THE OBSTINACY OF APPEARANCE
An analysis and attempted resolution of the explanatory gap debate about consciousness

by James Phillip Frank Tartaglia

University College, London
Ph.D.
ABSTRACT OF THESIS

Explanatory gap intuitions derive from the apparent conflict between views V#1, that consciousness is a natural phenomenon, and V#2, that consciousness is subjective. Deflationists presuppose V#1, rendering V#2 problematic; inflationists presuppose V#2, rendering V#1 problematic. Attributions of essences are based on individuation within a conceptual scheme, so physicalism combined with the essentialist claim that there is no appearance/reality distinction for conscious properties creates conceptual difficulties. Attempts to move away from a conception of consciousness as appearance fail, since all principal characteristics of mind presuppose it. Appearance cannot be functionalised because a fine-grained functional description saying what appearance is says it is physical, and a looser-grained description for recognising appearance allows room for error where there cannot be any. The concept of appearance, read into V#2, is apparently indispensable, but cannot be related to the physical world: deflationists cannot talk about consciousness without it, and inflationists cannot employ it without making physicalism problematic.

The Modes of Access Response says phenomenal concepts have no associated functional descriptions, express the same properties they refer to, and lack conceptual connections with physical concepts because of our psychologies. But with no leeway between how we conceive of something and what it is, phenomenal concepts cannot behave as recognitional concepts. The position belongs to a cycle of thought on the problem of conceptual difference, trying to avoid phenomenal concepts expressing intrinsic properties of appearance. There can be no resting place, given the epistemological role of appearance. But it is argued here that consciousness is not appearance. The mind is purely representational, so no properties of consciousness itself are presented to introspection. All recognitional concepts are of properties presented by consciousness. Confusion about the inescapability of the human perspective explains the apparent conceptual gulf. There are no conceptual obstacles to physicalism about consciousness.
Acknowledgements

Main thanks to Tim Crane and J.J. Valberg for all their help, as ever.

Thanks to Drs. Jon and Sophie for reading *The Obstinacy* through for me.

My studies were supported by a British Academy Studentship, and UCL awards and teaching, all of which I greatly appreciate.

This thesis is dedicated to Zoe Hoida,

who I married between chapters 6 and 7.

- James Tartaglia, Leytonstone, 26 July 2001
## CONTENTS

### Chapter 1: Inflationism and Deflationism about the Explanatory Gap

1.a. Introduction  
1.b. V#1  
1.c. V#2  
1.d. The Explanatory and Causal Mind-Body Problems  
1.e. The Idea of an Explanatory Gap  
1.f. Motives and Things to Come  

### Chapter 2: Scientific Essentialism and the Mind

2.a. A Framework of Natural Kinds  
2.b. Essence and Individuation  
2.c. Logical Consequences of Essentialism  
2.d. The Moulding of the Problem by the Framework  
2.e. Rejecting the Framework  

### Chapter 3: Consciousness as Appearance

3.a. Immediate, Phenomenal, and Occurrent  
3.b. The Appearance / Reality distinction  

### Chapter 4: The Entrenchment of Appearance

4.a. First and Second-Order Essence
4.b. Is the Mental Unified? A Taxonomy 72
4.c. Weak Intentionalism 74
4.d. Strong Intentionalism 81
4.e. Disunity of Mind 89
4.f. Duality of Mind 101
4.g. Appearance explains other Characteristics of Mind 108
4.h. So what is appearance? 114

Chapter 5: Why Appearance will not Functionalise 121
5.a. Appearances as Intrinsic Properties
5.b. Appearances as Functional Properties
5.c. Appearance and the Framework

Chapter 6: Strategies in a No-Win Situation 140
6.a. The Stand-Off between Inflationism and Deflationism
6.b. Offensive and Defensive Strategies
6.c. Conclusion

Chapter 7: The Modes of Access Response 160
7.a. Introduction
7.b. The Significance of MAR
7.c. Three Takes on Conceivability
7.d. Doing without Reference-Fixing Descriptions
7.e. Recognitional Concepts and Essential Conception
Chapter 8: The Problem of Conceptual Difference: A Cycle of Thought

8.f. Inappropriate Affiliations
204

Chapter 9: How could this be a state of my brain?

9.a. Disentangling Consciousness and Appearance
242
9.b. Introspective Access to Consciousness
247
9.c. Five Considerations
255
9.d. Human Representation
268
9.e. The Obsoletion of Appearance
276
9.f. Conclusion
288

Chapter 10: New View

10.a. Introduction
293
10.b. V#3
293

Bibliography
297
Chapter 1: Inflationism and Deflationism about the Explanatory Gap

1.a. Introduction

The topic of this thesis is the apparent explanatory gap within physicalist accounts of consciousness. My approach to this problem falls broadly into three parts. Firstly, I impose a structure upon the current debate, revealing key assumptions underlying opposing positions. Secondly, I develop a critique of the strategies employed by advocates of these positions. Thirdly, I try to provide an outline of a conception of consciousness which, without being implausibly revisionary, can still reject the assumptions giving the debate its impetus.

Our starting point, then, is the imposition of a structure upon the current debate, which will allow us to see the root of the problem. The justification for this interpretation of the debate should gradually emerge, as we see how many actual disputes it explains and how neatly it explains them. In the process, we will come to see what type of approach is needed if the problem is to be resolved.

Let us begin by distinguishing deflationism and inflationism about the explanatory gap. All the participants to the explanatory gap debate that I know of are either deflationists or inflationists, so evidently these are meant to be broad categorisations. The terminology simply reflects the idea that deflationists deflate the significance of the explanatory gap, and inflationists inflate it. The deflationist and the inflationist will be the central players throughout this work, and there will be many examples of both.
Deflationists and inflationists take differing attitudes to two views about consciousness, which I shall call V#1 and V#2. These views present very basic conceptions of consciousness, which philosophers of mind are strongly influenced by and inclined towards. V#1 and V#2 are the source of many of the intuitions which mould the infinitely more sophisticated theories of consciousness that actually get defended. The explanatory gap debate can be understood as springing from an apparent tension between these two views. Our last chapter will present a new view, V#3, which purports to be free from this tension.

