

## Does Mary know I experience plus rather than quus? A new hard problem

Philip Goff

Published online: 3 March 2011  
© Springer Science+Business Media B.V. 2011

**Abstract** Realism about cognitive or semantic phenomenology, the view that certain conscious states are intrinsically such as to ground thought or understanding, is increasingly being taken seriously in analytic philosophy. The principle aim of this paper is to argue that it is extremely difficult to be a physicalist about cognitive phenomenology. The general trend in later 20th century/early 21st century philosophy of mind has been to account for the content of thought in terms of facts outside the head of the thinker at the time of thought, e.g. in terms of causal relations between thinker and world, or in terms of the natural purposes for which mental representations have developed. However, on the assumption that consciousness is constitutively realised by what is going on inside the head of a thinker at the time of experience, the content of cognitive phenomenology cannot be accounted for in this way. Furthermore, any internalist account of content is particularly susceptible to Kripkensteinian rule following worries. It seems that if someone knew all the physical facts about what is going on in my head at the time I was having a given experience with cognitive phenomenology, they would not thereby know whether that state had ‘straight’ rather than ‘quus-like’ content, e.g. whether the experience was intrinsically such as the ground the thought that two plus two equals four or intrinsically such as to ground the thought that two quus two equals four. The project of naturalising consciousness is much harder for realists about cognitive phenomenology.

**Keywords** Consciousness · Hard problem · The knowledge argument · Cognitive phenomenology · Phenomenal intentionality

---

P. Goff (✉)  
University of Hertfordshire, Hertfordshire, UK  
e-mail: philipgoff@gmail.com

P. Goff  
King’s College London, London, UK

The standard examples used to explain what phenomenal consciousness is are things like *what it's like to see red*, *what it's like to feel pain*, *what it's like to taste marmite*. Such 'raw feels' have little to do with thought or understanding. But a commitment to *cognitive phenomenology*, conscious states which are intrinsically such as to ground thought or understanding—a view traditionally treated with scorn in analytic philosophy—is increasingly being taken seriously.<sup>1</sup>

There are well known difficulties facing the physicalist wanting to give an account of raw feels. The aim of this paper is to show that there is a further, quite distinct challenge for the physicalist wanting to give an account of cognitive phenomenology. A commitment to the existence of cognitive phenomenology brings with it severe, perhaps insuperable, difficulties for the physicalist.

In the Sect. 1 I will build a *prima facie* case for a commitment to *semantic phenomenology*, for the view that perceptual experience represents words as having specific meanings, perhaps the most plausible form of realism about cognitive phenomenology.<sup>2</sup> In the Sect. 2 I will tell a new story about Mary, intended to illustrate the difficulties involved in accounting for semantic phenomenology physicalistically. In the Sect. 3, I will give the moral of the tale.

## 1 A *prima facie* case for the reality of semantic phenomenology

There is semantic phenomenology: perceptual experience represents words as having specific meanings. As you look around you, your experience does not present you with brute shapes and colours, which you then try to make sense of. Rather you immediately experience things *as* tables, doors, people, faces with expression; so much is part of the character of conscious perceptual experience. In the same way, as you listen to someone speaking a language you understand, your experience does not present you with meaningless phonemes which you then interpret. Rather you immediately experience the words *as* meaningful. So much is basic phenomenological fact, apparent to introspection.

If I say to you, 'God is a friend to all', you have an experience as of somebody telling you *that God is a friend to all*. You could not convey to somebody the character of your experience at the time I am saying this to you without somehow conveying (although of course not necessarily asserting) the proposition *that God is a friend to all*. To experience an utterance of 'God as a friend to all' as the claim that God is a friend to all, involves experiencing the words contained within that sentence as having specific meanings, as contributing to the making of that claim in specific ways.

<sup>1</sup> For the past 15 years, there has been a slowly growing minority defending this view, e.g. Strawson (1994), Siewert (1998), Horgan and Tienson (2002), Pitt (2004). I expect the two forthcoming Oxford University Press collections by Bayne and Montague and Horgan and Kriegel to bring the view much more into the mainstream.

<sup>2</sup> Semantic phenomenology is a form of cognitive phenomenology, as it is intrinsically such as to ground linguistic understanding.

