unintelligible.

(2) On the other hand, if it is expected that the metaphysically necessary relation between processing a qualia-realising nature and processing a disposition-realising nature can be explained by some further fact, then I believe it is either (i) unlikely to be true or (ii) likely to be damaging to RM. This is because there will be an *explanatory gap* between the (possession of) disposition-realising nature and the (possession of) qualia-realising nature which is very similar to the one faced by the physicalist. The idea of the explanatory gap, as it is seminally introduced by Joseph Levine (1983), is as follows. While physical properties are always causal in nature, qualia such as pain and the experience of redness have a non-causal nature: the way the red quale feels (p. 357). Accordingly, the causal stories we can have by taking physical properties into account cannot tell why qualia feel the way they do. Even though Levine merely intends to point out that the explanatory gap leaves us inevitably agnostic about which particular mind-body theory is true (p. 359), many take the further step to reach the metaphysical conclusion that physicalism is false. For example, Chalmers develops the explanatory gap into a metaphysical principle which Stoljar calls *from-structure-only-structure*, according to which ‘from structure and dynamics, we can only get more structure and dynamics’ (Chalmers 1996, p. 118; Stoljar 2014, p. 29). According to Chalmers, all of the major arguments against physicalism, including his zombie argument and the Mary’s room argument, start by establishing it (2003, p. 107).

I take it that the explanatory gap applies equally to the metaphysically necessary relations between the disposition-realising nature and the qualia-realising nature of a categorical property, and between the possession of the two natures by the property: if these relations are considered to be explainable by some further fact instead of being brute and inexplicable, it is difficult to see what kind of fact can bridge the gap between disposition-related facts and qualia-related facts. Returning to my point that the further fact view is either (i) unlikely to be true or (ii) likely to be damaging to RM, the reason is as follows. (i) On the one hand, it is difficult to see how the gap can possibly be bridged. (ii) On the other hand, even if the gap can be bridged in some peculiar ways (though I take it to be very unlikely), it is probably bad news rather than good news for the RMist. This is because Chalmers’s from-structure-only-structure principle then collapses: namely, the gap between qualia and structure and dynamics is now not considered as, as Levine and Chalmers take it to be, in principle unbridgeable. The RMist will then face the risk that she might lose her motivation to get beyond physicalism and appeal to RM to explain qualia in in the first place.[[12]](#footnote-12)

In sum, appealing to metaphysically necessary relation between the possession of some qualia-realising nature and some disposition-realising nature to resist the zombie bases argument will lead one to fall into a *trilemma*. (1) If the necessity is brute and inexplicable, then it cannot really reject the conclusions of the zombie bases argument and PPJ. (2i) Taking the explanatory gap into account, if the necessity is considered to be grounded on some further fact, then it is difficult to see how there can possibly be such a fact. (2ii) Even if there can possibly be such a fact, it would then follow that the explanatory gap is not in principle unbridgeable, and this may undercut our motivation to go beyond physicalism and accept RM.

4.2. The weak conceivability objection

*The weak conceivability objection* can be developed from Stoljar’s work (2001), even though Stoljar does not have the zombie bases argument in mind. What he has in mind is, instead, the worry about the possibility of a peculiar kind of zombie that does not possess only our physical structure and dynamics but our categorical bases as well. In response to the worry, he argues that we cannot ‘truthfully be said to strongly conceive the possibility of something which is [categorically] identical’ (2001, p. 401). He explains, this is because, according to many metaphysicians, we are *irremediably ignorant* of categorical properties (see, for example, Armstrong 1961; Blackburn 1990; Langton 1998; Jackson 1998; Lewis 2009; Locke 2009).[[13]](#footnote-13) The point is: only strong conceivability – which is derived from a sufficiently informed and rational reasoning process – reveals possibility, while weak conceivability – which is derived from an insufficiently informed or rational reasoning process – does not; but it is widely accepted that we are never informed about categorical properties. If this is correct, we are not in a position to strongly conceive of the possibilities concerning categorical properties, such as the possibility of zombie bases. Accordingly, we should not assume that world *z* and zombie bases are possible.