V#1 is essentially a conception of consciousness as unproblematically a part of the physical world, like any other natural phenomenon. V#2 is essentially a conception of consciousness emphasising its subjectivity. Both strongly influence deflationist and inflationist positions, but in different ways. The following presentations of V#1 and V#2 can, of course, only claim to be representative of the sort of view I have in mind, but the purpose is just to remind ourselves of what such views amount to.

1.b. V#1

We live in a physical universe. All objects, large and small, are made up from and connected by the tiny particles, forces, waves, and suchlike, which physical science describes. We ourselves are no exceptions. Humans are physical things in the world, like anything else. What does mark us out as exceptional though, is consciousness. This special physical constitution means that we can know the rest of the physical world around us. Vases, by contrast, are unconscious, and as a consequence, they are oblivious to their surroundings.
Being conscious is a matter of having a certain physical structure, just as being made of wood is a matter of having a certain physical structure. However, the crucial aspect to that structure seems to be how the brain and nervous system function, rather than what they are made of. Perhaps the chemistry and physics of the massed neurons which make up the brain are so ingeniously adapted to their function that we cannot imagine anything else supporting consciousness. Nevertheless, it does seem that it is what they do, and not just what they are, which makes us conscious, since the same physical stuff remains present in those who have lost consciousness and died. It can no longer be functioning as it once did.

This functioning facilitates complex behaviours. Consciousness allows for discriminating responses to a variety of environmental conditions. The conscious being is not just causally pushed around by the surrounding world, and does not just make automatic responses to situations, from a repository of “built-in” set-pieces. It does all this too, but is distinctive in having goals which it actively pursues in light of what it takes the state of its surroundings to be, even if the organism is so primitive that its goals are limited to the avoidance of pain. Awareness of the state of the world which guides behaviour comes through perception. Having learnt to correlate perceptual changes and patterns of change with external conditions, we can engage in complex patterns of interaction with the environment and with other conscious organisms.

Anything that could learn, invent, reflect, and co-operate as we do, would be such a complex physical system that it would need to be conscious. When we imagine what an active human being would be like if it lacked consciousness, we imagine the fictional character familiar to the popular imagination as the “zombie”, which is slow, unresponsive, and stupid. Only the world’s more advanced biological systems have the
complex physical organisation we call “consciousness”. To find out more about
consciousness, we will have to find out more about the higher-level functioning of our
brains and nervous systems.

1.c. V#2

Physical things are “just there”. They can be completely described
impersonally, without any mention of an observer, or a point of view. A vase is a
physical thing, and to give a complete description of it, we need only talk about the
material from which it is composed, its shape, its dimensions, and so on. All the
elements of this description are valid for everyone; they hold true of the vase just as
much for someone viewing it from close up, or from a distance, or someone who has
only been told about the vase, and is thinking about it. They would even hold true if
there were no people. They claim universal validity.

This is all well and good, because an objective view of reality is desirable. For
the purposes of describing the world, we want to know how the vase is, not how the
vase is for Smith. However, if we want to describe consciousness, it looks as though
we must abandon the third person perspective. For if we employ the usual disinterested
scientific methodology of adopting a “view from nowhere”, we will not find any
consciousness in the world. All we will find are physical objects which behave as
though conscious. We will find human beings who exhibit all the paradigmatic outward
criteria for being conscious, in terms of their behaviour and physical-functional
constitution. But by viewing the world from no particular perspective, we have already
precluded the possibility of encountering consciousness itself. For consciousness is not
"just there", as vases and humans are. Rather, it is essentially tied to a particular
perspective on the world, for the trivial reason that it is a particular perspective on the
world. Conscious phenomena like perceptions, thoughts, and feelings are only available
to be described from a perspective, from within consciousness. They are not just there,
but are rather "there for X" or "there for Y".

Think about what it is like to smell coriander, to taste gin, to look at a Cézanne
in a gallery, to feel off balance whilst skating. All of these experiences are only
available to the person having the experience, that is, they are only available in the first
person. Now suppose we wanted to say what these experiences are in physical terms,
since experiences must be part of the physical world like anything else. To do so, we
would have to adopt the third person perspective, and describe what is actually going
on in the world. And this, surely, is that in the case of the smell of coriander,
something is going on between the coriander, your nose and your brain. More
precisely, we might say that molecules from the coriander enter your nostrils, causing
changes in your olfactory bulbs, which in turn cause changes in your entorhinal cortex.
Could this process be the very same thing as smelling coriander?

Well our ways of thinking about physical processes and experiences are
completely different. When you think about the physical process, you think about a
human being and their brain. When you think about what it would be like to be that
human being, you think about the smell of coriander. You could draw a diagram of the
first, but not the second. The first is a public, objective event, the second is a private,
subjective one. Objective events like horse races and brain events could hardly seem
any more different from subjective events like thoughts and perceptions. In fact, once
you recognise the subjectivity of consciousness, it is entirely obvious that it must be distinguished from anything objective.

Even so, we still have to grant a very significant link between the two. For, to revert to the example about smelling coriander, we think that whenever the right physical process is going on, the person undergoing that process will be smelling coriander. After all, the subjective phenomenon of smelling coriander is nothing more than what it is like to be an objective organism undergoing a certain physical process. So we need to elucidate this connection.

