Philos Stud
DOI 10.1007/s11098-008-9322-1

Swamp Mary's revenge: deviant phenomenal knowledge and physicalism

Pete Mandik

Abstract Deviant phenomenal knowledge is knowing what it's like to have experiences of, e.g., red without actually having had experiences of red. Such a knower is a deviant. Some physicalists have argued and some anti-physicalists have denied that the possibility of deviants undermines anti-physicalism and the Knowledge Argument. The current paper presents new arguments defending the deviant-based attacks on anti-physicalism. Central to my arguments are considerations concerning the psychosemantic underpinnings of deviant phenomenal knowledge. I argue that physicalists are in a superior position to account for the conditions in virtue of which states of deviants constitute representations of phenomenal facts.

Keywords Phenomenal knowledge · Physicalism · Knowledge argument · Psychosemantics

1 Introduction

Typically, people posses the phenomenal knowledge of what it's like to see red partially in virtue of their having seen red before. Following Alter (2008), let's call knowledge of what it's like to experience red possessed by someone who has never experienced (seen, after-imaged, hallucinated, or even imagined) red *deviant phenomenal knowledge*.[1] Let us call such a possessor a *deviant*.

[1] It will be useful, for brevity's sake, to count hallucinations and afterimages of red as episodes of seeing red.

P. Mandik (✉)
William Paterson University, 300 Pompton Road, Wayne, NJ 07470, USA
e-mail: mandikp@wpunj.edu

Published online: 10 December 2008

The majority opinion regarding Jackson's (1982) Mary is that she is not a deviant: despite her exhaustive physical knowledge of the neural underpinnings of red experience, she gains phenomenal knowledge of red experience only when or after she has red experience. This majority includes certain anti-physicalists as well as certain physicalists. The included anti-physicalists infer from an epistemic gap between physical and phenomenal facts a corresponding ontological gap. The included physicalists affirm an epistemic gap but deny a corresponding ontological gap. A minority physicalist opinion holds Mary to be a deviant: in knowing all of the physical facts Mary automatically knows what it's like to see red despite never having seen red before. This minority denies the epistemic gap definitive of the opposing majority. To label the groups so far described, let us use "qualia anti-physicalists," "gappy physicalists," and "non-gappy physicalists."[2]

I count myself among the non-gappy physicalists, though my present concern is not so much to provide arguments for my view as to provide arguments against certain opponents. My main aim in this paper is to articulate problems that arise for certain prominent qualia anti-physicalists. Additionally, though they will be far from my focus, I will have a few remarks to make about gappy physicalists. The problems I raise all concern deviants. Before saying more about the problems, more must be said about deviants.

Let us set Mary aside for now since controversy surrounds her status as a deviant. Deviants to grace the recent literature include Dennett's (2007) RoboMary, Mandik's (2009, in preparation) Hyperbolic Mary, and Love's Swamp Mary.[3]

A single deviant will do for present purposes, and a brief introduction should suffice for the uninitiated. Swamp Mary is a being that arises by quantum accident in classic Davidsonian Swamp-Man-style (Davidson 1987), but crucially, Swamp Mary is in a physical state intrinsically identical to the state Mary is in after Mary has seen, but is no longer seeing, red. Conducting a Swamp Mary thought experiment involves assuming that post-experiential Mary knows what it's like to see red and further, retains this knowledge even over stretches of time in which she is no longer having an experience of red. Also involved is an assumption that Swamp Mary has never herself experienced red. The conclusion urged is that Swamp Mary's physical similarity to post-experience Mary suffices for the relevant phenomenal knowledge despite Swamp Mary's never having had the target phenomenal experience.

Some non-gappy physicalists, e.g. Dennett (2007; Mandik 2009, in preparation), hold that the very possibility of deviants like Swamp Mary undermines both qualia anti-physicalism and the Knowledge Argument. In opposition, Alter (2008) argues that deviants are harmless in both cases. The purpose of the current paper is to present a pair of novel arguments to tip the balance against (certain) qualia

[2] This way of sorting philosophies of mind in terms of various attitudes toward epistemic and ontological gaps is due to Chalmers (2003a). The gappy physicalists are identified by Chalmers as "type-B materialists". The non-gappy physicalists are "type-A materialists," but there's also reason to regard what Chalmers calls "type-Q materialists" ("Q" for "Quinean") as non-gappy. For an extended defense of type-Q materialism, see Mandik and Weisberg (2008).

[3] Dennett (2007, p. 24) attributes the suggestion of Swamp Mary to Gabriel Love.

anti-physicalists. I will not be targeting all of the possible positions that merit being described as qualia anti-physicalists. I will instead be restricting my focus in ways to be detailed in Sect. 2. Also in Sect. 2 I make further preliminary remarks concerning the Knowledge Argument and deviant phenomenal knowledge, further setting the stage for the arguments to be developed in Sects. 3–4. I dedicate Sect. 3 to what I call the *Psychosemantic Argument*, the thrust of which is that the targeted qualia anti-physicalists are in a comparatively poor position to accommodate and account for the *psychosemantic* requirements on deviant phenomenal knowledge, the requirements that must be satisfied in order for putative knowledge states of beings who have not had certain phenomenal experiences to count as *representations* of the relevant phenomenal facts.

I develop further psychosemantic considerations against qualia anti-physicalism in Sect. 4's *Factivity Argument*, the thrust of which is that if a deviant like Swamp Mary is able to satisfy the psychosemantic requirements on phenomenal knowledge, then the way is cleared for pre-release Mary (non-Swamp Mary) to deduce the relevant phenomenal facts from the physical facts.

2 Further preliminaries

A wide variety of possible positions in the philosophy of mind may be described as versions of qualia anti-physicalism. Examples include certain versions of idealism, neutral monism, epiphenomenal dualism, and interactionist dualism. It enhances the tractability of my discussion without diminishing its interest to here restrict attention to a proper subset of the possible qualia anti-physicalist positions. Exposition will also by eased by referring, for now on, to the targeted position simply as "anti-physicalism".

