Closing (or at least narrowing) the explanatory gap[1]

Katalin Farkas
Central European University, Vienna

Forthcoming in: Peter Anstey & David Braddon-Mitchell (eds.) *Armstrong's Materialist Theory of Mind* Oxford: Oxford University Press

1. Introduction

Rereading the Preface to the 1993 paperback edition of David Armstrong's *A Materialist Theory of the Mind* (originally published in 1968) brings back vividly the personality of the author for those, who, like me, knew him. Armstrong had a strong feeling that he was an important part of the history of philosophy[2]; when he describes his theory in relation to those of other important figures like J.J. C. Smart and U.T. Place, this reads like a 20th century account of the culmination of the development of the World Spirit (except that nothing is further from the author's intention than declaring that the world is moved by a mind!). The book is indeed one of the finest expressions of a relentless materialist (or physicalist – I'll use these terms interchangeably) vision of the mind, and Armstrong's Causal theory of the mind has proved to be a powerful framework to explain various mental phenomena. But philosophers who accept some version of materialism still often think that their theory leaves something unexplained about the nature of conscious experiences. This contrasts with theories about physical phenomena, which are not considered to leave something unexplained in the same way. This phenomenon is known as the 'explanatory gap'.

In this paper, I revisit the issue of the explanatory gap. In section 2, I will recall a version of this idea from an influential paper which originated the very term: Joseph Levine's 1983 paper 'Materialism and qualia: and the explanatory gap'. In sections 3 and 4, I will relate the issue to Armstrong's discussion of secondary qualities and bodily sensations, and to Locke's idea that only the ideas of primary qualities resemble the quality that causes them. In section 5, I argue that the specific explanatory question Levine asks is this: 'why this phenomenal character, rather than any other, is attached to this physiological process?'. In section 6 and 7 I argue that this question can be answered. First, we should look for a fit not between the realizer of a role and the phenomenal character, but between

[1] Versions of this paper were presented at the *Annual meeting of the Italian Society for Analytic Philosophy* in Novara, and at the colloquium series at CEU. I am grateful to the audiences for pertinent questions; to Laszlo E. Szabo for advice, and to the editors, Peter Anstey and David-Braddon Mitchell for valuable comments on a draft. Special thanks to Tim Crane for many many discussions on the topic.

2 If an anecdote is permitted here: in a conversation, Armstrong once expressed scepticism about the advantages of blind refereeing for journals. Someone objected: 'But David, if you submitted a paper to a journal and it contained a mistake, you'd prefer if that mistake was identified by a referee, and the paper wasn't published just because you wrote it?' Armstrong replied: 'If I make a mistake, it is an event of philosophical significance.' This was of course largely a joke – but not entirely, I think.

the role itself and the character. Further, there is a natural fit between the phenomenal character of experiences and their functional roles: for example pains feel inherently unpleasant, and that explains why they cause avoidance behaviour. In section 8, I sketch a possible background theory that explains this fit, which I call Phenomenal Functionalism: a view that is meant to be similar to Phenomenal Intentionality, and holds that the functional role of an experience is grounded in its phenomenal features. In section 9, I discuss some other possible gaps in our understanding the relationship between the mental and the physical, and conclude that the fit between functional role and phenomenal character goes a long way, though probably not the whole way, to closing the explanatory gap.

2. The explanatory gap

As Armstrong explains in the Preface, the Causal theory, his own materialist account of the mind, has two steps. First, we give an analysis of mental features, based on the insight that a mental feature is apt to cause certain ranges of behaviour, and it is apt for being produced by certain ranges of stimuli. Second, we claim that what is apt for causing and being caused this way is a physical state of the brain. That physical state is then identified with the mental state (Armstrong 1968, xiv). These were originally regarded as contingent identities, but subsequently, Saul Kripke (1972) influentially argued that identity statements were necessary, even when empirically discoverable. Reflecting on this development, Armstrong notes that Kripke's argument works only for rigid designators, whereas he intends the physical description of these states as non-rigid designators. Hence he thinks that Kripke's argument against materialism fails (Armstrong 1968, xiv).

Kripke's argument against materialism has indeed left many philosophers unconvinced, but a subsequent reflection on the identity statements in the focus of Kripke's inquiry, by Joseph Levine, proved to be very influential. In this section, I will recall Levine's well-known argument for an 'explanatory gap' between the physical and mental description of certain phenomena.

Levine opens his paper (1983) by recalling Kripke's argument against mental-physical identities, such as pain is C-fibre firing. Following Kripke (1972), Levine compares the following identity statements:

(1) Pain is C-fibre firing.
(2) Heat is the motion of molecules.

As it is well known, Kripke argues that the first is false, the second is true. Levine has some doubts about Kripke's argument; he thinks that the metaphysical thesis of the distinctness of mind and body is not conclusively proved by Kripke. However, he wants to make a different point in his paper. He claims that even if Kripke were wrong, and contrary to what he says, both statements were true, there would still be a significant difference between them. The difference is not metaphysical, but rather epistemic: it relates to the explanatory power of the statements. The second is fully explanatory, Levine claims, but the first leaves something crucial unexplained.