1.d. The Explanatory and Causal Mind-Body Problems

The apparent tension between V#1 and V#2 is easy enough to see. V#1 assures us that consciousness must be physical, in some sense. When we consider the consciousness of others, this can strike us as obvious enough. V#2, on the other hand, urges the need to distinguish experiences from anything physical. When we consider our own consciousness, this can strike us as obvious enough. So for those who feel the pull of both views, there is a conflict: V#2 requires us to make a distinction which V#1 requires us to collapse. We can understand the explanatory gap debate in terms of the different responses that can be made to this conflict.

Inflationists are those philosophers so impressed by the considerations behind V#2, that they find V#1 problematic. Deflationists are those philosophers so impressed by V#1, that they think V#2 must be problematic. Inflationists are moved to reject elements of V#1, or to modify elements of V#1, or, what is most common, to regard V#1 as most likely true in some sense, but confusing, explanatorily incomplete,
mysterious. Deflationists consider V#1 obligatory, and seek to explain the intuitions which motivate V#2 in a way which does not make V#1 look problematic, or else they try to show that V#2 involves a fallacy.

It is the tension created by the intuitive draw of both V#1 and V#2 that gives the explanatory gap problem, the mind / body problem of our time, its particular character. Physicalism seems to require a view like V#1, but the reasoning presented in V#2 is seemingly unavoidable, its resilience demonstrated time and again in new thought experiments and conceivability arguments.

In presenting the problem in this way, we are setting apart the explanatory gap problem from what might be called the causal mind / body problem. The explanatory gap problem is a problem for physicalism. But physicalism is simply one possible position that can be taken on the causal mind / body problem. However, the two problems are closely connected; the explanatory problem results from the combination of V#1 and V#2, whilst the causal problem results from a combination of the reasoning embodied in V#2 and the need to account for mental efficacy in the physical world.

Once you have picked up the train of thought of V#2, and distinguished the subjective and the objective, there will immediately arise the task of putting them back together again, since the mind and body clearly interact. We do things on account of what we perceive and think, and our perceptions and thoughts are causally influenced by the objective world. Descartes distinguished the subjective from the objective, on account of the fact that he could doubt the existence of everything objective, but not everything subjective. He set this distinction in stone with his proposal that the mind and body are different substances.
Cartesian dualism offended the metaphysical principle that only things which share a nature can causally influence each other. Descartes' contemporaries urged this principle upon him, and it is still being appealed to today as a way of ruling out dualism.\(^1\) Of course, Hume's view of causation as constant conjunction constitutes an abandonment of this principle, but the problem for interactionism is easily reinstated through appeal to the well-founded empirical hypothesis that the physical domain is causally closed.\(^2\)

A number of imaginative alternatives to causal interactionism were devised in response to the causal mind/body problem, such as occasionalism, parallelism, epiphenomenalism, and a host of others. The best of these were the ones which worked according to the aforementioned metaphysical principle without the need to appeal to extraneous devices and conditions, that is, versions of substance monism. Of these, physicalism was able to accommodate the causal closure of the physical domain in the simplest possible way. If the mind is nothing over and above the physical world, then the way is smoothed for an account of its efficacy in that world.\(^3\)

But the acceptance of physicalism brings with it the explanatory mind/body problem. We may be able to account for mental efficacy in a physical world on the hypothesis that the mind is physical, but then the reasoning that tempted us to

\(^1\) Gassendi made this objection, which Descartes summarised as the idea that "if the soul and body are two substances whose nature is different, this prevents them from being able to act on each other": *The Philosophical Writings of Descartes, volume 2*, p. 275. See also Dennett (1991) *Consciousness Explained*, p. 35, and McGinn (1991) *The Problem of Consciousness*, p. 52.


\(^3\) Physicalism still has a problem of mental causation, of course, for there are substantial problems in explaining how mental states can be efficacious in virtue of their mental properties, so as to avoid the slide into epiphenomenalism. However, I do think that with the advent of physicalism, the causal mind/body problem has lost its bite, and the real conceptual trouble stemming from V#2 has moved on over into the explanatory mind/body problem.
distinguish the mind from anything physical in the first place, now entreats us to explain how the mind *could* be physical.

The explanatory gap is, as we have said, a problem for physicalism, and this current project is a physicalist project: our aim is to iron out a conceptual difficulty which exists within the physicalist framework. As such, we shall not be mounting a defence of physicalism. Of course, you might reject physicalism for reasons that have nothing to do with the explanatory gap, and not be particularly concerned by the problem. On a broad construal of deflationism and inflationism, this would be a non-physicalist, deflationist view. However, our concern is with that majority of participants to the explanatory gap debate who react to the tension they feel between V#1 and V#2. In analysing their responses, we hope to isolate the root cause of the problem.

Some inflationists react to the tension between V#1 and V#2 by giving up on V#1. These sort of non-physicalist positions are of interest to us here, because they are motivated by the uneasy combination of V#1 and V#2. Considering such positions is quite compatible with our claim that the explanatory gap is a problem for physicalism, for they are simply conceding the problem for physicalism, and hence giving up on physicalism. Such non-physicalists will generally continue to talk about an explanatory gap, however, but they do not mean what we mean; they have crossed over into a concern with something more like the causal mind/body problem.⁴

Since physicalism will be the backdrop for everything that follows, we should at least provide a statement of it. The basic way to get across the idea that the mind is nothing over and above the physical world would be to make an identity claim. But a