Contrast your experience with that of a monolingual Japanese person, call her Ayuko, as I say to both of you: ‘God is a friend to all’. Even though your respective perceptual experiences may be similar in many ways—they may be indiscernible in terms of their visual representational content—the fact that only one of you understands the sentence entails a phenomenological difference; what it is like to hear a language you understand is very different from what it is like to hear a language you don’t understand.<sup>3</sup> Once we admit that there is a phenomenological difference between you and Ayuko as I say to both of you ‘God is a friend to all’, one that is entailed by the fact that you but not she understands this sentence of English, the question is how to characterise that difference. The options seem to be as follows:

1. Both you and Ayuko experience the same phonemes, but only your experience involves a feeling of understanding.
2. Both you and Ayuko experience the same phonemes, but only you experience the utterance of those phonemes as the making of a claim, a claim which the representational content of your experience is silent on the specifics of.
3. Both you and Ayuko experience the same phonemes, but only you experience the utterance of those phonemes as the making of some claim in a certain range, which includes the claim *that God is a friend to all*, but also includes other claims, perhaps certain ‘quus-like’ variants, e.g. *that God is a friend to all unless it’s Tuesday in Timbuktu in which case God is evil, that God is a friend to all unless it’s after 2012 in which case God is an orange*.
4. Both you and Ayuko experience the same phonemes, but only you experience the utterance of those phonemes as the claim *that God is a friend to all*.

The first two options are to be rejected on the grounds that such a difference in phenomenology is not entailed by the fact that you but not she understands the sentence of English being spoken (remember we are looking for a phenomenological difference which is entailed by this difference in understanding); we can suppose that Ayuko has a veridical experience as of someone making a claim, just not one she understands, and if she is also delusional in certain ways, her failure to understand may be accompanied by a feeling of understanding. I am not sure whether option 3 describes intelligible phenomenology, but in any case I take it that, out of 3 and 4, 4 is by far the more plausible account of the phenomenological difference in question. Option 4 entails the existence of semantic phenomenology: to experience an utterance of ‘God is a friend to all’ as the claim *that God is a friend to all* involves experiencing the words contained within that sentence as having specific meanings (more on this below).

The following, then, is a powerful argument for semantic phenomenology:

*Premise* For any two people attentively listening to a sentence, such that one but not the other understands that sentence, there is a phenomenological difference

<sup>3</sup> Strawson (1994) makes use of this kind of comparison between hearing a language you understand and hearing a language you don’t understand. Siewert (forthcoming) develops the idea into an extremely rigorous argument.

between the auditory perceptual experiences of those people, a difference which is entailed by the fact that one but not the other understands the sentence.

*Premise* Any plausible characterisation of this phenomenological difference entails the existence of semantic phenomenology.

*Conclusion* There is semantic phenomenology.

I have already said twice that experiencing the utterance of a sentence as the making of a claim involves experiencing the words within that sentence as having specific meanings. I will now explain a tad more clearly what this amounts to.

One of Frege's great contributions to the philosophy of language was the insight that the sentence is explanatorily prior to the word. This priority is reflected in the representational structure of semantic phenomenology. When we consciously perceive someone using an assertoric sentence in a language we understand, our perceptual experience represents the speaker as making a certain claim; to return to the example, if I say to you 'God is a friend to all', your perceptual experience represents the utterance of that sentence as an act of literally claiming that God is a friend to all. This perceptual experience involves hearing the words in that sentence as meaningful, and this in turn is a matter of experience representing those words as contributing to the making of the claim in specific ways. Your experience represents the word 'God' as determining that God is the subject of the claim. Your experience represents the predicate 'is a friend to all' as determining that the utterance claims of the subject of the sentence that it is a friend to all.<sup>4</sup>

How is a *mere conscious experience*, something akin to 'headaches, tickles and nausea',<sup>5</sup> able to represent so much? This question, and the worry it expresses, sound reasonable only because our philosophical tradition, in so far as it accepts the reality of conscious experience at all, has become used to dealing with a caricature of consciousness: what it's like to see red, what it's like to taste lemons, what it's like to feel pain. It is customary for philosophers to talk as though such 'raw feels' exhausted the nature of consciousness.<sup>6</sup> But it is crazy to think that conscious experience is exhausted by raw feels. Even if we stick to perceptual experience, its representational powers vastly outstrip what the crude raw feels model allows for.

We experience sad faces, angry rampages, gluttonous displays of eating, joyous dancing, not just in the sense that such things are the objects of our experience, but in the sense that our experience represents them as such. We see the lion as about to pounce, the huge man as capable of overpowering, the mother's caress as expressing

<sup>4</sup> Of course we also perceive words in isolation from sentences, but even in these contexts what is represented is a matter of the capacities of words to be used in whole sentences. To experience 'God' as meaning *God* is to have experience which represents 'God' as *capable* of contributing to the making of claims (and other kinds of speech act, but I focus on assertions for the sake of simplicity) in certain specific ways, e.g. as capable of being used to make claims about God. The example I will be considering in the second section will be one in which a word is heard within a complete sentence (see footnote 9) and so avoids concerns about how much content is represented when a word is heard in isolation from a sentence.