In setting up the problem, Levine considers a causal-functionalist account of heat and pain, which consists of the two steps of Armstrong's Causal theory, as sketched in the first paragraph of this section (though Levine doesn't particularly mention Armstrong). First, we identify the causal or functional role of pain and heat. Pain typically 'warns us of damage, it

causes us to attempt to avoid situations we believe will result in it, etc' (Levine 1983, 357) Heat is 'responsible for the expansion and contraction of mercury in thermometers, causes some gases to rise and others sink, etc.' (Levine 1983, 357). Second, we look for the state that occupies this causal role and we identify these states with the original phenomena: it's the motion of molecules in the case of heat, and C-fibre firing in the case of pain.

According to Levine, the resulting statement (2), the identification of heat with the motion of molecule, expresses an identity that is *fully explanatory.* First, prior to the discovery of identity, the causal role of heat exhausts our notion of it: what we understand by the phenomenon of heat is fully accounted for by the causal role it plays. Second, knowledge of chemistry and physics makes it intelligible how the motion of molecules plays the causal role associated with heat. So once we identified the causal role, and discovered what plays this causal role, everything is explained.

The case of pain and C-fibre firing is different, Levine claims. What is similar is this. The causal role of pain is at least part of our concept of pain. And in the case of pain too, after we identify the causal role of pain, we discover a mechanism, C-fibre firing, that plays this causal role. But there is a crucial difference: our notion of pain is *not exhausted* by the causal role. The causal role is part of the notion, but there is more, namely, the phenomenal or qualitative character of pain. And while our knowledge of physiology makes it intelligible how C-fibre firing plays the causal role associated with pain, there is nothing in our knowledge of physiology that explains why C-fibre firing gives rise to the phenomenal properties of pain. As Levine puts it: '… there seems to be nothing about C-fiber firing which makes it naturally "fit" the phenomenal properties of pain, any more than it would fit some other set of phenomenal properties' (Levine 1983, 357).

According to Levine, the same story goes for example for the sensations of colours. Sensations of red and green can are apt to be caused by certain stimuli, which we may call physical red and physical green (first step). Now consider the receptors and physical processes that are responsible for responding to physical red and physical green, and call these physical stories R and G (second step). R is supposed to give rise to, or correlate with, the phenomenal feel of red experiences, and G the phenomenal feel of green experiences. But there is no explanation for this in R or G: 'R doesn't really explain why I have one kind of qualitative experience … and not the other' (Levine 1983, 358). Support for this claim is provided by imagining an inversion between the two qualitative characters, while the physical stories remain the same. The inversion seems perfectly conceivable, and this underlines the point that the phenomenal feel of red has no more intelligible connection to R than the phenomenal feel of green.

Talking of colour inversion may bring Locke's name to mind, and indeed, Levine says that the same point was captured by Locke who thought that 'two sets of phenomena – corpuscular process and simple ideas – are stuck together in an arbitrary manner' (Levine 1983, p. 359). Levine is right to trace back this idea to Locke, and he may have also traced it further back to Descartes. Neither Descartes nor Locke supported an identity theory, but both dealt with the question of the relationship between the ideas of secondary (and tertiary) qualities, and the physical and physiological processes that precede or cause the formation of these ideas. Both Descartes and Locke claims that *it is simply ordered by God which conscious experience goes with which physical process*. Invoking God's will in this way means, for both of them, that there is no intelligible explanation of the connection – that is, explanation that is intelligible for us.

3. Secondary and tertiary qualities

Levine's paper (1983) was published in the period between the first edition (1968) and the paperback edition (1993) of *A Materialist Theory of the Mind*, but it isn't included among the philosophical developments that Armstrong chose to reflect on in the Preface. Nor there is a mention of the explanatory gap in Armstrong's 1999 *The Mind-Body Problem*. However, it is instructive to recall Armstrong's discussion of secondary qualities and bodily sensations and compare it to the issue of the explanatory gap.

Both pain and colour sensations are analyzed by Armstrong as perceptual experiences (Chapters 12 and 14 of 1968), and are given a materialist account within his general materialist theory of perception as an acquisition of beliefs. The particular problems that need to be handled are the problems emerging from treating these experiences as perceptions.

In particular, if colour experiences are perceptions, then colours have to be analyzed as perceivable physical properties of physical objects. There are well-known objections to this view, but Armstrong thinks the objections can be overcome, and colours can be identified with micro-physical properties of objects. Bodily sensations are perceptions of various parts of the body. Pain, for example, is a perception of certain disturbances in the body, most probably of the stimulation of pain receptors.

The peculiarity of these experiences, according to Armstrong, is that they don't offer any clue about the nature of the physical or physiological properties that they are experiences of.