For the purposes of the current discussion, my anti-physicalist target can be picked out as embracing three key theses: (1) the ontological simplicity of qualia, (2) qualia epiphenomenalism, and (3) the soundness of the Knowledge Argument.

The claim of ontological simplicity of interest here is the claim, for at least some qualia, that they are not ontologically dependent on anything else. This is consistent with affirming of some other qualia that they are ontological complexes. For example, it is possible to hold that (a) there is a complex quale associated with seeing something with red and white stripes and (b) this complex is decomposable into component qualia such as a red quale and a white quale. It may be held that the red and white qualia are themselves ontological simples. One kind of view that embraces the simplicity claim would be a dualism that holds that there exist two kinds of irreducible fundamental entities or properties: irreducible *physical* entities or properties (the basic particles and forces of a completed microphysics) and fundamental *phenomenal* properties or entities that are not reducible to any physical entities (see, for example, Chalmers 1996, especially pp. 125–126).

The claim of epiphenomenalism is the claim that qualia lack (physical) causal effects. An example of someone who holds such an epiphenomenalism would be a dualist who affirms a causal closure principle for the physical whereby all physical events have physical causes and rejects an over-determination principle thus rejecting the possibility of having, in addition to physical causes, any non-physical causes (see, for example, Chalmers 1996, especially pp. 158–159).

Controversy surrounds the question of how best to represent the Knowledge Argument. I turn now to briefly review some aspects of the controversy. I note, however, that my aim in this brief review is not to settle the controversy over the Knowledge Argument, but to articulate some ideas that will be utilized in the Psychosemantic Argument and the Factivity Argument. I will begin by presenting a way of representing the Knowledge Argument that helps to convey the nature of the non-gappy physicalists' deviant-based attacks. Afterwards I turn to highlight some of Alter's (2008) criticisms of this way of thinking of the Knowledge Argument.

Let us consider the Knowledge Argument as having the following two premises.

Premise One	If physicalism is true then Mary, knowing all physical facts, also knows what it's like to see red, even though she has never experienced red before
Premise Two	However, Mary learns something new upon seeing red for the first time; she acquires knowledge of what it's like to see red

Deviants may be wielded against the Knowledge Argument by attacking Premise Two. Such an attack may be mounted in the following manner. We begin by noting that Premise Two needs an explanation to be true. It is a poor candidate for something that is *merely* true.[4] The only thing that looks like a plausible explanation of Premise Two is a thesis we can call the *Experience Requirement*, the thesis that, for some experiences at least,[5] and red experiences in particular, knowledge of what it's like to have such an experience requires that the knower has had or is currently

[4] We can, on perhaps some occasions, accept unexplained truths if they themselves are *explainers*. But there's no apparent explanatory work that acceptance of Premise Two does, so a proponent of it who denies it needs an explanation is offering us a putative truth that neither is explained nor does any explaining. It's difficult to see that there can be any grounds for accepting alleged facts that are entirely cut loose from the web of explanation.

[5] Formulating the Experience Requirement as being about *all* experiences would make it an implausibly strong claim. Consider, in opposition to this stronger version of the Experience Requirement, the Humean concern about a missing shade of blue. Plausibly, a person may know in advance what it is like to see a previously unseen shade of blue if they have previously seen other shades of blue. Similarly, a person who has never seen a red and white striped object before may nonetheless know in advance what it would be like if they had prior experiences of striped things, red things, and white things.

having such an experience.[6, 7] However, that deviants are possible makes clear that the Experience Requirement is false, even for prototypical instances of qualia such as phenomenal red. Thus, by falsifying the only plausible explanation of Premise Two, the possibility of deviants renders Premise Two intolerably inexplicable and thus false.

Alter (2008) has urged that a defender of the Knowledge Argument may embrace the possibility of deviants and regard such a possibility as irrelevant to the evaluation of the Knowledge Argument. Alter construes the physicalists' deviant-based attack as mistakenly taking the Knowledge Argument to depend on a claim that one can only gain phenomenal knowledge in one specific way. Instead, claims Alter, the Knowledge Argument depends on a claim concerning a way that phenomenal knowledge *cannot* be learned. As Alter (p. 249) puts the point, the intuition central to the Knowledge Argument "has the form 'one cannot learn Q in way W' not 'the only way to learn Q is in way W'". More specifically, what is crucial, according to this line, is a claim that phenomenal knowledge cannot be *deduced* from physical knowledge. Call this the *Nondeducibility Claim*. The essence of Alter's defense of the Knowledge Argument is to hold that the possibility of deviants is consistent with the Nondeducibility Claim.

It is worth mentioning, though I will not pursue such a strategy here, that one may question Alter's construal of the Knowledge Argument as hinging on claims concerning deduction. For example, Dennett responds to such a claim by writing "I just don't see that this is what matters. So far as I can see, this objection

[6] An anonymous reviewer asks why isn't another plausible explanation one that incorporates the disjunct "or the knower is a deviant". One possible response, that I won't develop here, is to point out general problems for disjunctive explanations. Another line of response that strikes me as more immediately promising is to raise problems for the proffered disjunct. I assume "deviant" to be a technical term whose stipulated meaning is exhausted by characterizations such as "being who knows what it's like, for example, to see red without having seen red". On such an assumption, then, "deviant" is no natural kind term and lacks what philosophers of science call the "surplus content" needed to do the required explanatory work. If this line about lacking surplus content is correct, then the proffered disjunctive explanation may be read as the following obviously unsatisfactory explanation of Premise Two: "knowledge of what it's like to see red requires that the knower either has experienced red before or is able to know what it's like without having experienced red before." Another way of putting my point may be by noting that what needs to be explained about Premise Two is why Mary isn't a deviant and so any explanation that amends the disjunct "or the knower is a deviant" gains no additional explanatory power.