<sup>5</sup> As Kripke (1982, p. 43) describes consciousness.

<sup>6</sup> Although there are voices against this tendency, e.g. Strawson (1994), Siewert (1998), Horgan and Tienson (2002), Pitt (2004). Regarding perceptual experience, representationalists are also opposed to this tendency, e.g. Dretske (1995), Tye (2009), Lycan (1996), Carruthers (2000), although they tend to be less sympathetic to realism about cognitive or semantic phenomenology.

compassion. If I am having an experience as of a face filled with angst, then any phenomenal duplicate of mine—my brain in a vat twin or my Cartesian ego twin—is also having an experience as of a face filled with angst. Our perceptual experience is *conceptually saturated*, to borrow a phrase from Strawson (1979). Being a functional duplicate of a human being, but one whose conscious experience was limited to raw feels, would be little better than being a zombie. Once the depth and remarkable versatility of human consciousness is fully appreciated, there is little difficulty in allowing that human conscious experience is capable of representing the semantic properties of words, complex though they are.

The claim is not that *all* the semantic properties of words are represented in conscious experience; to claim this would be to deny the persuasive intuitions which support some degree of semantic externalism, e.g. Putnam's Twin Earth thought experiments (Putnam 1975) (at least on the assumption that conscious experience is constitutively determined by what goes on inside the head). But we can accommodate Twin Earth intuitions, together with a commitment to semantic phenomenology, by distinguishing *narrow semantic content* from *broad semantic content*, and claiming that only the former is representing in perceptual experience of language. One way of making such a distinction is in terms of Jackson's and Chalmers' version of two-dimensional semantics (Chalmers 1996; Jackson 1998).

For example, we may claim that the assertion that I represent a speaker as making when they say 'Water is tasty' is that expressed by the *primary intension* associated with this sentence: roughly the assertion that the actual colourless, odourless stuff in oceans and lakes is transparent—this can be modelled as the set of worlds considered as actual where the colourless, odourless stuff in oceans and lakes is tasty. Given that the actual colourless, odourless stuff is H<sub>2</sub>O, the sentence is also thereby associated with a secondary intension, which can be modelled as the set of possible worlds considered as counterfactual where H<sub>2</sub>O is tasty. It is plausible to think that this latter intension is not represented in semantic phenomenology (at least not in the semantic phenomenology of a competent speaker ignorant of the facts of chemistry). Distinguishing between broad and narrow content according to the two-dimensional framework is just one way in which we can reconcile Twin Earth intuitions with the reality of semantic phenomenology.<sup>7</sup>

I have so far only talked about semantic phenomenology in so far as it concerns the representation of assertions. This is just to brush the surface; semantic phenomenology is at least as complex a phenomenon as natural language, and much, much more can be said about it. Moreover, I don't take what I have said here to be conclusive, or to constitute a full and adequate account of the phenomenon; the reality of semantic phenomenology has been defended elsewhere with much greater depth and rigour.<sup>8</sup> My aim in this section was merely to build a *prima facie* case for the reality of semantic phenomenology. I now proceed to the main purpose of the paper, which is to illustrate how difficult it is to account for semantic phenomenology in a physicalistically acceptable way.

<sup>7</sup> A similar strategy could be undertaken to accommodate semantic deference, building the deference to experts into the primary intension.

<sup>8</sup> Strawson (1994, 2008), Horgan and Tienson (2002), Pitt (2004), Siewart (1998, forthcoming).

## 2 Mary strikes back

Once upon a time, there was an evil demon. As well as being evil, this demon was quite philosophically curious. One day he woke up and thought to himself: ‘Right, I really need to figure out once and for all whether physicalism is true’. This particular philosophical issue had preoccupied the demon for some time. A few years earlier he had performed a little experiment to try to settle the matter. He had found a highly intelligent female baby called ‘Mary’ abandoned in a parking lot in Tucson Arizona, and had whisked her off to his realm and raised her in a black and white room, providing for all her needs, and teaching her, by means of black and white text books in conjunction with lectures which he performed covered head to toe in a black and white robe, all the physical facts about colour vision. On her 18th birthday the demon brought Mary out of the room and showed her a red rose for the first time. Upon seeing it she exclaimed, ‘Oh, so that’s what it’s like to see red!’