> ...while we recognize by sight that all red things have something in common, sight does not inform us what that common property is. In the same way, we recognize by bodily perceptions that the class of felt disturbances called 'bodily pains' all have something in common. But bodily perception does not inform us what that common feature is. (Armstrong 1968, 314)

> ...we recognize by sight that red things differ among themselves in respect of redness. There are different shades of red. In the same way, certain pain resemble each other and differ from other pains. We recognize different sorts of pain. In the case of colours, however, we need not concede that vision informs us of the nature of the differences involved in being different shades of the same colour. In the same way, we need not concede that bodily perceptions inform us of the nature of the difference of bodily disturbance involved in the different sorts of pains. We are simply informed that the disturbances do differ in some respect. (Armstrong 1968, 315)

Perhaps this is the point where something like the explanatory gap appears in Armstrong's theory. The issue is not the same as the one discussed by Levine. Armstrong's remarks concern the relationship between the felt quality of the experience and the property that's perceived in having that experience, while Levine is interested in the relationship between the experience and the realizing brain (or nervous system) state. Still, the issue is a certain lack of intelligible connection. Viewed from the side of the experiences, there is nothing that tells us about the nature of the physical property that is perceived in the experience. So

viewed from the other direction, the nature of the physical property will probably not explain why the experience feels the way it does.

Note the similarity between Armstrong's remarks and, again, Locke's ideas on primary and secondary qualities. Locke thought that the ideas of primary qualities *resemble* those qualities that cause them; whereas the ideas attached to what we call 'secondary qualities' don't resemble their causes, because those causes are in fact combinations of primary qualities of insensible parts:

> … the ideas of primary qualities of bodies are resemblances of them, and their patterns do really exist in the bodies themselves, but the ideas produced in us by these secondary qualities have no resemblance of them at all. There is nothing like our ideas, existing in the bodies themselves. (Locke 1690, Book II, Ch. VIII, 15)

Ideas of sensations like pain, sometimes called ideas of 'tertiary qualities' are treated by Locke in the same manner as ideas of secondary qualities. Tertiary qualities are bare powers of objects; that is, a combination of the primary qualities of insensible parts, that cause certain ideas in us, but again, the ideas don't resemble the causes. The difference between ideas of secondary and tertiary qualities is that in the latter case, we are not even tempted to think that their causes resemble them, while in the former case, we are so tempted.

4. An idea can be like nothing but an idea

As many have pointed out, the notion that ideas of primary qualities resemble the qualities that cause them is puzzling. Something has a circular shape if it takes up a certain space within a boundary whose points are equidistant from a central point. How could an idea resemble this, when ideas are either immaterial – as Locke thought – or realized in a brain state that is certainly not circular itself? As Berkeley aptly remarked, 'an idea can be like nothing but an idea'. Ideas of primary qualities will no more resemble the qualities that cause them than ideas of secondary qualities.

The suggestion that ideas could represent things in the world by resembling them is certainly discredited in contemporary theorizing about representations, and yet the Lockean idea survives in various forms. One, it seems to me, is Armstrong's view above: when we recognize that all red things have something in common, the idea does not inform us of the nature of the shared property – that is, that it's some microphysical property of surfaces.

The situation is supposed to be different for primary qualities. But is it? Locke's list of primary qualities is solidity, extension, figure (shape), motion or rest, and number. Consider the primary qualities involved in my current experience. I see *one stationary* coffee cup on the table, with a *continuous smooth* surface and a *cylindrical* shape. As a matter of fact, these qualities are realized by a combination of the primary qualities of insensible particles: a structure of not one, but *many* particles, with *space between them*, and these particles may *move* while the cup is stationary. Where is the resemblance? And we could hardly discover the microphysical realization of these perceivable properties just from the ideas of the properties (perceived shape, size and motion) themselves. So primary qualities seem to have the same status in this respect as secondary qualities. The point, as it is sometimes remarked, is the difference between the manifest and the scientific image, rather than between primary and secondary qualities.

Interestingly, elsewhere Armstrong acknowledges what seems to me a similar point. He introduces the topic of secondary qualities with the well-known observation that the scientific image of the world, at least since Galileo, provides no place for the secondary qualities, since physical theories do not attribute such qualities to the fundamental constituents to the physical world (like particles or fields). He adds that some philosophers hold that therefore secondary qualities must qualify the mind, rather than the physical world. Armstrong responds by endorsing an objection made by Berkeley: in a visual experience, he says, for example colour and visible extension 'are inextricably bound up with each other' (Armstrong 1968, 272). So if we relegate colour to the mind, we should do the same with visible extension.

All these points seem to be manifestations of the same underlying issue. Our awareness of primary qualities in perceptual experiences surely contributes to the phenomenal character of these experiences alongside with our awareness of secondary and tertiary qualities. Insofar as these are all ideas, that is, (properties of) experiences, they have the same ontological status. And even if they are all material, none of them will 'resemble' their extra-experiential causes more than any other: a perception of a blue circle will be no more circular than it is blue. Further, neither of them will inform us, just by the way they appear in our awareness, about their microphysical realization.

When Locke says that for secondary qualities, God assigned simple ideas to corpuscular processes, but could have chosen to assign different ones, he is motivated by the thought that the ideas of secondary qualities do not resemble their causes. One of the places that Levine references in Locke's *Essay* (Book II, Ch. VIII, sec. 13) explicitly mentions the lack of 'similitude' between these ideas and the movement of insensible particles affecting our senses which cause these ideas. Although Levine says that Locke captures the same point that he (that is, Levine) is trying to make, surely he doesn't mean the view that only primary qualities resemble their causes. So we should see if Levine is more successful than Locke and Armstrong in drawing a contrast between primary and secondary qualities along these lines.