---

⁴ I have in mind David Chalmers as an example of a non-physicalist inflationist, and Tim Crane as an example of a non-physicalist deflationist.
formulation of physicalism in terms of property identity is made unattractive by the multiple realisability of mental properties. With a very relaxed attitude to properties, we could perhaps get around this by saying that different instantiations of mental properties, at different times in the same and in different subjects, are all to be identified with different complex physical properties. So, for instance, my pain one minute and my pain the next minute (indistinguishable to me, perhaps), might be identified with different physical properties of my brain. However, though such a position might be possible, it would sacrifice the unity we presuppose in talking of mental properties. We should favour an approach according to which different individuals we describe as “pains” can share a property.\footnote{This objection might be met by identifying mental properties with the disjunction of all their possible physical realisers, see Kim (1993) “Non-reductivism and Mental Causation”. But the need to identify an item with a potentially infinite disjunction suggests nothing but the inappropriateness of an identity claim. The relation of realisation, with which we shall formulate physicalism, is far better suited to the problem in hand (dealing with multiple realisability), allocates more autonomy to mental properties (they are not the same as this or that physical property), and leaves the idea of an identity claim suitably simple.}

An alternative identity claim would be that every token mental event is a physical event. This is fine in and of itself, but it tells us nothing about the connection between mental and physical properties, except that all events with mental properties have physical properties.\footnote{This is assuming Davidson’s conception of events as concrete particulars, to be counted as basic constituents of the world alongside physical objects. On Kim’s approach to events, they are structured particulars consisting of property instantiations by objects at times. If events are structured particulars, then event identities imply property identities. If the identity of a mental event to a physical event means that something having a mental property at t is the same as that thing having a physical property at t, then the properties must be identical. For these different approaches to events, see Kim (1976) “Events as Property Exemplifications”, and Davidson (1970) “The Individuation of Events”.) This leaves an unacceptable independence between the two. If the mind is physical, then physical properties should determine mental properties, but for all that has been said, two events with the same physical properties might have different mental properties.
The standard response to this is to add in a supervenience claim. But apart from the well known problems with getting the strength of a supervenience claim right, I am inclined to agree with Kim when he says that supervenience does not provide an account of the mind / body relation. All it does is assert a certain pattern of covariance between mental and physical properties; it says that supervenient mental properties cannot vary without some change in the physical base properties upon which they supervene, and consequently that any mental change necessarily co-varies with some physical change. We need some positive account of the metaphysical relation between mental and physical properties, to which supervenience would be a consequence.

The ontological relation of realisation meets this requirement. Levine has proposed the following statement of physicalism in terms of realisation, which we shall adopt:

Only the fundamental properties of physics are instantiated in a basic way; all others, particularly mental properties, are instantiated by being realized by the instantiation of other properties.

To be instantiated “in a basic way” is just to be instantiated without being realised through the instantiation of another property. A property p realises a property q if and only if the instantiation of p constitutes the instantiation of q.

A familiar example of realisation is that of a functional property by a physical property; if the physical property plays a certain functional role, then its instantiation

---

8 Levine (2001) *Purple Haze: The Puzzle of Consciousness*, p. 12. Levine later qualifies this formulation to just say that “only non-mental properties are instantiated in a basic way” (p. 21), so as to obviate qualms about defining the “fundamental properties of physics”. However, it would be preferable not to tie a formulation of physicalism in too closely with its consequences for the mind, since its remit is wider than this. Let us simply record a preference for Levine’s stronger formulation, then, without going too far afield with the question of whether it is tenable, since this has little or no bearing on our project.
constitutes the instantiation of the functional property. The realisation relation is one-
many, since many different physical properties could play the same functional role, and
the instantiation of any of them would constitute the instantiation of that same
functional property. Talk of constitution reflects the idea that the realised property is
instantiated simply in virtue of the presence of the realiser property. The physical
property does not cause, or bring it about that the functional property is realised: it is
its realisation.

From now on, then, we will understand physicalism as claiming that all mental
events are physical events, and that all mental properties are realised by physical
properties. We shall also follow the standard and convenient practice of talking about
states, and sometimes processes. These are to be understood as particulars in the same
way as events, with the possible distinction that there is no implication of a happening,
or sudden change. However, the issues for physicalism about consciousness concern
the relation between mental and physical properties, and so our main interest will be in
the types that these particulars fall under.

1.e. The Idea of an Explanatory Gap

Philosophers of both inflationary and deflationary persuasion often make the
intuitive case for an explanatory gap in the same way, by asking “How could a physical
process be an experience?” or “How could a physical process be an experience?”

---

9 See Kim (1996) *The Philosophy of Mind*, p. 6, and Tye (1995) *Ten Problems of Consciousness*, p. 85. A state is a stage of existence of something. The terminology of “states” is popular because it is broad enough to encompass both mental phenomena for which it is not natural to talk of an “event”, such as a state of anxiety that lasts all day, and also those for which it is: we may properly speak of a man’s mental state at the moment he suddenly realises that he is trapped, for example.
Sometimes they stress the possibility, sometimes the identity. But either way, it looks as though what is being asked for is an explanation of an identity.

You cannot explain quantitative identities in any straightforward sense. Explanations have to stop somewhere, and whenever a request for an explanation is answered by a statement of identity, it is always time to stop. If someone asks why $X$ is $Y$, "It just is" will be the appropriate response, because true identities are necessarily true and no contingent state of affairs could shed light on the matter. This simple fact has provided David Papineau with the quickest possible deflationist means of short-circuiting the reasoning of V#2 which leads to an explanatory gap. The mind is the brain, he says, "and there's an end on it".\(^{10}\)

Still, the "how could $X$ be $Y$?" locution is widely used by those willing to accept that $X$ is indeed $Y$, not just in the explanatory gap debate but in life generally, so there must be some way of making sense of it. Take Nagel's example of a pre-Socratic believing that all matter is energy.\(^{11}\) The pre-Socratic may believe it all right, and what he believes may also be true, as we now think it is. What appears to be lacking is any justification for his belief. So perhaps explanatory gap intuitions result from our inability to justify physicalism about consciousness.