My reply

A problem with this objection is that it is too strong. We rarely think about possibilities and necessities by taking full and comprehensive knowledge of the matter of concern into account. As Chalmers remarks, it is *ideal* conceivability – which is derived from an all-informed, perfectly rational reasoning process – that entails possibility (2002, p. 204), but ‘it is not obvious that an ideal reasoner is possible or coherent’ (p. 148). If we were to accept the fact that our (practically possible) conceivability does not imply possibility, then there would be a disastrous overkill. We may, suddenly, lose our strongest arguments against physicalism and thus have no reason to find RM attractive. As Chalmers notes, all major arguments against physicalism begin with assuming the explanatory gap (2003, p. 107); and the explanatory gap, as I understand it, rests on the inconceivability and thus impossibility that features of the physical and qualia can be bridged. If conceivability does not imply possibility, then it seems that we should just agree with Patricia Churchland (1996) that the gap is merely a matter of our current ignorance and nothing metaphysical: given the recent progress of neuroscience, it is difficult to see why we should not expect that neuroscience will eventually provide the solution (see Hohwy & Frith 2004; Chan & Latham forthcoming).

Chalmers provides a convincing proposal that allows us to attain – or, in my view, attain something closer to – ideal conceivability. The proposal is that we may have a justification of the expectation that a belief is undefeatable by better reasoning. As he puts it:

[O]ne can […] simply invoke the notion of undefeatability by better reasoning. […] We can say that S is ideally conceivable when there is a possible subject for whom S is prima facie conceivable, with justification that is undefeatable by better reasoning. The idea is that when prima facie conceivability falls short of ideal conceivability, then the claim that the relevant tests are passed will either be unjustified, or the justification will be defeatable by further reasoning. For ideal conceivability, one needs justification that cannot be rationally defeated. (2002, p. 148)

This line of reasoning reflects what he has (or tacitly has) in mind in his works on the conceivability argument. For example, he writes:

[W]e do not need to imagine each of the neurons to make the case. Mere complexity among neurons could not conceptually entail consciousness; if all that neural structure is to be relevant to consciousness, it must be relevant in virtue of some higher-level properties that it enables. So it is enough to imagine the system at a coarse level. […] No matter how sophisticated we imagine these mechanisms to be, the zombie scenario remains as coherent as ever. (1996, p. 98)

We can see that Chalmers assumes the from-structure-only-structure principle we discussed earlier. I believe that his line of reasoning is as follows. Our currently unbridgeable explanatory gap relates to *a principle*, which is the from-structure-only-structure principle. If something is a principle, then it is undefeatable. Hence, the from-structure-only-structure principle, in turn, implies ideal conceivability of zombies, which, in turn, implies the possibility of zombies (see also Chalmers 2003, p. 107). The crucial step here is *to discover some principle(s).*

Let us return to the conceivability of world *z* and zombie bases bearing our dispositional properties. With Chalmers’s line of reasoning, to guarantee their possibility, we do not need to have perfect knowledge about categorical properties; we only need to discover some principle(s) with regard to their nature. Of course, the matter of concern, as Stoljar points out, is whether our irremediable ignorance of categorical properties allows us to discover it. But I think that the answer is obviously yes. The irremediable ignorance of categorical properties, if true, only prevents us from knowing their exact identity or what they are precisely like (see Lewis 2009; Locke 2009). Metaphysical issues of them such as their realisation relation, identity relation, and so forth are more abstract. We can still assess which kinds are plausible or implausible, in terms of whether they make sense in logic and metaphysics, whether they are compatible with natural science, and so on – after all, it would be amazing if we cannot even know that our categorical bases are self-identical or are not Mickey Mouse or square-circles. This allows us to posit some metaphysical principles with regard to the nature of categorical properties.