[7] Alter (2008, pp. 263, 5) presents an interesting case that there is relatively sparse evidence of any philosophers explicitly affirming a thesis Alter identifies as "the experience requirement". Two points are worth noting in the present context. First, the thesis Alter calls "the experience requirement" is "the idea that seeing in color is required for knowing what it's like to see in color" (p. 248) and it is not clear whether this is to be read as equivalent to what I am presently calling "the Experience Requirement". A person who has seen colors other than red before and knows what it's like to see red prior to seeing red would satisfy at least one reading of Alter's requirement but not satisfy mine. The second, and perhaps more important, point worth noting is that it is not *directly* relevant to the present discussion how many philosophers explicitly embrace the Experience Requirement. Crucial to the deviant-based attack is the claim that the Experience Requirement is the only plausible explanation for Premise Two. This claim is compatible with the claim that the Experience Requirement is seldom *explicitly* affirmed. Of course, there might be some way in which Alter's sociological claim is indirectly relevant. Perhaps an argument can be constructed against the deviant-based attack that has as a premise "If the Experience Requirement were the only plausible explanation, more philosophers would have explicitly affirmed it". However, I will not consider this line of thought further.

presupposes an improbable and extravagant distinction between (pure?) deduction and other varieties of knowledgeable self-enlightenment" (Dennett 2007, p. 29).[8] The strategy I instead pursue is to argue, in the Factivity Argument, against the truth of the Nondeducibility Claim. However, the Factivity Argument depends on points to be developed first in the Psychosemantic Argument.

3 The Psychosemantic Argument

At the core of the Psychosemantic Argument is the following idea: If the possibility of deviants is consistent with anti-physicalism, then there must be some account, consistent with anti-physicalism, of how it is possible for a state of a deviant to constitute a representation of the relevant phenomenal facts. I assume as a truism about knowledge that one can only know that *such-and-such* is the case if one can represent—that is, have a state that represents—that *such-and-such* is the case. A straightforward application of this assumption to phenomenal knowledge yields as a truism that one can only know that what it's like to see red is such-and-such if one can have a state that represents that what it's like to see red is such-and-such.[9]

Setting aside for the moment both anti-physicalists and deviants, let's survey the available kinds of psychosemantic accounts of how the state of a person can suffice for representing what it's like to see red. After the survey I will address which, if any, of the four options are consistent with the conjunction of anti-physicalism and the possibility of deviants.

The past several decades gave rise to a large literature in which many various theories of content were offered, shot down, refurbished, and resurrected. Nonetheless, for the purposes of this paper this great variety of psychosemantic options may be reduced to only four. They are: 1. **Quotation**, 2. **Actual Cause**, 3. **Descriptive-isomorphism**, and 4. **Nomological**. Of course, I cannot be absolutely certain that there is not some as yet undiscovered theory of content that would demand adding a fifth option to my list. Nonetheless, I will proceed under the

[8] For Alter's (2008) detailed response to this remark of Dennnett's see pp. 250–253.

[9] Further, this representational requirement on phenomenal knowledge seems to be the most important requirement as opposed to, say, requirements concerning justification. I think that Alter (1998) is exactly right when he writes:

> [L]et us ask what exactly [Mary's] lack of factual knowledge consists in. We color-sighted folk in the outside world are supposed to know facts that she does not, but what distinguishes her epistemic state from ours? The difference does not seem to turn on *justification*. That is, her problem is not that she has the same beliefs as we, but her beliefs, unlike ours, are unjustified—as though she suspects that seeing red has a certain distinctive phenomenal quality, the same one we know it to have, but she cannot confirm her suspicion. Rather, if she lacks knowledge of facts about color experiences, this would seem to be because she lacks the appropriate beliefs: certain propositions, which we grasp, are inaccessible to her. (p. 46)

I should clarify that I agree with Alter here on what failing to know what it's like consists in. I do not, of course, agree that pre-experience Mary fails to know what it's like. For a view in opposition to mine and Alter's, one that construes Mary's knowledge failure as hinging crucially on *justification*, see Beisecker (2005).

assumption that my four-item list is exhaustive. I will also indicate how some notable extant theories fit into my taxonomy.

According to **Quotation**, the states that constitute representations of phenomenal properties themselves instantiate or incorporate phenomenal properties. Versions of **Quotation** include Papineau's (2002) account of phenomenal concepts.[10] One's thought about experiences when one is having the experiences include those very experiences as constituents. When one thinks of those experiences *after* having had them, the thoughts incorporate imaginative recreations of the experiences where the recreation is quite literal and involves literal replication of the items thereby thought about. Just as a carpet sample represents a red shag carpet by both the sample and the carpet of which it is a sample—the sample and the sampl*ed*—instantiating the same properties (redness, shagginess), the deployment of phenomenal concepts involves both the thoughts and the thought-about co-instantiating the same properties (red qualia and perhaps even shaggy qualia). Other examples of **Quotation** include Gertler's (2001) "embedding" account of the reference of introspective states and Chalmers's (2003b) account of "direct phenomenal concepts" wherein they are "based on the attention to the quality, 'taking up' the quality into the concept" (p.235).

According to **Actual Cause**, a representation of a red quale need not currently coincide with the instantiation of a red quale. However, the instantiation of a red quale must be one of the antecedent links of a causal chain eventuating in the representation of a red quale. If **Actual Cause** is to be modeled on the causal theories of reference of Kripke (1980) and Putnam (1981), then the crucial initial link in the chain may be regarded as a reference-fixing dubbing or "baptismal" event. If **Actual Cause** is to be modeled on the teleosemantics of Millikan (1984) then the crucial initial link in the chain may be a long-past event in the representer's evolutionary history.