5. Why this rather than that?

We have seen in section 2 that Levine thinks that there is an explanatory gap between the felt quality of experiences and their physical realization. In this section, I will look more closely at the precise nature of the missing explanation.

The unanswered question that is identified by both Locke and Levine has this form: **'**Why this, rather than any other?' What is missing, according to both of them, is the contrastive explanation of *why this phenomenal character rather than any other* is identical to, or produced by, a certain physiological process. That Levine is missing this contrastive explanation is clear from his formulations of the problem that I quoted above:

> … there seems to be nothing about C-fiber firing which makes it naturally "fit" the phenomenal properties of pain, *any more than it would fit some other set of phenomenal properties.* (Levine 1983, 357, emphasis added)

> R doesn't really explain why I have one kind of qualitative experience … *and not the other.* (Levine 1983, 358, emphasis added)

The same thought is present in Levine's reference to Locke, and the suggestion that corpuscular processes and ideas are stuck together in an arbitrary manner:

> The simple ideas go with their respective corpuscular configurations because God chose to so attach them. *He could have chosen differently*. (Levine 1983, 359, emphasis added)

In other, words, there is no explanation that would make it intelligible for us why God attached this, rather than another idea to the physical process. The lack of such contrastive explanation, at least on the face of it, suggests that phenomenal characters and physiological processes are wholly different kinds of existents. They don't fit into the same order of beings, but instead they exist side-by-side.

This, according to Levine and many other commentators, is merely suggestive at this point. One way to pursue the matter is to develop the argument for the explanatory gap into *an argument for an ontological gap*. This is the anti-physicalist route.

One alternative is a certain type of physicalist views, which acknowledges that there is an explanatory gap, but resists the further development, for example by criticizing the anti-physicalist's positive argument. Some of these physicalists felt that they still owe an account of *why* there is an explanatory gap, since the gap seems to point towards distinctness where no real distinctness exists (Papineau 2002).

In what follows, I would like to pursue a strategy that's different from both the above physicalist and anti-physicalist strategy. I want to suggest that contrary to what Levine and Locke say, in many cases, there *is* a contrastive explanation, and the gap can be closed, or it is at least not as wide as it first appears. The explanation is provided in the framework of the Causal theory of the mind. This, in itself, is not a supporting positive argument for physicalism. Just as asserting the explanatory gap does not amount to the assertion of ontological distinctness, denying the explanatory gap does not amount to denying ontological distinctness.

The 'why this rather than that' is a particular version of an explanatory demand, and there are other possible explanatory demands. For example, we could ask: why anything, rather than nothing? That is, why do some physiological processes constitute, or give rise to conscious experiences *at all*? Another possible question: why this, rather than something slightly different? Why does this physiological process give rise to a sensation of precisely this shade of phenomenal red, rather than of a somewhat darker shade?

The focus of this paper is the first kind of question. I believe this is the question asked in Levine's paper, as the emphasis in the quotes above show. The original statement of the explanatory gap was a plea for a missing 'why this, rather than that' explanation, and this is the explanation I aim to give. As I will clarify below, my strategy may not suitable for satisfying the other two explanatory demands. What is achieved by my argument, if these other questions are left unexplained? In the last section, I will return to this question and try to answer what will have been achieved by then. Here I simply ask the reader to keep in mind that I am deliberately targeting the first type of question, and not the second two.

6. The contingency of the realizer

Levine claims that we don't have an intelligible explanation of why the qualitative feel of pain, rather than another phenomenal property, goes together with C-fibre firing. In

contrast, I would like to argue that we do have such an explanation. The explanation is based on two crucial ideas: the first is the Causal theory of the mind, and the second is a thesis about the natural fit between the phenomenal character of experiences and their functional roles. This section expounds the first part, the next two sections the second.

The first step of the Causal theory, as we have seen, is to identify the causal (or functional) role of a mental phenomena. Pain is normally caused by some damage to the body, and it normally causes avoidance behaviour. Of course, the actual causal role of pain is much more complex, and is relative to other mental states, but these can be included in the analysis.

Once we identified the functional role of an experience, we will look for the mechanism or process that plays this functional role in human beings. Here philosophers tend to behave with a certain amount of nonchalance about the empirical details, assuming that scientists will identify some or other process. Kripke himself admits that he knows virtually nothing about C-fibres, except that their stimulation is said to correlate with pain (Kripke 1972, 149). One thing we cannot take for granted is that there will be a single mechanism or process type. For example, it seems very likely that pain has a sensory and an affective component, and these can be dissociated through selective damage (Grahek 2011). Even the sensory component is likely to be realized by more than one type of process. I'm going to appeal to the same device as Levine does when he calls the 'physical story' of seeing red 'R': I am going to call the 'physical story' for pain 'C-fibre firing', and note that this story may turn out to be quite complex.