Now it should be clear how we can respond to the worry that zombie bases are not strongly conceivable. In Section 4.1, based on some general metaphysical principles, I argued that the possible scenarios in which zombie bases are metaphysically impossible force the RMist into a trilemma. Hence, one must either accept that world *z* is possible, or one must face the trilemma. Since this is based on some general metaphysical principles, this appears to be undefeatable by better reasoning or knowledge: *to think otherwise is inconceivable*. It is therefore inadequate to respond to the zombie bases argument by appealing to our irremediable ignorance of categorical properties.

4.3. The dynamic qualia objection

Recall the categorical qualia thesis, a component of RM, that our qualia are *solely or partially* constituted by their categorical bases. The *dynamic qualia objection* points out that qualia could be constituted partially by their categorical bases, *and* partially by the causal, dispositional, and structural properties borne by those bases. By positing those dispositional properties, not only categorical properties, as constituents of qualia, RM can attribute causal powers and thereby powers to be tracked to qualia.[[14]](#footnote-14)

My reply

We can understand the point of view of the RMist with more sympathy and agree that, from her point of view, for something to be called a quale, it might have two components: the *dynamic component* and the *categorical component.* The dynamic component can be fully and exclusively accounted for by physical facts and, in addition, causally efficacious to other physical entities, while the categorical component is non-physical and causally impotent. This is different from most physicalists who believe that the dynamic component alone (if there is any) is sufficient to be called a quale.

However, what actually follows is that our phenomenal judgments are still not sensitive to the difference between a qualitative categorical component and its zombie counterpart. As a result, our phenomenal judgments track only the dynamic component of qualia (and maybe the fact that it is borne by *some* categorical bases), and nonetheless fail to track the qualitative nature of the categorical component.

With this in mind, it also seems that Braddon-Mitchell and Jackson’s argument that the epiphenomenalist would inevitably be led to convert to physicalism, which we have discussed earlier, applies perfectly to this view. If its advocate wants the nature of qualia to be a part of the story about how we come to have our beliefs about and theories of qualia, she would be forced to expect that it is solely the work of the dynamic component. Unfortunately, this move inevitably converts her to physicalism. For the nature of qualia must now be fully and exclusively accounted for by the physical properties that constitute the dynamic component, even though we might not yet know what the account is. The categorical component, then, must play no role at all, with the exception of bearing those physical properties. To say that its qualitative nature is a necessary component of qualia commits oneself to a useless ‘nomological dangler’, to use Smart’s famous phrase (1959, p. 142), just as the epiphenomenalist does.

*4.4. The ‘this is just another radical scepticism’ objection*

Unlike the previous objections which target the metaphysical assumptions of the zombie bases argument, the ‘this is just another radical scepticism’ objection targets an epistemological assumption of the argument, according to which insensitivity to the possibility of swapping entities, properties, or states of affairs implies a failure of informational tracking. The worry is that the assumption might be too strong and is one shared by many arguments for radical scepticism, including the brain in a vat argument, the Cartesian demon argument, and so on. For many, the brain in a vat story is entirely possible; and if the story were true, we would nevertheless have the same beliefs about our world. Nonetheless, it is absurd to think that the possibility of this kind of story demonstrates that our beliefs about our world do not track. According to the very same epistemic standard, we should not think that the possibility of zombie bases shows that we cannot track qualia if RM is true.

My reply

This objection appeals to an analogy between the zombie bases argument and some of the arguments for radical scepticism, but the analogy is not obviously successful. The problem is as follows. We should not assume that the conclusion of the brain in a vat argument, or arguments alike, is absurd simply because it would render most of our beliefs unjustified. Making such an assumption seems to be nothing more than a wishful thinking. The correct way we should respond to those arguments is: to offer good objections to them. Accordingly, it can be questioned whether those good objections apply equally to the zombie bases argument as well; and the answer is very likely to be ‘No!’.