According to **Descriptive-isomorphism**, the key idea, for the purposes of this paper, is that both representations and representeds are complexes and representation is conceived of as hinging on the structural complexity of both represented and representation. Examples of this sort of semantic approach include the descriptive theory of names (e.g., Russell 1905), conceptual role semantics (e.g., Block 1986; Sellars 1953), classical resemblance theories (see Watson (1995) for a review), and contemporary isomorphism-based theories (e.g., Cummins 1996).

According to **Nomological**, causation plays a crucial psychosemantic role but, unlike **Actual Cause**, counterfactual and not actual causes do the heavy lifting. Aficionados of the theory of content literature will recognize Fodor's (1987) Asymmetric Dependency theory as the key exemplar of this approach to psychosemantics. Fodor's theory is designed with an eye toward solving the disjunction problem for causal theories of content: the problem of explaining how a representation that can be caused by either cows or horses comes to have as its

[10] Though note that Papineau (2007) abandons this account of phenomenal concepts for one better construed as either an instance of **Actual cause** or some combination of **Actual cause** and **Descriptive-isomorphism**.

content just COW and not COW-OR-HORSE. The gist of Fodor's solution is to hold that representation tokens, R's, represent cows and not horses if and only if…

1. It's a law that cows cause R's,
2. It's a law that horses cause R's,
3. If it were not the case that (1), then not (2),
4. If it were not the case that (2) it would still be the case that (1).

Nomological spelled out Fodor-style achieves its key distinctness from **Actual Cause** by it being possible that there is a particular R that was not actually caused by a cow even though it's a law that cows caused R's.

Let us turn now to consider which of these four psychosemantic approaches can be embraced by the anti-physicalist in need of an account of how a state of a deviant can represent the phenomenal facts she has knowledge of.

It should be immediately obvious that **Quotation** and **Actual Cause** are unavailable to anyone—physicalist or anti-physicalist—embracing the possibility of deviants. A deviant like Swamp Mary is stipulated to know what it's like to have a red experience without ever having had a red experience. Also, she's stipulated to have popped into existence fully formed and thus no state of her has any causal antecedants. There's no red experience, then, to serve as either a quotational constituent or causal antecedent of Swamp Mary's representation of what it's like to have a red experience.

Turning then to **Descriptive-isomorphism**, we find a third psychosemantic option unavailable to the anti-physicalist. In this case the unavailability is due to something internal to anti-physicalism. In particular, the conflict that arises is between the anti-physicalist doctrine that qualia are ontological simples and the **Descriptive-isomorphism** requirement that the targets of representation be complexes. A crucial plank of the anti-physicalist platform is that at least some qualia, especially a red quale, are not decomposable into simpler constituents nor are they constituted by any relations, causal, functional, or otherwise, that they bear to anything else. It is this ontological simplicity that, according to the anti-physicalist, bars the reduction of qualia via any physical or functional analysis. This ontological simplicity likewise accounts for the ineffability of qualia—the alleged impossibility of conveying what it's like to have phenomenally conscious experiences via any description. Ontological simples like a red quale cannot be conveyed by description in the way that ontological complexes like experiences of red and white striped objects can. And, crucially, this ontological simplicity bars the anti-physicalist proponent of deviants from explicating deviant phenomenal representation along the lines of **Descriptive-isomorphism**.

The last remaining psychosemantic option to consider is **Nomological**. Is *it* consistent with the conjunction of anti-physicalism and the possibility of deviants? It will be instructive here to see how far we can go in attempting to adapt Fodor's Asymmetric Dependence theory to fit with anti-physicalist doctrines. The first obstacle to surmount is that Fodor's theory requires that representeds have causal effects and anti-physicalists are also epiphenomenalists and thus hold qualia to be without effects.

A potential way out for the anti-physicalist/epiphenomenalist seeking to adapt and adopt Asymmetric Dependence is to exploit the anti-physicalist doctrine that even though qualia do not *logically* supervene on the physical, they nonetheless, in this and nearby worlds at least, *nomologically* supervene on the physical (Chalmers 1996). Perhaps, then, the anti-physicalist/epiphenomenalist can harness nomological supervenience for a version of **Nomological** psychosemantics in the following fashion. The suggestion is to restate the Fodorian theory with a modification made by replacing all instances of "cause" with "kause" where the technical notion of kausation is stipulated to have the following analysis:

C kauses E = df (C supervenes on B) & (B causes E).

So, R's represent Q's and not Z's if and only if

1. It's a law that Q's kause R's
2. It's a law that Z's kause R's
3. If it were not the case that (1), then not (2)
4. If it were not the case that (2) it would still be the case that (1)

Spelling this out just a bit more, suppose the following: Q is a red quale, B_1 is a brain state that Q supervenes on, Z is either a non-red quale or a non-quale, B_2 is Z's supervenience base, and R is a concept of a red quale. The anti-physicalist may suppose that, on many occasions of introspection, the application of R is kaused (with a 'k') by Q meaning that R is caused (with a 'c') by B_1, the supervenience base of Q. The anti-physicalist may also suppose that introspective error is possible and such occasions of error may be illustrated by R's being kaused (with a 'k') by Z. The adapted Asymmetric Dependence theory is designed to account for why it is that R's content is Q and not the disjunctive content Q-or-Z.

Prima facie, the adapted Asymmetric Dependence theory that swaps causation-with-a-'c' for kausation-with-a-'k' surmounts the difficulty that epiphenomenalism initially posed. However, I think the adaptation contains a much deeper problem: it is *unknowable* whether any of the four propositions that make up adapted Asymmetric Dependence are ever true. The source of the trouble here is that if *epiphenomenal* qualia nomologically supervene on anything it's *unknowable* that this is the case.

I'll say more to substantiate my unknowability charge in just a moment, but I first want to note what appear to be the anti-physicalist's only options in responding to such a charge. The first line of response is to build a case that the requisite laws are knowable after all. The second line of response is to claim that the knowability of the laws is irrelevant. I turn now to briefly sketch some grounds for being pessimistic about the first line of response.