There is something in this suggestion, but it is not right as it stands, since physicalism about consciousness is justified by the wealth of experimental results which predict and explain consciousness in terms of brain functioning.\(^{12}\) And the pre-Socratic's belief that matter is energy could have been justified: suppose Athena had

---

\(^{10}\) Papineau (1993) *Philosophical Naturalism*, p. 121. Papineau develops this idea in greater depth in his (1998) "Mind the Gap". The point has also been made by Ned Block (1978) "Reductionism: Philosophical Analysis", and more recently in Ned Block and Robert Stalnaker (1998) "Conceptual Analysis, Dualism and the Explanatory Gap".

\(^{11}\) Nagel (1974) "What is it like to be a bat"?

\(^{12}\) There are many such examples in Patricia Churchland's (1986) *Neurophilosophy*, and her (1994) "Can Neurobiology teach us anything about consciousness?".
appeared and told him so. But even if she had, a crucial discrepancy would still remain between his understanding, and ours or the Goddess's. This is because the pre-Socratic has the answer, but lacks the workings-out. And when you do maths at school, it is the workings-out that get you most of the marks, because they show you understand the arithmetic you are practising.

The pre-Socratic cannot provide the workings-out which lead to the conclusion that matter is energy, because he lacks a background in the physical theory which explains matter in terms of energy. We, on the other hand, can provide a theoretical justification for our belief in this identity, or could at least get hold of one. The ability to provide such a theoretical justification shows our understanding of the relation between the concept of energy and the concept of matter. This is the understanding which the pre-Socratic lacks; he understands the identity of matter with energy just as well as we do, but he does not understand how to relate his concept of matter to his concept of energy.¹³

The relation between the concept of matter and the concept of energy which the theory provides is not a conceptual connection. There are no conceptual connections between matter and energy, in the sense that the way we ordinarily conceive of matter does not involve or in any way imply anything about energy. The way such theories relate concepts is rather more indirect. The theory shows how the behavioural characteristics which lead you to apply a concept to something, follow from the underlying microproperties of that something. When you know the theory, you understand how the behaviour which leads you to apply the concept matter to something, and the behaviour which leads you to apply the concept energy to

¹³ His concepts will need considerable updating before he can relate them as we do.
something, can both be explained by the same property. You now know not only that the concepts refer to the same property, but you also know how that property gave rise to your differing concepts.

The pre-Socratic who takes the Goddess at her word, has no idea how that single property referred to by his concepts *matter* and *energy*, gives rise to both the behaviour characteristic of matter and the behaviour characteristic of energy. He cannot justify his belief that the two concepts refer to the same property in terms of the property itself, by explaining how that property gave him his two concepts. He can only justify it by invoking something other than the property, namely the Goddess. And part of the idea of the explanatory gap is that we are in a similar situation as regards our concepts of experiences and physical processes. We cannot justify our belief that the concept of pain and the concept of cortico-thalamic oscillation both refer to the same property, in terms of that property itself. We have no theory which explains the feeling of pain in terms of physical microproperties. We have to appeal to the evidence provided by our discovery that whenever it is appropriate to apply the mental concept, it turns out to be appropriate to apply the physical concept.

We should, then, recognise two relevant sorts of justification. There is external justification for a belief, of the sort that is gained by looking up the answer to a mathematical problem in the back of the text-book. And there is internal justification for a belief, of the sort that is gained by working out the answer by using mathematical principles. It is someone's ability to provide an internal justification for their belief which inclines us to say that they understand what they believe. But we can only externally justify physicalism about the mind, by means of general theoretical considerations supporting physicalism, such as the causal closure of the physical
domain, and also from evidence that mental properties are co-instantiated with physical properties.

The relevant internal justification for physicalism about the mind would be the ability to provide an account of how physical properties could be such as to satisfy the mental concepts we have formed of them. So we should interpret the question “How could a physical process be an experience?” as asking “How could a physical process be such that it satisfies the concept of an experience?” What is being requested is a theory which justifies our belief that a mental and physical concept can both refer to the same property by explaining how each of the two concepts are applicable to the property in terms of that property. In other words, we need to explain what it is about the property which makes both concepts applicable to it.\(^\text{14}\)

This explains why the conjunction of V#1 and V#2 generates an explanatory gap. V#1 presented a conception of consciousness as a physical process. Then the reasoning in V#2 brought out another conception of consciousness, that of experience from the point of view of the experiencer. Here we have two differing conceptions of consciousness, with no explanation of how they can both apply to the same property. This worries the inflationists and deflationists alike, who are troubled by the apparent need for an explanation of the way physical processes of the sort talked about in V#1, gave rise to the conception of experience presented in V#2. Inflationists will emphasise the difficulty of explaining the ascendancy of a conception of consciousness replete with feelings, qualities, and a point of view, by reference only to objective physical

\(^\text{14}\) We want to explain our possession of the concept on the basis of the property it applies to. Our concern is primarily the property, not the concept; we do not want a sociological history of the concept’s origin, for instance.
objects and processes. And deflationists will play down the difficulties, put forward their own solutions, or make moves to discredit the problem.

Understanding the explanatory gap with reference to our inability to justify physicalism in a certain way, has the added advantage of tying in with the epistemological rather than metaphysical intent of the originator of the term “explanatory gap”, Levine. For Levine, there are gaps in our understanding which stop us explaining why certain neural processes constitute an experience such as the smell of coriander.

We can imagine the smell without the neural process, and vice versa, and this is the basis for what Levine calls “Cartesian conceivability arguments”, which employ conceivability as a premise to establish the metaphysical thesis that the two are distinct. But Levine takes the line that epistemological premises should only issue in epistemological conclusions, and that the conceivability in question only shows that qualitative character has not been adequately explained by physicalism. An adequate explanation would show how a particular neural process necessitated a particular smell of coriander. With such an explanation in hand, we could not imagine the neural process without the experience; the latter would be “epistemologically necessitated” by the former.