The connection between C-fibre firing and the causal role of pain involves some contingency, in two directions. First there is the widely accepted thesis of multiple realizability, meaning that processes other than C-fibre firing could play the pain-role. The connection is not entirely arbitrary: it's not true that just any physical or physiological process could do the job. But there is almost certainly a lack of necessary connection. The other direction of contingency concerns the functions that C-fibres could play. Plausibly, they could play roles other than the pain-role – again, not just any roles, but probably more than one. This is not news: most philosophers accept that the connection between the role and realizer of mental functions is, to some extent, contingent.

Levine himself is of course aware of this: at the beginning of the paper, he mentions the functionalist proposal that 'the intuition that pain could exist without C-fibers is explained in terms of the multiple realizability of mental states' (Levine 1983, 355). Since he thinks the functionalist will have other problems, he abandons this explanation. This, I believe is a mistake; what is partly (only partly, but still) responsible for the appearance of an explanatory gap for 'Pain is C-fibre firing' is an attempt to connect directly the realizer of the role with the phenomenal property

When Levine asks what it is about C-fibre firing that fits the phenomenal properties of pain, the functional role momentarily drops out of the picture. The direct connection between C-fibre firing and pain would be difficult to establish. Levine says that the lack of intelligible connection between C-fibre firing and pain explains why in some other arguments, philosophers rely on the possibility of imagining one without the other. Continuing the presentation of Locke's view that the ideas and corpuscular processes are attached only by God's will, he says that

> so long as the two states of affairs seem arbitrarily stuck together this way, imagination will pry them apart. Thus it is the non-intelligibility of the connection between the feeling of pain and its physical correlate that underlies the apparent contingency of that connection. (Levine 1983, 359)

But this diagnosis can be questioned. I propose to return to the familiar functionalist point: that the 'apparent contingency' is a consequence of the contingent connection between C-fibre firing and the pain-role. Even if there was some intelligible connection between the feeling of pain and the *pain-role* (as it is proposed in the next section), the realizer would still occupy this role contingently. This is the first step in closing or narrowing the explanatory gap.

.

7. The natural fit between qualitative character and functional role

The first part in narrowing the explanatory gap was to identify the functional role of an experience and look for the physical process that plays this role. The second part is to point out that there is a *natural fit between the phenomenal character and the functional role of conscious experiences.*

If we follow Levine, the explanatory gap opens when we ask why C-fibre firing is identical to (or gives rise to, or correlated with) pain, rather than with the experience of pleasure, or the experience of smelling gasoline. As we saw in the previous section, this isn't really the question we should be asking. The important question is rather this: why does an experience with the phenomenal character of pain (rather than with the character of feeling pleasure, or the character of smelling gasoline) *play the pain-role*? Levine claims that we have no intelligible explanation for this either. I think he is wrong: the answer is that the phenomenal character of pain is eminently suitable for an experience that plays this functional role.

First, pain is inherently unpleasant. This is part of the phenomenal character of pain. That pain is unpleasant makes it no surprise that it causes avoidance behaviour. Other things being equal, we prefer not having unpleasant experiences, so we try to avoid them or stop them once we have them. (If pain has a distinct sensory and affective component, then unpleasantness is part of the affective component. And that's precisely the component that seems to give rise to avoidance behaviour; see Grahek 2011). Second, it is good to try to avoid or stop pain because pain alerts us to a damage in the body and – other things being equal – we would like to avoid or stop damage to the body. So it is a good idea that damage in the body causes this unpleasant experience which, in turn, causes avoidance. Third, physical pains normally have a felt location in the body. The felt location is part of the phenomenal character of the experience, and it directs the agent towards the possibly damaged bodily part. Fourth, by and large, more serious damage causes more intense pain, which is again part of the phenomenal character of pain.

Hence far from being arbitrary, it was in fact a very wise decision by God to choose the unpleasant experience of pain which is felt in a certain part of the body to accompany the state that is caused by damage in that part of the body, and causes avoidance behaviour.

If God is all powerful, maybe he could have associated the experience of pleasure or of smelling gasoline with the functional role of pain, but it is really hard to see how that would have worked for us. First, it's hard to see how the experience of smelling gasoline could play the role of alerting to damages in various parts of our body. Smells don't have a

felt location in the way pains do. Perhaps it could be suggested that we could have different smells associated with damage in different parts of the body. That means the smell of gasoline alone is already disqualified, since it doesn't have the complexity that pain sensations do. To inform us of damage of the many different parts of the body, we would need a lot of different smells. Integration would be a problem. We are not very good in distinguishing different parts of a complex smell; if we fell and hurt our knee and hand, and we would have the sensations of the smell of onion and of burnt toast (supposing these are the simple ideas arbitrarily associated with damage in the knee and hand) it's not clear that we could discern the elements of the complex resulting sensation. They would also have to be integrated with sensations of pressure and the sensation of temperature. The whole enterprise seems rather hopeless.