Chalmers (1996, p. 215) is live to the worry that the kind of anti-physicalism he embraces leads to worries of the unknowability of the posited psychophysical laws. As Chalmers conceives of the nature of consciousness, it constitutes a domain wherein "rigorous intersubjective testing is impossible" and thus a science of consciousness "will probably always lack the strong empirical credentials of other sciences" (p. 218).

In response to such worries, Chalmers presents a detailed set of considerations, especially in chapters 6 and 7, for why he remains optimistic about the scientific tenability of his "naturalistic dualism". I find the case for optimism uncompelling and will here briefly convey why I remain pessimistic.

For starters, it certainly doesn't help the anti-physicalist that no psychophysical laws have yet been discovered. Of course, that no psychophysical laws are *known* doesn't constitute a principled reason for thinking that psychophysical laws relating epiphenomenal qualia to brain states are *unknowable*. Nonetheless, principled reasons for such pessimism can be given.

One sort of principled reason points out a very serious asymmetry between the entities posited in the physical sciences and allegedly epiphenomenal qualia. Entities such as the stars and planets of astronomy, the electrons and protons of particle physics, and the brain areas of neuroanatomy enter into a rich set of detection-supporting causal interactions. It is this multiple-detectability that is crucial for deciding whether an investigative technique has yielded a genuine datum or a mere artifact of the technique (see, for example, Bechtel, in press). In contrast, if the anti-physicalist is right, the only detection-supporting relation a quale enters into is the relation of acquaintance between a quale and its lone introspector. This puts the anti-physicalists at a serious disadvantage in responding to worries that what they are actually acquainted with are artifacts of their chosen investigative technique.[11]

A related worry concerns whether heavily first-personal techniques can ever provide the kind of *correlational* data—data concerning correlations between qualia and brain states—needed to underwrite psychophysical laws. Even if Chalmers were right that "we have a rich set of data in our own case" (p. 215) the data here are just data about the presence of my own current qualia, not data correlating my qualia to my brain states. Worse, in seeking to gain the requisite correlational data by, for instance, applying a futuristic cerebroscope to my own brain, the worry arises that I'll only be accessing data about what it's like to see red while simultaneously observing the read-out of a brain scanner. It's difficult to see how I could gain, from the first-person perspective, correlational data concerning what it's like to see red while *not* reading a brain scanner's output. Attempts to circumvent this problem by having someone else read the brain scanner and record my verbal reports about my current qualia lead to further worries about artifacts, like the worry that what I'm really accessing is what it's like to give a verbal report. Unsullied data correlating brain states with the quale of simply seeing something red will continue to elude us.

Chalmers, no doubt, will remain optimistic that some future considerations of theoretical simplicity and plausibility will decide between competing theories underdetermined by the admittedly relatively small set of data. Chalmers writes, of the worry of untestability, that "[t]his worry will only come into play in a strong way if it turns out that there are two equally simple theories, both of which fit the data perfectly, and both of which meet the relevant plausibility constraints" (p. 217). He further states that such a worry about multiple competing theories is

[11] See Schwitzgebel (2008) for an excellent recent discussion of the pitfalls of introspective artifacts.

premature since we currently lack "even a single theory that can handle the phenomena in a remotely satisfactory way" (p. 218).

My attitude about the relative scarcity of equally adequate epiphenomenalist theories of consciousness is just about the exact opposite of Chalmers's. It strikes me that, given the conceptual toolbox that Chalmers utilizes in arguing against the logical supervenience of the phenomenal on the physical—a toolbox populated by zombies, inverted spectra, dancing and fading qualia, and the like—it ought to be *exceedingly easy* to construct multiple epiphenomenalistic theories that are equal with respect to empirical and theoretical virtues. I turn now to a quick sketch of how such constructions might proceed.

Let us begin by supposing the following. Suppose that I am currently attending a red quale right now, Q1, and form a concept of that determinate quale. Suppose further that there is a phenomenal determinable whose determinate instances, Q1–Qn, may be ordered with respect to being more or less "faded". I suppose we understand what relative fadedness is insofar as we understand Chalmers's "fading qualia" thought experiments (1996 pp. 253–263) or Hume's suggestion that an idea is a less vivid copy of a sensory impression. Suppose further that the current read-out of my cerebroscope tells me that brain state B is currently co-instantiated with Q1.

All of the suppositions in the previous paragraph are consistent with theory T1, which says, among other things, that it is a law that Q1 is co-instantiated with B. However, these suppositions are also consistent with T2 which differs from T1 only in saying that Q2 is nomically co-instantiated with B. Relative to T2, the observation that on this occasion Q1 and not Q2 was co-instantiated with B must be regarded as a bit of noise or a data point that doesn't fit the curve. By similar reasoning we can generate T3, T4, T5, and so on which respectively nomically associate Q3, Q4, Q5, and so on with B. (And if we make the further assumption that there are countably many qualitatitve determinates that are more faded than Q1, then we can generate countably many competitors for T1.)

I don't see that there can be any difference in *simplicity* between T1-Tn, since they differ only with respect to whether they say it is Q1 or Q2–Qn that is nomically related to B. Will collecting more data help? I don't see how. No matter how many more instances of Q1 I access from my first-person perspective and confirm to be co-instantiated with B, there are more than six billion people alive today who all may have Q2 (or Q3…) co-instantiated with B. Relative to the six billion other people alive today, my own data may just be a small number of points that fall outside of the curve. And I don't see that a *plausibility* consideration is really going to serve to rule out the possibility that my qualia are mildly faded with respect to the rest of the population, or that roughly half of the population is slightly faded, and so on.