The sort of explanation which Levine envisions closing the explanatory gap, would show how the qualitative characteristics which lead you to apply experiential concepts, follow deductively from the underlying microproperties of the neural

---

16 Levine discusses “Cartesian conceivability arguments” (such as the famous anti-physicalist arguments of Kripke, Jackson and Chalmers) in his (1993) “On Leaving Out What It’s Like”.
17 This is of course controversial; Chalmers (1996) presents some sophisticated arguments which purport to show that epistemological premises can yield metaphysical conclusions in the mind / body case. We will look at his view of conceivability in Chapter 7.
process, within a Hempelian Deductive-Nomological explanation.\textsuperscript{19} When you have this explanation, you know that the qualities which lead you to apply the concept to the experience, and the observable behaviour which leads you to apply a neurophysiological concept, both follow from the same physical properties. You now not only know that the concepts refer to the same property, but you also know how that property gave rise to your differing concepts. Explicitly formulating the explanatory gap with reference to the derivation and aptness of concepts is little more than a variation upon Levine’s idea, then. But it does have the advantage of obviating qualms about explaining identities, as well as underlining the point that the gap is conceptual, and internal to physicalism.

1.f. Motives and Things to Come

In this introductory chapter, I first presented two views about consciousness which provide the intuitive backdrop for the explanatory gap debate. I divided all the philosophers engaged in the debate into two camps, the inflationists and the deflationists, on the basis of their attitudes to these views. I distinguished the causal from the explanatory mind / body problem, explaining how the favoured solution to the first problem has yielded the second problem. Then I presented a formulation of physicalism, and made the case that the explanatory mind / body problem is best understood as the epistemological problem of justifying our belief in physicalism in such a way as to show that we fully understand what we are committing ourselves to.

\textsuperscript{19} Levine considers qualifications to Hempel’s model, but holds onto the central idea that an explanatorily adequate physicalism must make mental properties deductively derivable from their microphysical realisers. See his (2001) \textit{Purple Haze: The Puzzle of Consciousness}, pp. 70-6.
My concern with V#1 and V#2 reflects my suspicion that the explanatory gap debate may be a product of specifically philosophical background assumptions. Inflationists and deflationists alike are struck by an obvious incoherence between V#1 and V#2. But many people lacking a background in philosophy demonstrate what can seem to be a complete inability to understand what the explanatory gap is supposed to be. And those with a background in science can come across as impenetrably insulated from intuitions of an explanatory gap. Such people just cannot ‘see’ the problem, and will endlessly frustrate the philosopher who tries to explain it to them. Even philosophy students, au fait with terminology like ‘explanatory gap’, ‘qualia’ and ‘what it is like’, can turn out to lack the intuitions these terms were coined to describe. But then they suddenly ‘see’ the explanatory gap problem for the first time, and just as suddenly, the thesis that consciousness is a part of the physical world does not seem so obvious anymore. They did not realise that physicalism was saying that.

Understanding V#1 and V#2 does not automatically yield intuitions of an explanatory gap. As a rule, the intuitions have to be prompted and brought to light. However, for the initiated, it is almost impossible to think of the source of the intuitions as anything but entirely obvious. And the way inflationists and deflationists alike react to those intuitions is nothing if not extreme. It is a common denominator between participants to the explanatory gap debate, that they find the idea of a conscious experience being identified with a segment of the physical world, utterly mystifying, totally baffling.\^20\footnote{At least to start with; deflationists typically go on to present a demystification.} Not only “seeing” this problem, but being “worried” or “bothered” by it, feeling its “pull”: these are the rites of initiation into the debate.

Dennett concedes that “It does seem impossible”, McGinn is struck by “a vertiginous
sense of ultimate mystery”, and Chalmers has no doubt that it is “the biggest mystery”. 21

These philosophers talk as though ordinary conceptions of the world are incoherent on a grand scale. We all go around immersed in a world of colours, sounds, feelings and thoughts, without the vaguest notion of how those colours, sounds, feelings and thoughts could possibly form a part of the physical world, whilst we nevertheless maintain an unshakeable conviction that the physical world is all inclusive. It is hard to see how anybody could miss this mystery, and easy to see why philosophers react so strongly to it. Yet most people remain oblivious to the mind / body problem, and are much more worried by philosophically inert questions like how the universe could be infinite. Perhaps they are just missing the point. Or perhaps there is simple way of understanding experience which does not clash so obviously with physicalism. 22 Maybe non-philosophers are missing something. Maybe philosophers are assuming something.

These are just some of the considerations which make me suspect that the explanatory gap debate ought not to be taken at face value. Another reason that disinclines me from jumping in on the side of inflationism or deflationism is that the actual debate is completely dead-locked between these factions. This intransigence looks to be a matter of principle, rather than a temporary lull for wont of a decisive argument. Particularly suggestive of this assessment is the increasing tendency of inflationist philosophers to deny that deflationist theories of consciousness are really


22 McGinn, Chalmers, and Dennett all disagree with me about whether non-philosophers are mystified by consciousness. The prefaces to their books all announce the universal appeal of the problem. Whoever is right (who knows?), it is still worth considering whether it might be possible to understand V#1 and V#2, and yet not be worried by the explanatory gap.
about consciousness. In an exchange between Searle and Dennett, we find the denial going both ways, as an inflationist and a deflationist explicitly disown each other's conceptions of consciousness. For Searle, Dennett "keeps the vocabulary of consciousness, while denying its existence", and so his theories are not really about consciousness at all. Dennett replies that his theory is indeed about consciousness, just not "Searle's brand of consciousness", for no scientific theory could be. Perhaps these are insignificant rhetorical turns, but if the two sides really are arguing at cross purposes, it would pay to get clear about why that is, rather than going straight to their arguments against each other.