As for pleasure: what we are to imagine is that pleasant feelings would cause immediate distress and avoidance behaviour. Now, it is possible to imagine that a person exhibits such reactions: suppose someone is deeply convinced that feeling any kind of pleasure is a mortal sin. For such a person, feeling pleasure could cause immediate distress. But note that this isn't a scenario where pleasure occupies the functional role of pain: rather, this explanation simply makes the story intelligible to us relative to *our* experience of pleasure, with its ordinary functional role. We noted above the familiar point that functional roles are always relative to other mental states, and this is just another instance of that point.

But what about the other case Levine mentions, that of red and green? Levine here relies on the familiar idea that red and green sensations could be inverted. Surely, here God could have chosen to attach either of the two ideas with the physical story R. In response, note that the case of red and green are very special. It is simply not true that you could invoke any other pairs of phenomenal properties and easily imagine an inversion, while retaining functional equivalence. For example, a pain-pleasure inversion, or a pain and experience-of-smelling-gasoline inversion are not plausible at all.

One reason inversion is not plausible in these cases, as mentioned above, is that different affective elements – unpleasant, pleasant, neutral – are associated with each experience, and these are plausibly responsible for different effects the experience has on our behaviour. The other reason why many inversions are not plausible is that phenomenal properties are part of a quality-space, with certain structural features (Clark 1993). Pains vary in intensity, in felt location, and in qualities such as throbbing, dull-aching, burning, or stabbing. These are all phenomenal features, and they are responsible for the felt similarities and differences among episodes of pain experiences. The structure of the quality space shows a natural fit with the typical causes of the experience, and it is reflected in our discriminatory behaviour in responding to these causes. (The concept is presented in much more detail in David Rosenthal's contribution to this volume.)

The structure of the pain quality space is quite different from the structure of the odour quality space, which in turn is different from the structure of the colour quality space and the sound quality space. So it's hard to see how we could replace a quality by a completely different one while keeping the same structure of causes and the same range of behavioural manifestations. In fact, it was suggested that not even the red-green inversion is made possible by the phenomenal structure of colours, and that this has relevance for the explanatory gap (Hardin 1987).

8. Phenomenal Functionalism

The homomorphism between quality spaces and physical causes, and the fact that quality spaces are reflected in discriminatory behaviour, suggested to some that we could offer a reductive theory of phenomenal qualities. David Rosenthal, in his contribution to this volume, presents exactly this proposal (see also Clark 1993and Rosenthal 2005). But we need not assume such a reductive view to make use of the explanatory potential of these homomorphisms. I will sketch a different background theory that goes well with the insight about the natural fit between phenomenal characters and functional roles. (Space doesn't permit me to defend this view here in detail, so I will restrict myself to its presentation.)

I propose a view which may be called Phenomenal Functionalism. It is meant to be parallel to the Phenomenal Intentionality proposal (Horgan and Tienson 2002, Loar 2003), on which the intentional features of conscious experiences are grounded in their phenomenal character. Similarly, I put forward that *(some of) the functional features of conscious experiences are grounded in their phenomenal character*. We try to avoid unpleasant experiences and seek out pleasant ones. Phenomenal qualities are located in quality spaces, and this makes them suitable to be responses to a homomorphic structure of causes which characterizes a certain portion of reality. The structure of these spaces is quite different for different kinds of qualities (for example, in different perceptual modalities), hence the functional roles of experiences – their typical causes, and the discriminatory consequences in behaviour – will be quite different. For these reasons, seamless inversions of qualities while we keep functional or physical states the same are very rare.

In addition to bodily sensations, phenomenal functionalism works well for occurrent moods like anxiety or elation. Anxiety is characterized by a characteristic behaviour: for example the inability to concentrate, the tendency to get irritated, to jump on unexpected stimuli, and so on. These reactions are different from the characteristic behavioural patterns accompanying for example carefree contentment (Crane 1998). And the way these moods feel seems to be intrinsically connected to these roles: as in the case of pain and pleasure, it's difficult to imagine an inversion of the feeling of anxiety and contentment.

The qualitative feel of experiences is often characterized as something that's left once we extract their intentional and functional features. Tim Crane calls this the 'residue conception' of consciousness (Crane 2018). Levine seems to be suggesting something along these lines. After he acknowledges that the causal role of pain is a crucial part of our concept of pain, he says that 'there is more to our concept of pain than its causal role, there is its qualitative character, how it feels; and what is left unexplained by the discovery of C-fibre firing is *why pain should feel the way it does*!' (Levine 1983, 357, emphasis in the original).

The idea that the qualitative feel is 'more' than the causal role can be interpreted in different ways. It could mean that the qualitative feel is not reducible to the functional role, and in this sense, I am in agreement. But it could also mean that the phenomenal character is completely independent of the functional role – that once the functional role is fixed, any kind of feeling could be added. As I explained above, I find this kind of independence very implausible, but this is exactly what seems to be suggested by Levine when he talks about the explanatory gap.

The Phenomenal Intentionality program was born out of a dissatisfaction between the sharp separation between the phenomenal and the intentional features of conscious experiences. In opposition to this separation, defenders of the program suggest that the

intentionality of mental states is inextricably linked to their phenomenal character. Phenomenal Functionalism would be a similar protest against the divide between the phenomenal and functional features, and an affirmation of their deep connection.