I turn now to the other strategy the anti-physicalist might pursue in responding to the unknowability charge, namely that it is irrelevant to the anti-physicalist adherent of **Nomological** whether the requisite psychophysical laws be knowable. The anti-physicalist might assert instead that the statements concerning laws in their version of **Nomological** psychosemantics need only be *true*.

The problem that arises for this sort of anti-physicalist strategy is that it makes the anti-physicalist who embraces deviants *especially* vulnerable to the *Factivity Argument*. Before turning to elaborate this line of thought, it may be useful to take

stock so far. The anti-physicalist who embraces the possibility of deviants faces the problem of explaining how deviants can satisfy the psychosemantic requirements on phenomenal knowledge. The general approaches of **Quotation**, **Actual Cause**, and **Descriptive-isomorphism** are not at all promising. The least problematic option, **Nomological**, is nonetheless deeply problematic for the anti-physicalist. The problems hinge on whether nomic relations between the physical and the phenomenal need to be *knowable* in order for **Nomological** to be promising. If they do need to be knowable, then the problem arises that their knowability has not yet been convincingly demonstrated. If they do not need to be knowable, then the anti-physicalist is rendered especially vulnerable to the line to be developed in the next section.

4 The factivity argument

The argument that follows is designed to target the anti-physicalist's Nondeducibility claim and show instead that phenomenal facts are deducible from physical facts.

Let us stipulate the following. Let "D" stand for a massive conjunction exhaustively describing, in an exclusively physical vocabulary, the total current state of a deviant. Let "Q" stand for what it's like to see red, a proposition of the form **What it's like to see red is *such-and-such***. Let "KQ" be the application of an epistemic modal operator to "Q" so that "KQ" is read as "It is known that Q". Let "→" be, at a minimum, an implication operator exhibiting transitivity. I will interpret it as material implication for now and will address a bit later whether it needs some other interpretation for the argument to adequately target Nondeducibility. Let "Deducibility" be the negation of the dualist's Nondeducibility claim. The argument, then, is as follows.

P1. D → KQ
P2. KQ → Q
C1. D → Q
C2. Deducibility

I take it as obvious that C1 follows straightforwardly from P1 and P2 via the transitivity of implication and thus no serious questions arise concerning the validity of this inference. I also regard the truth of P2 as not worth questioning since I assume the factivity of knowledge and thus that it is known that Q only if Q.

I take C2 to follow pretty straightforwardly from C1. If it is indeed the case that a phenomenal fact is entailed by a physical fact, then Deducibility is thereby established.

I take it, then, that the interesting questions surround the truth of P1. One might object to P1 on the grounds that while a deviant can lack a red quale a deviant cannot lack *all* qualia. David Chalmers (personal communication) has advocated the possibility of deviants while also insisting that the deviant cannot be a *zombie*—a being devoid of qualia while physically/functionally identical to a being with qualia. Not being a zombie, the deviant has *some* qualia and so the physicalist may seem to

be begging a crucial question in P1. In other words, according to this line of objection developed from Chalmers' suggestion, D fails to entail KQ since D's being stated in an exclusively physicalistic vocabulary means it fails to capture relevant aspects of the deviant's total state. In other words, D could be satisfied by a zombie and a zombie wouldn't have phenomenal knowledge, so P1 is false.

However, let's consider a defense against this objection to P1 not in terms of zombies but in terms of Swamp Mary. Swamp Mary, let's suppose, has existed long enough to have had various non-red qualia, and thus despite lacking red qualia, is no zombie. Now consider what happens when we knock her out with a general anesthetic. Several key things are true of anesthetized Swamp Mary. First, since she's under a general anesthetic, she has *no* qualia. Second, general anesthetics aren't, generally, *amnestics*. I've had various surgeries and thus received anesthesia on various occasions. On none of these occasions did I forget my home address, mother's maiden name, or, most importantly, what it's like to see red. We may suppose a similar sort of point applies to Swamp Mary. Third, Swamp Mary is no zombie. In her generally anesthetized state, she, like a zombie, lacks qualia, but unlike a zombie, isn't physically/functionally identical to someone currently enjoying qualia.

The crucial upshot of anaesthetized Swamp Mary is that in her anesthetized state, she has no qualia and thus, an exclusively physical vocabulary should suffice to capture her total current state. However, since anesthetics aren't amnestics, anesthetized Swamp Mary retains her deviant phenomenal knowledge.

I turn now to consider a line of objection against the Factivity Argument that may perhaps be interpreted as hinging on what the relevant notion of entailment is that is needed to defeat Nondeducibility. This line questions whether "→" is best interpreted as material implication or should instead be read as something stronger. One way of conveying what this something stronger should be is that it should be such that Mary, prior to experiencing red, should be able to make a transition from grasping the contents of any of the relevant antecedents of the conditionals in question to grasping the contents of the consequents of these conditionals. She should, upon examining Swamp Mary, be able to know what it is that Swamp Mary knows.

Alter (2008) discusses a similar scenario involving what pre-experience Mary can know about Swamp Mary. Alter urges that intuitively, pre-experience Mary cannot know the content of Swamp Mary's total state. Putting Alter's point in terms of the current argument, pre-experience Mary can know D, but not the content of Swamp Mary's knowledge state, namely Q. Another way of putting this objection to the Factivity Argument is to say that intuitively, pre-experience Mary cannot herself have a state with the same content as the deviant's state.

The line of objection under consideration depends crucially on claiming a necessary asymmetry between the contents of Swamp Mary's states and the states of pre-experience Mary. However, there's a crucial flaw in this line of objection and it can be drawn out by asking what could possibly *explain* this alleged psychosemantic asymmetry between Mary and Swamp Mary.