There are many possible objections to sceptical arguments that appeal to stories like the brain in a vat. Seminal ones include that the brain in a vat story violates the principle of simplicity, that we can be justified in believing our intuitions unless they are proved false by evidence, and so forth. A detailed consideration of every objection is impossible in this paper. Nevertheless, a very brief overall analysis can be offered. If, on the one hand, we, in response to radical scepticism, appeal to the difference between the *causal stories* of the brain in a vat story and the non-sceptical story, the analogy between the zombie bases argument and radical scepticism then collapses. This is because the difference between qualitative and zombie bases cannot ever be causal ones. If, on the other hand, we, in response to radical scepticism, appeal to the reliability of our intuitions, the analogy between the zombie bases argument and radical scepticism again collapses. This is because RM is not intuitively true: it is a sophisticated metaphysical hypothesis used to explain our intuitions about qualia, but is not itself an intuition about qualia. In any case, in order to assume that the analogy between the zombie bases argument and radical scepticism is successful, we must carefully assess each possible objection to radical scepticism. Even though I cannot offer such a detailed assessment in this paper, I have provided some reasons for believing that many of those objections may not apply to the zombie bases argument.

5 Conclusion

The implications of the zombie bases argument are simple and straightforward: that all of the implications of PPJ to epiphenomenalism apply equally to RM as well, regardless of whether or not RM is a version of epiphenomenalism, and that Chalmers’s and Seager’s arguments against this fail. In fact, when Braddon-Mitchell and Jackson put forward their version of PPJ and argue that it inevitably leads the epiphenomenalist to convert to physicalism, they also have RM in mind (2007, pp. 140-141). The difference between their argument and my argument here is that they take for granted that RM to be a version of epiphenomenalism, while I avoid this assumption, and critically assess – and even accept – Chalmers’s and Seager’s arguments that RM is not a version of epiphenomenalism. Nonetheless, I reached the same conclusion as Braddon-Mitchell and Jackson do.

Let us return to the topic of this special issue: mental powers. Ever since the debate between Descartes and Princess Elisabeth of Bohemia, the interaction problem – namely, how the mental and the physical could ever have powers to causally interact – has been the most difficult problem facing dualism. Interactionist dualism can hardly offer a solution that is not in conflict with our neuroscience and the causal closure of the physical. Epiphenomenal dualism deprives the mental of its causal powers, and thereby retains its consistency with our neuroscience and the causal closure of the physical. But it then faces a far worse problem: PPJ, and can therefore be said as a step backward rather than forward. RM has recently attracted attention in the literature, and is considered by many influential writers to be a significant improvement over dualism or epiphenomenalism. Stoljar, for example, believes that it is a view that can preserve the ‘spirit and structure’ of physicalism and ‘so avoids the inelegance of dualism’ (2014, p. 19). Heil believes that RM abandons the mental-material distinction of dualism without falling into functionalism, which cannot take qualia seriously (2004, pp. 239-240). Seager believes that RM is one of the most prominent alternatives to dualism, the ‘deeply unsatisfactory’ doctrine that commits to a ‘totally mysterious causal interaction between completely disparate substances’ (2009, p. 208). Chalmers, seeing PPJ as a problem to be avoided because of its counterintuitiveness, considers RM to be one of the ‘strategies for avoiding epiphenomenalism’ (1996, p. 151) and ‘the strategy to which [he is] most drawn’ (p. 153). However, the zombie bases argument shows that RM, in facing the interaction problem, is just as bad as epiphenomenalism is. While qualia under this view can be said to have causal powers, this is not at all useful – for the view cannot escape PPJ. With this in mind, RM is, contrary to the beliefs of many, as untenable and inelegant as dualism is. Its improvement over dualism might be very much overrated.

**Acknowledgments**

The author would like to acknowledge David Braddon-Mitchell, Belinda Calderone, Michael Duncan, Frank Jackson, Justine Kingsbury, Andrew James Latham, James Norton, Luke Roelofs, Daniel Stoljar, Lei Zhong, the philosophy of mind group at the Australian National University, and participants in my presentations of this paper at the University of Sydney and the National Taiwan University for their useful discussions or comments. Special thanks are due to Belinda Rickard.