At first the demon thought this settled the matter. It seemed clear that, upon seeing the rose, Mary learnt something new: she learnt *what it’s like to see red*. Before she left her black and white room Mary already knew all the physical facts about colour vision, so the new fact she learnt upon seeing the rose must be a non-physical fact about colour vision. Therefore, there are non-physical facts; therefore, physicalism is false. But soon the demon began to have his doubts. Perhaps, he thought, what Mary learnt when she saw the rose was just a matter of *know how* rather than *knowledge that*, or perhaps she learnt a new way of conceiving of a fact she had known already. Both of these hypotheses are consistent with physicalism.

‘What I need’, the demon thought, ‘is a new experiment, one that’ll settle the matter beyond doubt!’ The demon clicked his fingers, and, suddenly Mary appeared before him. ‘Right, young Mary’, said the demon, ‘I have a new mission for you. There is this young Welsh chap, Cuthbert, who lives in Aberystwyth. Your mission is to try to work out the nature of Cuthbert’s semantic phenomenology. Specifically I want you to try work out what Cuthbert’s perceptual experience represents ‘plus’ as meaning. Now I have a hunch that Cuthbert experiences “plus” to mean plus. But that’s just a hunch. I want you to work it out for yourself.

‘The experiment is going to take place in two stages. In the first stage you will have access only to the physical facts. And because consciousness is inside the head, you’re only going to need access to a certain limited set of facts. I will take a certain time  $t$ , during which Cuthbert is consciously perceiving someone saying ‘plus’, and reveal to you all the facts about Cuthbert’s intrinsic physical nature at  $t$ . In the first stage of the experiment you must see what you can work out about Cuthbert’s semantic phenomenology from these facts. During the second stage of the experiment I will reveal to you all the phenomenal facts concerning Cuthbert at  $t$ , i.e. everything there is to know about what it’s like to be Cuthbert at  $t$ .<sup>9</sup>

‘So let’s begin’, the demon cried and clicked his fingers. Suddenly a big book appeared which contained all the physical facts concerning Cuthbert at  $t$ . Mary set to work. ‘A difficult task’, she thought, despite her extensive neurophysiological

<sup>9</sup> We can take it that  $t$  is some time during which Cuthbert hears a full sentence containing ‘plus’ with understanding.

knowledge, ‘The trouble is knowing where to start...what kind of physical facts are going to reveal to me Cuthbert’s semantic phenomenology?’ A long period of mulling and pacing later, the only option which suggested itself was to turn to the physical facts which determine meaning or semantic content. So Mary went to the demon’s rather extensive library and pulled out a load of books on philosophy of language and theory of meaning.

Now if there’s one thing you could say about Mary, she could sure read fast, especially when she knew what she was looking for. She zoomed through Davidson and Dretske, Millikan and McDowell, pulling out what she needed. Under the title of ‘potential meaning determining facts’, she made a big list of all the facts which philosophers have taken to determine meaning or semantic content. She found that her list divided neatly into two categories: potential meaning determining facts outside of the head of the thinker/speaker, and potential meaning determining facts inside of the head of the thinker/speaker. In the first category were: facts about teleology, facts about causal relations or asymmetric causal dependencies, fact about which properties are natural, facts about usage/behaviour over time (as opposed to dispositions to use/behave), facts about the linguistic community at large. In the second category were: behavioural dispositions and narrow functional states.

Mary thought: ‘Well the first category is no use to me, as I only have access to what’s under Cuthbert’s skin at  $t$ . So I need to work with Cuthbert’s dispositions and narrow functional states’. There seemed to be progress; Mary felt a rush. ‘Right’, she thought, ‘I need to get methodical about this. What I need is a general theory of how semantic phenomenology is determined by physical facts, and then I need to apply that theory in the specific case of Cuthbert at  $t$ ’. Mary racked her brains some more: ‘If semantic phenomenology is determined by functional states or behavioural dispositions, then it must be something to do with how the thinker is disposed to use language’. And so Mary formed the following hypothesis:

**Mary’s general theory** How a thinker is disposed to experience a term is determined by how the thinker is disposed to use the term.

Now she needed to apply this general theory in the case of Cuthbert. Remembering that the demon had had a hunch that Cuthbert experienced ‘plus’ to mean plus, Mary formed the following hypothesis:

**Mary’s first hypothesis** Cuthbert is disposed such that, if he applies ‘plus’ to numbers  $x$  and  $y$ , he will give the sum of  $x$  and  $y$ , and hence he is disposed to experience ‘plus’ as meaning plus.