Let us turn again to the four psychosemantic options from the previous section, this time asking whether any account can be given of the alleged asymmetry

between Mary and Swamp Mary. It should be immediately obvious that neither **Quotation** nor **Actual Cause** can do the trick, since both Mary and Swamp Mary are equally incapable of satisfying the requisite relations to a red quale. And since this line of objection is being advanced by an the anti-physicalist, **Descriptive-isomorphism** is unavailable since the anti-physicalist affirms the ontological simplicity of the relevant qualia. No asymmetry between Mary and Swamp Mary thereby obtains because both are equally incapable of having states descriptive of or appropriately isomorphic to ontological simples.

Let us turn, then, to ask whether **Nomological** can provide any grounds for the anti-physicalist's claim of a psychosemantic asymmetry between Mary and Swamp Mary. If such a maneuver is available to the anti-physicalist, then it is going to involve a semantics-supporting nomological relation that obtains between the relevant phenomenal facts and a state of Swamp Mary while at the same time necessarily failing to obtain between the phenomenal facts and every possible (candidate representational) state of (pre-experience) Mary. It's not clear, however, that the anti-physicalist can establish that such an asymmetry must obtain. The question I want to press here is: What is to prevent *Mary* from "piggy-backing" on whatever nomological connections relate *Swamp Mary's* states to the phenomenal facts? And the answer I want to press here is: Nothing prevents Mary from such "piggy-backing".

The notion of "piggy-backing" at play here may be conveyed by an example. I presume that **Nomological** lends itself quite naturally to the following account of how scientists are able to think about entities such as electrons and radio-waves that are imperceptible to the unaided senses yet detectable via certain instruments. First, an instrument's detection functions are sustained via nomological relations that obtain between the entities in question and states of the instrument. Second, instrument states are able to be read by the scientist in virtue of nomological relations that obtain between instrument states and states of the scientist's sensory systems. (Of course, if the scientist wears hearing aids or eyeglasses, we need to introduce a third set of nomological relations, but I will ignore such complications for simplicity here.) It is in virtue of the second sort of nomological relations that the scientist is able to "piggy-back" on the first sort and thus secure a semantic grip on the imperceptible.

With that sketch of "piggy-backing" in hand, let's turn to see whether Mary can exploit Swamp Mary the way scientists generally exploit their instruments. The question of whether "piggy-backing" may take place may be phrased in terms of whether the two sorts of nomological relations may obtain, the first being between the distal target and the instrument and the second being between the instrument and the scientist. That the first sort of relation obtains between Swamp Mary and the physical facts is beyond question here since it must be assumed by the application of **Nomological** to account for Swamp Mary's phenomenal knowledge. And we have every reason to believe that the second sort of relation obtains—the one between Swamp Mary and Mary—since it's stipulated by both anti-physicalists and their opponents that Mary is physically omniscient. So there's no physical state of Swamp Mary that Mary can't nomologically relate to exactly as well as she relates to the various states of her scientific instruments (she is a *super* neuroscientist, after

all). Of course, if the requisite aspects of Swamp Mary involve e.g. microscopic features of her nervous system, we need to posit a third set of relations, those between Mary and e.g. her microscopes, but such complicating factors do nothing to diminish the main point about "piggy-backing". Mary, being physically omnisicient, is "locked on" to every physical state of Swamp Mary, and it's the core supposition of **Nomological** that Swamp Mary is "locked on" to every phenomenal fact she has deviant phenomenal knowledge of.

The anti-physicalist may be tempted here to resist the "piggy-backing" argument on the grounds that Mary is in no position to verify whether her "instrument" is correctly calibrated. The intuition may persist that, unless pre-release Mary is able to perform this direct verification, she is in no position to understand what it is like to see red. I have several responses to this.

My first response is to point out that this attempt at resistance simply begs the question against those non-gappy physicalists (e.g. Dennett 2007; Mandik 2009, in preparation) who argue that Mary herself is a deviant, since this line of anti-physicalist resistance *just is* the insistence that pre-experiential Mary cannot know what it's like to see red. What this line of resistance needs but fails to provide is an explanation of how **Nomological** can suffice to explain Swamp Mary's phenomenal knowledge without also granting the same phenomenal knowledge to pre-release Mary.

My second response is that it is quite common for scientists to utilize instruments that they themselves neither calibrated nor are in any position to do so. Likely most thinkers that have a semantic grasp on the various imperceptibles that populate scientific discourse are not themselves in a position to verify the calibrations of the instruments that constitute key links in the relevant nomological chains. I see no reason, then, why a similar point wouldn't also be true of Mary's relation to what Swamp Mary knows.

My third response is to point out the tension between the current anti-physicalist line and the one scouted in the previous section concerning how the obtaining of nomological relations don't have to be knowable, just true. If the anti-physicalist is going to insist that **Nomological** can be utilized without the laws in question being knowable, then the physicalist may make a similar maneuver in the current context and point out that it doesn't matter if Mary is in a position to *verify* the "calibration" of Swamp Mary. As long as the requisite laws are true, Mary's states that result from her deductive inference satisfy the psychosemantic conditions on phenomenal knowledge.

5 Conclusion

I have argued that the possibility of deviants spells trouble for anti-physicalism on the grounds that the anti-physicalists are in no position to account for the psychosemantic grounding of the deviants' phenomenal knowledge in the phenomenal facts in a way that is consistent with denying that pre-release Mary is a deviant. I close by addressing whether physicalists are in any better position to account for the requisite psychosemantic grounding.

As already pointed out, no one embracing the possibility of deviants may avail themselves of **Quotation** or **Actual cause**. The question that remains, then, is whether the physicalists who embrace the possibility of deviants may avail themselves of **Descriptive-isomorphism** or **Nomological**. And I think that things look pretty good here for the physicalists. Not being committed to the ontological simplicity of qualia, it's open for a physicalist to account for the deviant's psychosemantic grounding along the lines of **Descriptive-isomorphism**. And if there are physicalists who, for independent reasons, are not particularly fond of the kind of psychosemantic story offered by **Descriptive-isomorphism**, then they may avail themselves of **Nomological** with ease. Such ease comes with the physicalist denial that qualia are epiphenomenal. Such a physicalist, then, will have a much easier time dealing with the questions of *knowability* that arise for the various laws they will need deal with.