**References:**

Armstrong, D 1961, *Perception and the physical world*, Routledge, London.

Blackburn, S 1990, ‘Filling in space’, *Analysis*, vol.50, no. 2, pp. 62-65.

Braddon-Mitchell, D & Jackson, F 2007, *The philosophy of mind and cognition*, 2nd edn, Blackwell, Malden.

Campbell, K 1984, *Body and mind*, 2nd edn, University of Notre Dame Press, Notre Dame.

Chalmers, D 1996, *The conscious mind: in search of a fundamental theory*, Oxford University Press, New York.

–––– 2002, ‘Does conceivability entail possibility?’, in TS Gendler & J Hawthorne (eds.), *Conceivability and possibility*, Oxford University Press, Oxford, pp. 145-200.

–––– 2003, ‘Consciousness and its place in nature’, in DS Stich & F Warfield (eds.), *The Blackwell guide to philosophy of mind*, Blackwell Publishing, Malden, pp. 102-142.

Chan, LC & Latham AJ (Forthcoming), ‘Four Meta-methods for the Study of Qualia’, *Erkenntnis*.

Churchland, PS 1996, ‘The hornswoggle problem’, *Journal of Consciousness Studies*, vol. 3, no. 5-6, pp. 402-408.

Dennett, D 1991, *Consciousness explained*, Back Bay Books, Boston.

Harman, G 1986 ‘Moral explanations of natural facts – can moral claims be tested against moral reality?’, *Southern Journal of Philosophy*, vol. 24, pp. 57-68.

Hawthorne, J 2001, ‘Causal structuralism’, *Philosophical Perspectives*, vol. 15, pp. 361-378.

Heil, J 2003, *From an Ontological Point of View*, Oxford University Press, Oxford.

–––– 2004, *Philosophy of mind*, 2nd edn, Routledge, New York.

–––– 2005, ‘Dispositions’, *Synthese*, vol. 144, no. 3, pp. 343-356.

Hohwy, J & Frith, C 2004, ‘Can neuroscience explain consciousness?’*, Journal of Consciousness Studies*, vol.11, no. 7/8, pp. 180-198.

Jackson, F 1982, ‘Epiphenomenal qualia’, *Philosophical Quarterly*, vol. 32, pp. 127-136.

–––– 1998, *From metaphysics to ethics: a defence of conceptual analysis*, Oxford University Press, Oxford.

Joyce, R 2016, ‘Evolution, truth-tracking, and moral skepticism’, in *Essays in moral skepticism*, Oxford University Press, Oxford, pp. 142.158

Kim, J 2005, *Physicalism, or something near enough*, Princeton University Press, Princeton.

Langton, R 1998, *Kantian humility: our ignorance of things in themselves*, Oxford University Press, Oxford.

Levine, J 1983, ‘Materialism and qualia: the explanatory gap’, *Pacific Philosophical Quarterly*,vol. 64, pp. 354-361.

Lewis, D 2009, ‘Ramseyan humility’, in D Braddon-Mitchell & R Nola (eds.), *Conceptual analysis and philosophical naturalism*, MIT Press, Cambridge, MA, pp. 203-222.

Locke, D 2009, ‘A partial defense of Ramseyan humility’, in D Braddon-Mitchell & R Nola (eds.), *Conceptual analysis and philosophical naturalism*, MIT Press, Cambridge, MA, pp. 223-242.

Nozick, R 1981, *Philosophical explanations*, Harvard University Press, Cambridge.

Oderberg, D 2009, ‘The non-identity of the categorical and the dispositional’, *Analysis,* vol. 69, no. 4, pp. 677-684.

Robinson, W 2012, ‘Phenomenal realist physicalism implies coherency of epiphenomenalist meaning’, The Journal of Consciousness Studies, vol. 19, no. 3-4, pp. 145–163.