Mary set about testing her hypothesis. At first things worked out pretty well. Looking at the details of the workings of Cuthbert’s brain, she discovered that if Cuthbert applied ‘plus’ to 2 and 2, he would give the answer 4. If Cuthbert applied ‘plus’ to 50 and 50, he would give the answer 100. But she soon began to encounter problems. As it happened, Cuthbert was a pretty dodgy mathematician. He was fine with numbers under 1000, but with numbers greater than 1000 he was systematically in error. Sometimes he fluked on the right answer, but equally often he was 5 numbers out either way. And of course this was reflected in what Mary found in Cuthbert’s dispositions; Mary discovered that:

For numbers  $x$  and  $y$  greater than 1000, such that Cuthbert applies ‘plus’ to those numbers, Cuthbert is disposed to give, not the sum of  $x$  and  $y$ , but a random number between 5 below the sum of  $x$  and  $y$  and 5 above the sum of  $x$  and  $y$ .

This discovery falsified Mary’s first hypothesis: it entails that Cuthbert is not disposed such that if he applies ‘plus’ to two numbers, he will give their sum.

Despite having her first hypothesis falsified, Mary still wanted to stick to her overall theory. She decided the demon’s hunch must have been wrong: Cuthbert does not experience ‘plus’ to mean plus, rather he experiences ‘plus’ to mean *quus*, a function defined as follows:

$$\begin{aligned} x \text{ quus } y &= x + y && \text{if } x, y < \text{ a thousand} \\ &= \text{ a random number between} \\ & (x + y) - 5 \text{ and } (x + y) + 5, && \text{otherwise} \end{aligned}$$

Thus, Mary formulated her second hypothesis:

**Mary’s second hypothesis** Cuthbert is disposed such that, if he applies ‘plus’ to numbers  $x$  and  $y$  such that  $x$  and  $y$  are less than 1000, he will give the sum of  $x$  and  $y$ , and if he applies ‘plus’ to numbers  $x$  and  $y$  such that  $x$  and  $y$  are greater than 1000, he will give a random number between  $(x + y) - 5$  and  $(x + y) + 5$ , and hence experiences ‘plus’ to mean *quus*.

‘Time to test’, thought Mary, confident of greater success this time. Again, things went well to start off with. She worked out that if Cuthbert applied ‘plus’ to 1000 and 1000, he would give a random number 5 either side of 2000. If he applied ‘plus’ to 1,000,000 and 1,000,000, he would give a random number 5 either side of 2,000,000. But before long Mary began to encounter more difficulties.

Sadly, Cuthbert’s brain was a finite system. And due to its finitude, there were certain numbers, numbers greater than a gillion, which were just too huge for Cuthbert to perform functions on. This was problematic for her second hypothesis, for there was no fact of the matter as to what Cuthbert would do if he applied ‘plus’ to numbers greater than a gillion, given that it was physically impossible for him to perform functions on such numbers. So Mary now found that she had no way of knowing whether Cuthbert experiences ‘plus’ to mean *quus* or *zuus*, where *zuus* is a function defined as follows:

$$\begin{aligned} x \text{ zuus } y &= x \text{ quus } y && \text{if } x, y < \text{ a gillion} \\ &= 5 && \text{otherwise} \end{aligned}$$

Mary was stuck. She racked her brains harder than they had hitherto been racked. ‘Maybe I’ve been working with too crude a notion of a disposition’, she mused in desperation, ‘Perhaps there are truths about what Cuthbert *would* do in certain counterfactual situations, e.g. if his brain were complicated enough to perform functions on numbers greater than a gillion, and maybe these counterfactual truths determine his semantic phenomenology’. Mary tried her best to make this new strategy work, but when she thought about it she realised that there were a great many possible worlds where Cuthbert’s brain was more complicated such that he could perform functions on numbers greater than a gillion, and in which he applies

‘plus’ to two such numbers. In some of those worlds, call them the ‘plus worlds’, Cuthbert gave the sum of those numbers. In some of those worlds, call them the ‘quus worlds’, Cuthbert gave a random number five either side of the sum of those numbers. In some or those worlds, call them the ‘zuus worlds’, Cuthbert gave the answer 5.

The question Mary needed to know the answer to was: Which of those worlds are accessible in evaluating the relevant counterfactuals? The trouble was that, on reflection, Mary realised that which worlds are accessible is dependent on how she interprets the physical system that is Cuthbert’s brain. If she interpreted that system as performing the *plus function*, then the *plus worlds* became accessible. If she interpreted that system as performing the *quus function*, then the *quus worlds* became accessible. If she interpreted that system as performing the *zuus function*, then the *zuus worlds* became accessible. But nothing to be found in the physical facts could tell her which of these interpretations was the correct one. Mary was well and truly stuck.