The Phenomenal Intentionality program is one way of opposing the sharp separation between phenomenal and intentional features. Another is representationalism, a view that also sees a strong connection between the two, but reverses their order of priority. Instead of holding that intentional features are constituted by phenomenal features, it asserts the supervenience of the phenomenal on the representational. Phenomenal Functionalism also has such a counterpart: good old functionalism. This form of functionalism would not try to eliminate phenomenal features, but would also deny the residue conception of consciousness.

Both functionalism and Phenomenal Functionalism immediately entail that there is a natural fit between the phenomenal features of an experience and its functional role. Hence they are good background theories for the explanatory task carried out in section 6.

My sympathies lie with Phenomenal Functionalism, and someone may ask how this background theory fits into the overall project of closing the explanatory gap. Though as far as I can see, Phenomenal Functionalism, at least as characterized so far, is neutral on the issue of physicalism, it is most plausibly seen as opposed to a reductive view of phenomenal features (just like the Phenomenal Intentionality program). One may ask whether this aspect would render its services useless in the current project. The thought is that closing the gap is a project for physicalists. A view that posits phenomenal features as basic in the explanatory theory will hardly serve this purpose.

On the picture suggested by Levine, by Locke, by the residue conception of consciousness, by the epiphenomenal qualia view, there is absolutely no intelligible connection between the functional and phenomenal features of experiences. These two aspects of reality exist side-by-side. Even if the explanatory gap is not in itself an argument for anti-physicalism, its existence is meant to be a concern for physicalists. Consequently, we may expect that anti-physicalists would cheerfully acknowledge the existence of the gap. I've been arguing in this paper that this would be wrong, on independent grounds: whether physicalism is correct or not, the sharp separation of the phenomenal and the functional is simply implausible. Our experiences fit very well into the physical world and our place in it. Instead of arbitrarily assigning simple ideas to physical processes, God, in his wisdom, assigned phenomenal characters to experiences that are eminently suitable for playing certain functional roles.

9. Various gaps

Conscious experiences have a phenomenal character, and this character has a natural fit with the functional role of the experience. Hence the explanatory demand 'why this, rather than that?', posed by Levine, can be satisfied. C-fibre firing correlates with the experience of pain, rather than, say, pleasure, because in human beings, C-fibre firing occupies the functional role of pain, and the phenomenal character of pain naturally fits this functional role (much better than it would fit the functional role of pleasure).

Thus we can answer the 'why this, rather than that' question. Where does this leave us with the other possible explanatory questions we identified in section 4? Let us first consider the question 'why this, rather than something slightly different?' I am not sure what to say about this case. If we identify the functional role of experiences on the basis of a

common-sense analysis, then it seems that there is no answer to this question. Yes, the unpleasantness of pain is suitable for causing avoidance behaviour, but a slightly more or less intense pain would also seem suitable to cause the same. Yes, these causes and this discriminatory behaviour perfectly fit the inner similarity relations among the sensations of shades of red, but if all shades felt a touch darker, the fit would seem to be the same. This would suggest that the strategy pursued here cannot answer this explanatory question.

However, what makes me hesitant is the prospect of a scientific inquiry for example into quality spaces and their corresponding causes that would reveal subtle differences undetected by common-sense analysis. Something like this happened in the case of red-green inversion: common sense seems to license the possibility of such an inversion, but it has been argued that a more subtle investigation of the colour quality space shows that a perfect functional equivalent is not possible (eg. Hardin 1987). Maybe more subtle investigations would also exclude the possibility of shifts.

Can we answer the 'why anything, rather than nothing' question – in other words, can we explain why some physiological processes give rise to conscious experiences at all? According to the line of thought presented in this paper, once we have the functional role of states or events on one side, and the phenomenal characters on the other, we can pair them according to the natural fit between character and role. But the same strategy does not explain why there were phenomenal characters in the first place. Some of our bodily functions seem to work perfectly well without the involvement of consciousness. For example, arguably, processes that occur in the dorsal visual stream, which takes visual stimuli as input and issues motor commands as an output, do not give rise to consciousness (Goodale and Milner 1992). In contrast, processes in the ventral stream do give rise to consciousness. The strategy pursued here does not explain why.

In fact, this challenge can also be divided to further, more specific tasks. One is to give an account of the general function of consciousness (for example having to do with combining information from several sources), and try to answer questions like the one above. When certain interactions with the world fell in the category of requiring consciousness, we could detect this from their functional role. This theory would assign a uniform account for all conscious processes.

There is a different task that Mary, the scientist in the black and white room faces (Jackson 1982, Howard Robinson 1982 has a similar example with a deaf scientist). Let us first consider a modified Mary story, where Mary lives in a normal world and is accorded the full range of experiences, but her studies are curiously detached from her experiences. She studies in detail the physics and physiology of sensations, perceptions, and emotions, but she is not told which of her experiences are connected to which physiological process. For example, she studies 'C-fibres' (or whatever complex physiological process which underlies pain), but is never told that this process in fact underlies pain. According to the argument pursued in this paper, after a while, she would be able to make the connection herself.