In what follows I shall try to be fairly even-handed in my treatment of inflationism and deflationism: our purpose is only to understand them in order to provide a diagnosis of explanatory gap intuitions. If the diagnosis is right, then we should see how to lose the intuitions: this will be the acid test. The explanatory gap, and consequently inflationism, and deflationism, might be undermined at a stroke if we can reject a problematic assumption subserving them all. Then we could get on with the job of reconciling V#1 and V#2, so as to come up with a V#3 conception of consciousness.

---

23 See for instance Lycan's complaint, in footnote 17 of his (1997) "Consciousness as Internal Monitoring".
25 Dennett's reply to Searle in ibid., p. 118.
Chapter 2: Scientific Essentialism and the Mind

2.a. A Framework of Natural Kinds

Locke put Aristotelian essentialism into shape for the modern age. He distinguished between nominal and real essence. Nominal essence was the complex of ideas associated with the name of a type of thing. For example “gold” stands for ideas such as “a Body yellow, of a certain weight, malleable, fusible, and fixed.” Nominal essence is thus interest relative, in the sense that the ideas by which we pick out gold are contingent upon our interactions with gold. If we did not have colour vision, we would have a different nominal essence for the same thing, gold. What would make it the same thing would be its real essence, which Locke tells us is:

the constitution of the insensible parts of that Body, on which those Qualities, and all the other Properties of Gold depend.\textsuperscript{27}

We form the ideas that comprise nominal essences, which we use to distinguish between different kinds of thing, from the sensible properties of things. These sensible properties, as he says above, “depend” upon the insensible parts which comprise the real Essence of the thing. Elsewhere, he says that the sensible properties “flow” from the real Essence.\textsuperscript{28}

\textsuperscript{26} An Essay Concerning Human Understanding, Book 3, Chapter 6, §2.
\textsuperscript{27} op. cit., Book 3, Chapter 6, §2.
\textsuperscript{28} op. cit., Book 3, Chapter 3, §17.
Substitute "insensible parts" for "microproperties" and "sensible properties" for "macroproperties", and we find this idea at the heart of the contemporary explanatory gap debate. According to Levine, we explain the macroproperties of water by reducing it to $\text{H}_2\text{O}$, because this microproperty, conjoined with the laws of chemical microtheory, necessitate the instantiation of the macroproperties by which we identify samples of water. Water’s boiling at 100°C, one of its macroproperties, could in principle be deduced from the laws governing chemical microproperties, and initial conditions involving $\text{H}_2\text{O}$, kinetic energy, atmospheric pressure, and the like. The macroproperties would be the explanandum, and the microproperties the explanans within a D-N explanation.

On Levine’s account, we are unable to make physical or functional reductions of conscious states accord with this pattern of explanation, because we are unable to provide a theoretical account which reveals the “bottom-up necessity” making possible a deduction of macroproperties from microproperties and laws. We can reduce a macroproperty, pain, to a microproperty, cortico-thalamic oscillation, but we cannot then provide a D-N explanation which deduces the feeling of pain from cortico-thalamic oscillation. Physical microtheory can provide the details needed for an explanation of boiling, but not for an explanation of consciousness, and hence we have an explanatory gap.

This idea of an explanatory gap recalls the Lockean picture of sensible properties, like boiling or conscious feelings, flowing from an underlying structure which they are dependent upon. Locke himself despaired of deriving, and thereby

---

29 Levine (1993) “On Leaving Out What It’s Like”, p. 132. I have curtailed Levine’s model considerably, mentioning only what is needed to make the current point.
understanding the connection between consciousness and physical states of the body.

Hence we find him saying,

> the Ideas of sensible secondary Qualities, which we have in our Mind, can, by us, be no way deduced from Bodily Causes, nor any correspondence or connexion be found between them and those primary Qualities which (Experience shews us) produce them in us... 

This is a poignant turn of phrase for those familiar with standard formulations of the explanatory gap. And Locke arrives at it through his doctrine of real essence. Locke takes mental qualities such as redness to depend upon a real essence, which he considers, on empirical grounds, to be a primary quality of the body of the person experiencing red. He is sure there is a connection, but he is unable to deduce the experience from the bodily state, and so does not understand what the connection is. However, Locke faced an even more daunting explanatory gap than we do. For the crucial difference between Locke’s real essences and physical microproperties is that Locke thought of real essences as unknowable, or at the very least unknown, by mankind. He thought of our ideas of gold, or of the human species itself, as grossly inadequate approximations to real essences. Only God’s nominal essences would match real essences exactly; if we had God’s conception of humankind,

> our idea of any individual man would be as far different from what it is now, as is his who knows all the springs and wheels and other contrivances within of the famous clock at Strasbourg, from that which a gazing countryman has of it...

Thus Locke’s problem was not just the lack of a deduction from a physical essence to

---

31 Locke, op. cit., Book 3, Chapter 6, §3.
a conscious property; the physical essences themselves were also unavailable, on his view.  

Contemporary scientific essentialism, on the other hand, takes it that science has discovered real essences, or at least that current scientific microproperties provide a good approximation towards them. In a classic statement of scientific essentialism, Irving Copi is explicit in equating Locke’s real essences with scientific microproperties such as H$_2$O and DNA. Locke’s essentialism is adopted minus the pessimism about human understanding; the rapid progress of science since Locke’s time showing, according to Copi, that we actually can know real essences.

Interestingly, science plays the same role for Copi as God does for Locke, because scientific interests close the gap between nominal and real essence, scientific interests being so wide. Scientists want to be able to predict and explain varieties of behaviour under specified conditions, and such interests are best served by classifying things according to those properties which will facilitate the framing of a maximal number of causal laws and explanatory theories. Water is classified according to its microstructure, from which all the other properties it has in its relations to the multiplicity of human interests, can be causally explained. So science takes H$_2$O as the essence of water because more of the properties of water can be inferred from this property, than from any other.