At this point the demon returned and asked Mary how she was getting on. ‘I’m so sorry’, Mary pleaded, avoiding the demon’s gaze, ‘From the facts you have given me, there is just no way of knowing whether Cuthbert experiences “plus” to mean plus or quus or zuus. I’m afraid I have failed’. ‘Don’t worry!’ said the demon, ‘that was just the first part of the experiment. It’s now time for the second’. He clicked his fingers, and instantly Mary knew everything there was to know about what it’s like to be Cuthbert at *t*. She now had access to Cuthbert’s semantic phenomenology at *t*; to what meaning Cuthbert’s perceptual experience represented the word he was hearing at *t* as having. Mary had only to reflect for a moment before declaring, ‘Cuthbert experiences ‘plus’ to mean plus’. The experiment was over.

The demon went home and wrote up the experiment. He read over it and reflected. ‘Well that’s it’, he thought, ‘I’ve finally shown conclusively that physicalism is false. In the second part of the experiment Mary learnt a new fact about Cuthbert at *t*: she learnt the meaning Cuthbert’s perceptual experience represents ‘plus’ as having. In the first part of the experiment she already knew all the *physical* facts about Cuthbert at *t*, so this new fact must be a *non-physical* fact about Cuthbert at *t*. Therefore there are non-physical facts; therefore physicalism is false’.

In the first experiment the demon had come to doubt the findings of his experiment, on the grounds that he couldn’t rule out that Mary had merely learnt new know how, or new ways of thinking about a fact she already knew. But in the case of this new experiment, these options seemed implausible. It was impossible to deny that, in learning that Cuthbert’s perceptual experience represents ‘plus’ to mean plus, rather than quus or zuus, Mary learnt *new information*, in the sense of ruling out *genuine possibilities*, i.e. the possibility that Cuthbert’s experience represents ‘plus’ to mean quus, and the possibility that Cuthbert’s perceptual experience represents ‘plus’ to mean zuus.

And so, philosophically satisfied, the demon decided to get an early night. The end.<sup>10</sup>

<sup>10</sup> This story self-consciously steals and merges ideas from Jackson (1982, 1986) and Kripke (1982).

### 3 The moral of the story

The general trend in late 20th century/early 21st century has been to account for the content of thought and language ‘outside the head’, i.e. in terms of facts outside of the skull of the thinker at the time it’s thinking a particular thought or understanding a particular token sentence. Information theorists explain mental representation in terms of causal relationships between internal states and features of the environment; obviously such causal relationships extend outside the head of the thinker (Dretske 1988, 1994; Fodor 1990).<sup>11</sup> Teleo-semanticist accounts explain the content of thought in terms of the natural functions of representations, or the natural functions of systems which make use of representations; such natural functions are dependent on the history of the organism or of its species. Again, the explananda involves facts outside of the organism at a given time of employing a particular representation (Millikan 1984, 1989; Papineau 1984, 1993). David Lewis (1983) suggests we supplement a use based theory of meaning with a role for degree of naturalness of properties in determining content; at least in most cases the properties whose degree of naturalness is relevant for a certain thinker of a certain thought, will not be instantiated in the brain of that thinker at the time of that thought.<sup>12</sup> All these theories go beyond the dispositionalist, or merely use based, accounts of meaning which Kripke considers in his discussion of Wittgenstein and rule following, and hence are plausibly not subject to the concerns his discussion raises for such accounts.

However, on the assumption that consciousness is ‘inside the head’, i.e. constitutively determined by what is going on inside the head of the thinker, externalist theories cannot account for its existence or nature. If my semantic phenomenology is determined by what’s going on in my head right now, then my evolutionary history, or causal relations between me and the world, or facts about how natural the candidate objects of thought are, have no direct bearing on it (although they may of course have indirect bearing on it by shaping what goes on inside my head). If consciousness is inside the head, then, in explaining semantic phenomenology, we must confine ourselves to facts about what’s going on inside the head; to what me and my brain in a vat twin have in common.<sup>13</sup>

The trouble is that any physicalist account of the content which restricts itself to what’s going on inside the head, is particularly susceptible to rule following worries.

<sup>11</sup> The difficulty for information based accounts is finding some way of account for *misrepresentations*, for situations where a representation is caused by  $x$  but does not represent  $x$ , e.g. when my thought that there is a cow in the field is caused by a horse. Dretske (1988, 1994) does this by adding teleology to his account. Fodor (1990) offers a non-teleological way of distinguishing causes of a representation which the representation represents, from causes of a representation which the representation doesn’t represent.