–––– 2018, ‘Russellian monism and epiphenomenalism’, *Pacific Philosophical Quarterly*, vol. 99, no. 1, pp. 100-117.

Russell, B 1992/1927, *The analysis of matter*, Routledge, London.

Seager, W 2009, ‘Panpsychism’, in A Beckermann & BP McLaughlin (eds.), *The Oxford handbook of philosophy of mind*, Oxford University Press, New York, pp. 206-219.

Smart, JJC 1959, ‘Sensations and brain processes’, *Philosophical Review*, vol. 68, no. 2, pp. 141-156.

–––– 2004, ‘Consciousness and awareness’, *Journal of Consciousness Studies*, vol. 11, no. 2, pp. 41-50.

–––– 2007, ‘The Mind/brain identity theory’, in E Zalta (ed.), *Stanford encyclopedia of philosophy*, retrieved 25 October 2017, <http://plato.stanford.edu/entries/mind-identity/>

Stoljar, D 2001, ‘The conceivability argument and two conceptions of the physical’, *Philosophical Perspectives*, vol. 15, no. s15, pp. 393-413.

–––– 2014, ‘Four kinds of Russellian monism’, in U Kriegel (ed.), *Current controversies in philosophy of mind*, Routledge, New York, pp. 17-39.

1. Some philosophers call these properties ‘intrinsic properties’ instead. That does not affect the structure of RM and the arguments I will discuss. [↑](#footnote-ref-1)
2. William Robinson (2018) provides another interesting comparison between RM and epiphenomenalism. While my discussion is compatible with and potentially relevant to his, his concern is not PPJ. [↑](#footnote-ref-2)
3. See Chalmers (1996) for a detailed overview. [↑](#footnote-ref-3)
4. A rival view is that the connections between categorical and dispositional properties are contingent: it is possible for the same categorical property to play different dispositional roles under different laws of nature (see, for example, Langton 1998; Lewis 2009). [↑](#footnote-ref-4)
5. Another possible interpretation is that the two kinds of properties are *identical*. See Section 4.1 for a discussion of this view. [↑](#footnote-ref-5)
6. I assume that most readers are familiar with Chalmers’s zombie argument and thus will not introduce it here. [↑](#footnote-ref-6)
7. Our actual categorical bases are assumed not to be zombie bases for the sake of the argument, because we are assuming that RM is true. However, without the assumption that RM is true, it is possible that our actual categorical bases are zombie bases. [↑](#footnote-ref-7)
8. I am heavily indebted to Frank Jackson for a discussion with respect to this point. [↑](#footnote-ref-8)
9. While Nozick’s account is seminal, it has some more developed variants in the literature which take probability, distance between possible worlds, and so on into account. For the sake of simplicity, however, I shall not discuss them. [↑](#footnote-ref-9)
10. Heil’s identity theory does not allow for multiple realisability and is thus much stronger than the identity theory used by Smart (2007) to explain the mind/body relation. [↑](#footnote-ref-10)
11. It is noteworthy that, in a seminal paper on metaphysics (2005), Heil provides some independent reasons for rejecting the idea of multiple realisability. With this in mind, he might reasonably reject the implication. Nonetheless, the rejection of multiple realisability is, doubtlessly, a very radical view which very few would accept. [↑](#footnote-ref-11)
12. Of course, we may not know exactly how to bridge the gap, but a view widely accepted among scientists and physicalist philosophers is: contemporary brain sciences lead us to expect that advances in those disciplines will eventually reveal the answer to the question, and to thereby believe that physicalism is true (see, for example, Churchland 1996; Hohwy & Frith 2004; Chan & Latham forthcoming). [↑](#footnote-ref-12)
13. This view is in fact very similar to an assumption of RM: the scientific categorical ignorance thesis (see Section 1). [↑](#footnote-ref-13)
14. I am indebted to Daniel Stoljar for raising this objection. [↑](#footnote-ref-14)