Original Mary is in a less favourable position. Her task is not to pair experiences with physical processes, but rather to somehow deduce the nature of the experience from the physical process, without any further help. The consensus is that she won't be able to do that. She may be able to form some idea: she will know it's a colour experience, it's not an emotion or a bodily sensation, and since she has had those, she will have some reasonable expectations of the phenomenal character of experience. But she won't know it exactly.

One immediate consequence is that if there is an epistemic gap between the physical and the phenomenal, it will be the kind of gap that we find in original Mary's understanding. Indeed, Levine himself, in his book *Purple Haze* (2001), shifts focus to the absent qualia argument and the knowledge argument from the kind of explanatory question that he considers in the original paper. Note, however, that Mary's predicament is not very naturally characterized as lacking some *explanation*. Explanations are usually answers to 'why' questions; they have an explicandum: a fact, or an event. Mary's task is to deduce the character of the experience from its physical description. It doesn't look like she is trying to answer a why-question, or that there is a fact or event that she is trying to explain. In contrast, in Levine's 1983 paper, there was a clear why-question waiting for an answer: why this, rather than that? There is also the other general explanatory question – why are some processes conscious? – but that is not the question Mary is trying to answer either.

So while Mary has a gap in her understanding or knowledge, I would be reluctant to call it an 'explanatory gap'. In any case, no matter what we call it, the question is whether this gap threatens our understanding of the mind. The answer will partly depend on our assessment of what is achieved by explanations about physical phenomena.

For example, do we have an explanation of why the fundamental constituents of the universe are what they are? Couldn't we have a world with a slightly different assortment of elementary particles? It is arguable that certain natural constants that determine the strength of interactions, have, according to our best theories, taken spontaneous values during the evolution of the universe, and could have had other values within a certain range. And many of these facts affect the world of the elementary parts. Perhaps it's just a fact that about our world that it contains certain basic ingredients. Similarly, we could argue that it is a fact about our world that it contains certain conscious experiences.

If this is right, then answering Levine's question, 'why this, rather than another?' actually goes a long way to close or narrow the explanatory gap. We have a world where certain beings have conscious experiences with phenomenal characters. These phenomenal characters don't belong to a realm that it entirely disconnected from the physical realm: in fact, there is a natural fit between the two kinds of features. There isn't a huge gap between the physical and phenomenal aspects of the world.

References

Armstrong, D. M. (1968) *A Materialist Theory of Mind*, London: Routledge and Kegan Paul

Armstrong, D. M. (1999) *The Mind-Body Problem: An Opinionated Introduction*. Boulder, CO.: Westview Press.

Clark, Austen (1993) *Sensory Qualities*. Oxford: Oxford University Press

Crane, Tim (1998) 'Intentionality as the Mark of the Mental' in *Royal Institute of Philosophy Supplement* 43. Cambridge: Cambridge University Press, 229–251.

Crane, Tim (2018) 'A Short History of the Philosophy of Consciousness in the Twentieth Century' in Amy Kind (ed.), *Philosophy of Mind in the Twentieth and Twenty-First Centuries: The History of the Philosophy of Mind, Volume 6*. London: Routledge.

Grahek, Nikola (2011) *Feeling pain and being in pain*. Cambridge, MA: MIT Press.

Hardin, Clyde L. (1987) 'Qualia and Materialism: Closing the Explanatory Gap' *Philosophy and Phenomenological Research* 48 (December), 281–98.

Horgan, Terence & Tienson, John (2002) 'The Intentionality of Phenomenology and the Phenomenology of Intentionality' in David J. Chalmers (ed.), *Philosophy of Mind: Classical and Contemporary Readings*. Oxford University Press, 520–533.

Jackson, Frank (1982) 'Epiphenomenal Qualia' *Philosophical Quarterly* 32 (April),127–136.

Levine, Joseph (1983) 'Materialism and Qualia: The Explanatory Gap' *Pacific Philosophical Quarterly* 64 (October), 354–61.

Levine, Joseph (2001) *Purple Haze: The Puzzle of Consciousness*. New York: Oxford University Press

Loar, Brian (2003) 'Phenomenal Intentionality as the Basis of Mental Content' in Martin Hahn & B. Ramberg (eds.), *Reflections and Replies: Essays on the Philosophy of Tyler Burge*, Cambridge, MA: MIT Press, 229–258

Locke, John (1690) *An Essay on Human Understanding.* Peter H. Nidditch (ed.), Oxford Clarendon Press 1975

Kripke, Saul A. (1972) 'Naming and Necessity' in Donald Davidson & Gilbert Harman (eds.), *Semantics for Natural Language*. Dordrecht: Reidel. Reprinted in book form by Blackwell, Oxford 1980.

Papineau, David (2002) *Thinking about Consciousness.* Oxford: Clarendon Press

Robinson, Howard (1982) *Matter and Sense: A Critique of Contemporary Materialism*. Cambridge: Cambridge University Press.

Rosenthal, David M. (2005) *Consciousness and Mind*. Oxford|: Oxford University Press.

Rosenthal, David M. (xxxx) 'Armstrong and perception' (this volume)