<sup>12</sup> Even if the relevant properties are instantiated in the brain, their being instantiated in the brain is not essential to their role in fixing content. In contrast, I take it that, for those who believe that consciousness is ‘inside the head’, it is essential that the properties which realise consciousness are instantiated where the subject of experience is.

<sup>13</sup> If the facts about which properties are natural and to what degree obtain in all possible worlds, then there will be no difference, on Lewis’s view, between my brain in a vat twin and I vis-à-vis content determining facts. However, content is still ‘outside the head’, in the sense that it is not *constitutively determined* by goings on inside the head.

Any physical system that can be interpreted as representing *that two plus two*, can also be interpreted as representing *that two quus two equals four*. It is difficult to see what in the intrinsic physical facts about a system could speak in favour of ‘straight’ rather than a ‘quus-like’ interpretation. Perhaps if there were contingent psycho-physical laws of nature which entailed that physical systems realise ‘straight’ rather than ‘quus-like’ semantic phenomenology, then we might explain how semantic phenomenology is determined by the physical facts (in conjunction with such laws). However, physicalism, by definition, is (or at least involves) the view that all facts supervene with metaphysical necessity on the physical facts, rather than being the view that all facts supervene with merely nomic necessity on the physical facts. Therefore, physicalists must try to explain how the physical facts determine semantic phenomenology without appeal to psycho-physical laws of nature which obtain in some worlds but not others. But without such an appeal, it is hard to see how we can give an internalist physicalist account of why semantic phenomenology has, as it generally does have, straight rather than quus-like content.<sup>14</sup>

We can put the difficulty in the form of the following anti-physicalist argument:

1. There is semantic phenomenology, with straight rather than quus-like content.
2. If there is semantic phenomenology, then it is determined by what’s going on inside the head (as consciousness is constitutively determined by what’s going on inside the head).
3. No physical facts concerning what’s going on inside the head fully determine semantic phenomenology (as no physical facts inside the head can determine that my semantic phenomenology has straight rather than quus-like content), at least not independently of contingent psycho-physical laws of nature.
4. Therefore, no physical facts, independently of contingent psycho-physical laws of nature, fully determine semantic phenomenology (by 2 and 3).
5. Physicalism is false (because 1 and 4 entail that there is some aspect of my conscious experience which the physical facts do not fully—independently of contingent psycho-physical laws of nature—determine, which is inconsistent with physicalism).

If there is semantic phenomenology, then such phenomenology, contra Kripke, solves the rule following challenge, at least in the sense that it determines that the meaning represented in semantic phenomenology is straight rather than quus-like.<sup>15</sup> Even so, the rest of Kripke’s discussion of rule following makes a strong case that internalist physical facts cannot determine that we have a case of straight rather than quus-like content. And if that is the case, then we are unable to explain the straightness of the phenomenology in terms of the physical facts. We no longer have a puzzle as to how the semantic facts are determined—this is done by the semantic phenomenology—but we now have a puzzle as to how the semantic phenomenology is determined.

<sup>14</sup> Only narrow content is represented in semantic phenomenology, but we still need an account of what makes the narrow content straight rather than quus-like.

<sup>15</sup> More work would need to be done to get from *Jack experiences ‘plus’ to mean plus* to *Jack means ‘plus’ by plus*.

There are several ways one can respond to this argument. One might accept it as a sound argument against physicalism. Or one might ‘tollenise’ the above into an argument against the existence of semantic phenomenology: all conscious experience is determined by the physical facts, if there were semantic phenomenology it wouldn’t be determined by the physical facts, therefore there is no semantic phenomenology. Of course, this argument against semantic phenomenology would need to be balanced against the arguments in favour of semantic phenomenology. For those philosophers for whom the existence of semantic phenomenology is introspectively evident, the case for eliminativism about semantic phenomenology on the grounds that we can’t give a physicalist account of it will have little or no force (it would be like the physicalist who responds to the knowledge argument by denying the existence of colour experience).

Alternately, one might try to hold on to both semantic phenomenology and physicalism by denying either premise 2 or premise 3. I will consider both of these options in turn. From one perspective, denying premise 2 is a very attractive option. If we accept that semantic phenomenology is, at least in part, determined by facts outside the head of the thinker, then we are free to adopt an externalist theory of content to account for semantic phenomenology. We might, for example, explain the fact that I experience ‘plus’ as plus rather than quus in terms of the natural purpose for which my plus concept has developed. The trouble is, that premise 2 is a very plausible premise indeed, and denying it is a huge intuitive cost.