My main aim in this paper has been to wield deviants against anti-physicalists. Although it likely merits the treatment of an entire separate paper, I do want to briefly address the potential relevance of deviants in debates between gappy and non-gappy physicalists. Gappy physicalists, in denying that Mary is a deviant, would incur special psychosemantic burdens if they wanted to also admit that there could be deviants such as Swamp Mary. In past discussions of Mary, at least some gappy physicalists, in particular Papineu (2002, 2007) have wanted to supply psychosemantic accounts of her phenomenal concepts that conform to what I've called **Quotation** and **Actual Cause**. If the gappy physicalist wants to claim additionally that Swamp Mary also has phenomenal concepts, then this puts a strain on the gappy physicalist's psychosemantics. If my above arguments are correct, then the gappy physicalist, if attributing phenomenal concepts to Swamp Mary, would be forced to adopt a psychosemantics conforming to either **Descriptive-isomorphism** or **Nomological**. However, such a move incurs the further burden of explaining why pre-release Mary fails to satisfy the psychosemantic requirements on phenomenal knowledge. It is beyond the scope of the current project to argue that the gappy physicalist cannot solve this problem. But I close in noting that it is a problem that the non-gappy physicalist does not have to solve.

Acknowledgements For especially useful comments on earlier versions and precursors, I thank Torin Alter, David Chalmers, Josh Weisberg, and an anonymous reviewer. I'm also grateful for discussions with Patricia Churchland, Paul Churchland, Paula Droege, Brian Fiala, Ellen Fridland, Chris Gauker, Uriah Kriegel, Eric Schwitzgebel, John Martin, Tom Polger, William Robinson, William Seager, Ben Young, and audiences of presentations at the University of Cincinnati Philosophy Colloquium on the Churchlands and the 2008 meeting of Toward a Science of Consciousness in Tucson, Arizona.

References

Alter, T. (1998). A limited defense of the knowledge argument. *Philosophical Studies, 90*, 35–56. doi: 10.1023/A:1004290020847.

Alter, T. (2008). Phenomenal knowledge without experience. In E. Wright (Ed.), *The case for qualia*. Cambridge, MA: MIT Press.

Bechtel, W. (in press). The epistemology of evidence in cognitive neuroscience. In R. Skipper Jr., C. Allen, R. A. Ankeny, C. F. Craver, L. Darden, G. Mikkelson, & R. Richardson (eds.), *Philosophy and the life sciences: A reader. Cambridge.* MA: MIT Press.

Beisecker, D. (2005). Phenomenal consciousness, sense impressions, and the logic of 'what it's like. *Consciousness and Emotion*, 137–153.

Block, N. (1986). Advertisement for a semantics for psychology. In P. French, T. Uehling, & H. Wettstein (Eds.), *Midwest studies in philosophy, Vol. 10, studies in the philosophy of mind* (pp. 615–678). Minneapolis: University of Minnesota Press.

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

Chalmers, D. (2003a). Consciousness and its place in nature. In S. Stich & F. Warfield (Eds.), *Blackwell guide to philosophy of mind*. Oxford: Blackwell.

Chalmers, D. (2003b). The content and epistemology of phenomenal belief. In Q. Smith & A. Jokic (Eds.), *Consciousness: New philosophical essays* (pp. 220–272). Oxford: Oxford University Press.

Cummins, R. (1996). *Representations, targets, and attitudes*. Cambridge, MA: MIT Press.

Davidson, D. (1987). Knowing one's own mind. In *Proceedings and Addresses of the American Philosophical Association* (Vol. 60). pp. 441–458.

Dennett, D. (2007). What RoboMary knows. In T. Alter & S. Walter (Eds.), *Phenomenal concepts and phenomenal knowledge* (pp. 15–32). Oxford: Oxford University Press.

Fodor, J. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind*. Cambridge, MA: MIT Press.

Gertler, B. (2001). Introspecting phenomenal states. *Philosophy and Phenomenological Research, 63*, 305–328. doi:10.1111/j.1933-1592.2001.tb00105.x.

Jackson, F. (1982). Epiphenomenal qualia. *The Philosophical Quarterly, 32*, 127–136. doi:10.2307/2960077.

Kripke, S. (1980). *Naming and necessity*. Cambridge, MA: Harvard University Press.

Mandik, P. (2009, in preparation). The neurophilosophy of subjectivity. In J. Bickle (Ed.), *The Oxford handbook of philosophy and neuroscience*. Oxford: Oxford University Press.

Mandik, P., & Weisberg, J. (2008). Type-Q materialism. In C. B. Wrenn (Ed.), *Naturalism, reference and ontology: Essays in honor of Roger F. Gibson*. New York: Peter Lang Publishing Group.

Millikan, R. (1984). *Language, thought, and other biological categories*. Cambridge, MA: MIT Press.

Papineau, D. (2002). *Thinking about consciousness*. Oxford: Oxford University Press.

Papineau, D. (2007). Phenomenal and perceptual concepts. In T. Alter & S. Walter (Eds.), *Phenomenal concepts and phenomenal knowledge*. Oxford: Oxford University Press.

Putnam, H. (1981). *Reason, truth, and history*. Cambridge: Cambridge University Press.

Russell, B. (1905). On denoting. *Mind, XIV*(4), 479–493.

Schwitzgebel, E. (2008). The unreliability of naive introspection. *Philosophical Review, 117*, 245–273.

Sellars, W. (1953). Inference and meaning. *Mind, LXII*(247), 313–338.

Watson, R. (1995). *Representational ideas: From Plato to Patricia Churchland*. Dordrecht, Netherlands: Kluwer Academic Publishers.