The ideal for the physicalist wanting to embrace semantic phenomenology would be to deny premise 3. But doing so requires finding a plausible internalist physicalist solution to the rule-following challenge (in order to give a physicalist account of the fact that semantic phenomenology has straight rather than quus-like content). This is a tall order. Once we accept the existence of cognitive phenomenology, the Kripkenstein rule following considerations, in conjunction with the premise that consciousness is inside the head, constitute a powerful argument against physicalism.

I have focused in this paper on semantic phenomenology. However, the same difficulty will arise for any kind of cognitive phenomenology. Suppose we believe in a phenomenology of thought: there is a conscious state which is intrinsically such as to ground the thought *that two plus two equals four*. On the assumption that consciousness is inside the head, the physicalist will need to give an account, in terms of the physical goings on inside the head of the thinker, of what makes it the case this experience is intrinsically such as to ground the thought *that two plus two equals four* rather than intrinsically such as to ground the thought *that two quus two equals four*. Again, rule following considerations make this very difficult, perhaps impossible, to do.

I conclude that realism about cognitive phenomenology and physicalism are difficult, perhaps impossible, to marry.

**Acknowledgment** I would like to thank David Papineau, Barry Smith, Meredith Williams, David Chalmers, Robert van Gulick, William Seager, Chris Schriener, Emma Bullock, Kirk Surgener, Constantine Sandis, Stephen Boulter, Holly Lawford-Smith and Rory Madden for comments and discussion. I wrote this paper as a Research Fellow with the AHRC project ‘Phenomenal Qualities’.

## References

- Bayne, T., & Montague, M. (Eds.). (forthcoming). *Cognitive phenomenology*. Oxford: Oxford University Press.
- Carruthers, P. (2000). *Phenomenal consciousness: A naturalistic theory*. Cambridge: Cambridge University Press.
- Chalmers, D. J. (1996). *The conscious mind: Towards a fundamental theory*. Oxford: Oxford University Press.
- Dretske, F. (1988). *Explaining behaviour*. Cambridge, MA: MIT Press.
- Dretske, F. (1994). If you can't make one, you don't know how it works. *Midwest Studies in Philosophy*, 19, 468–482.
- Dretske, F. (1995). *Naturalizing the mind*. Cambridge, MA: MIT Press.
- Fodor, J. (1990). *Theory of content and other essays*. Cambridge, MA: MIT Press.
- Horgan, T., & Kriegal, U. (Eds.). (forthcoming). *Phenomenal intentionality*. Oxford: Oxford University Press.
- Horgan, T., & Tienson, J. (2002). The intentionality of phenomenology and the phenomenology of intentionality. In D. J. Chalmers (Ed.), *Philosophy of mind: Classical and contemporary readings*. Oxford, New York: Oxford University Press.
- Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly*, 32(127), 127–136.
- Jackson, F. (1986). What Mary didn't know. *Journal of Philosophy*, 83(5), 291–295.
- Jackson, F. (1998). *From metaphysics to ethics: A defence of conceptual analysis*. Oxford: Clarendon Press.
- Kripke, S. (1982). *Wittgenstein on rules and private language*. Oxford: Basil Blackwell.
- Lewis, D. (1983). New work for a theory of universals. *Australasian Journal of Philosophy*, 61(4), 343–377.
- Lycan, W. (1996). *Consciousness and experience*. Cambridge, MA: MIT Press.
- Millikan, R. (1984). *Language, thought and other biological categories*. Cambridge, MA: MIT Press.
- Millikan, R. (1989). Biosemantics. *Journal of Philosophy*, 86(6), 281–297.
- Papineau, D. (1984). Representation and explanation. *Philosophy of Science*, 51(4), 550–572.
- Papineau, D. (1993). *Philosophical naturalism*. Oxford: Blackwell.
- Pitt, D. (2004). The phenomenology of cognition or 'What is it like to think that P?'. *Philosophy and Phenomenological Research*, 69(1), 1–36.
- Putnam, H. (1975). *The meaning of meaning*. In *his language, meaning and reality*. Cambridge: Cambridge University Press.
- Siewart, C. (1998). *The significance of consciousness*. Princeton: Princeton University Press.
- Siewart, C. (forthcoming). Phenomenal thought, in Bayne and Montague.
- Strawson, P. F. (1979). Perception and its objects. In G. F. Macdonald (Ed.), *Perception and identity: Essays presented to A. J. Ayer*. London: Macmillan.
- Strawson, G. (1994). *Mental reality*. Cambridge, MA: MIT Press.
- Strawson, G. (2008). *Real intentionality 3: Why intentionality entails consciousness*. In *his real materialism and other essays*. Oxford: Oxford University Press.
- Tye, M. (2009). *Consciousness revisited*. Cambridge, MA: MIT Press.