Putnam’s externalist conception of the meaning of natural kind terms draws upon this intuition that deep structure makes things what they are, though Putnam’s concern was with the intuition itself, with speaker intention, and is thus neutral with

---


33 Irving Copi (1954) “Essence and Accident”.
Putnam’s idea was that in using natural kind terms, we intend to refer to whatever has the deep structure of certain paradigms of our acquaintance, thus introducing an indexical component into speaker intention. Users of natural kind terms are usually not the best placed to know whether some particular sample has such a structure, and so Putnam adds that part of speaker intention is for the extension of natural kind terms to be settled by the judgement of experts. The opinion of experts about the deep structure of a natural kind may change, but nevertheless, we mean to refer to whatever has the essence of that kind, whether or not anyone knows what it is exactly.

2.b. Essence and Individuation

Scientific essentialism is a realist position. It holds that science uncovers what things really are; those necessary properties a thing has just in virtue of being that type of thing. Essences are there to be discovered, existing independently of human concerns, but underpinning the disparate properties by which we may come to know them. However, the idea that in using natural kind terms it is our intention to refer to essences, suggests a link between the essence of the kind, and our way of conceiving of it.

David Wiggins provides a clear account of this link, saying that attributions of essences to objects are based upon our inability to conceive of them lacking those properties. Thus he provides the following necessary and sufficient conditions for being a possible accidental property and for being an essential property of an object:

---

X can be $\varphi$ just in case it is possible to conceive of x that x is $\varphi$; x must be $\varphi$
just in case it is not possible to conceive of x that x is not $\varphi$.\textsuperscript{35}

The reason that we cannot conceive of an object without its essence is that we
individuate objects by way of their essences. Take away the essence, and we take away
the object’s identity: there is no longer an object for us to conceive of.

Essential properties are necessary properties of an object. But it is necessary
for an essential property to hold of an object however that object might be described.
So the necessity is not de dicto, but de re. This is because essentialism is a realist
position: the necessity is out there waiting to be discovered, rather than something
which might or might not arise as a result of our descriptive practices. Wiggins
suggests that “we should see the de re necessity of essence as the limiting case of the
other de re necessities with which their form seems to conjoin them”.\textsuperscript{36} The other de re
necessities Wiggins has in mind are based upon capacities and obligations, for example
someone’s having to pay a debt, or someone’s not being able to stop themselves from
sliding down an incline. The debt and the sliding are unalterable, unavoidable. Essence,
for Wiggins, is so unalterable that a thing’s very existence depends upon it. And since
we individuate objects with essences, then we are absolutely constrained to describe
and think of them in such a way as not to exclude their having the essence which is
theirs, for otherwise we would no longer be talking or thinking about them at all, but
rather about some different sort of object.

“Essentialism, Continuity and Identity”.

\textsuperscript{36} Wiggins (1979) “Ayer on Monism, Pluralism and Essence”, p. 149.
This restraint on how we must conceive of things comes from our conceptual scheme: once such a scheme is in place, we are objectively constrained as to how we conceive of a thing, or we may fail to individuate it. But this in no way implies that it is not also to be found in the world. For the essence of an object is “bound up with the whole mode of articulating reality to discover such an object in reality”.\(^{37}\) We “articulate” reality by means of a conceptual scheme, but according to the realism inherent in essentialism, that scheme is tailored to and constrained by the world which we “discover”. We cannot know objects at all unless a conceptual scheme is in place, providing the necessary conceptual articulation for us to think about the world, and so if an object cannot lack its essence within our conceptual scheme, it thereby cannot lack it in reality as we know it either. Hence the essentialist can hold both that the origin of de re necessity is conceptual, and that objects themselves have de re necessary properties.

So we have here a general explanation of the basis for essentialism. It is that given our need to articulate reality, to individuate objects according to their properties, there must be some properties of any particular object that are prerequisite for individuation to occur at all. We can tell the essential properties according to whether we can conceive of the object not having that property. If we cannot, then the reason is that the object has been incorporated into our conceptual scheme because it has that property, and so could not exist without it.

The reason scientific essentialism has been so enduring is that it respects the basic intuition that what makes something the sort of thing it is, is how it is made up, whilst allowing room for empirical discoveries about types of make-up. We can refer

\(^{37}\) Ibid., p. 149.
to the essence of an object once we have categorised it with a natural kind term, whilst leaving open the possibility of finding new and better ways of describing that essence. And according to scientific essentialism, physical microtheory provides our best available descriptions of the inner make-ups responsible for the observable patterns which led us to distinguish the natural kinds in the first place.

2.c. Logical Consequences of Essentialism

If the origin of de re necessity is conceptual, then the difficulty we feel in talking about objects without their essential properties is a straightforwardly conceptual difficulty. If objects are individuated by their essential properties, if essential properties are what make an object the object that it is, then any talk of that object lacking its essential properties is conceptually incoherent. To talk of an object x without its essence F is to presuppose that x has F whilst simultaneously denying that it does. This is the same sort of conceptual incoherence that can be demonstrated more blatantly in terms of accidental properties, as if someone were to say “that man, who is a millionaire, is not a millionaire”, or in terms of artefact essence, as if someone were to say “that carburettor is not supposed to mix fuel and air”. Hence the incoherence, pointed out by Wiggins, of saying “Caesar was not a man”. For given that we think of Caesar as a man, there is no immediate answer to the question: “of what [are we] predicking ‘not a man’?”

---

38 Wiggins (1974) “Essentialism, Continuity and Identity”, p. 353. A plausible answer might be “whatever it was that people at the time referred to as ‘Caesar’.” Wiggins of course considers such objections, but our purpose here is not to assess essentialism, only to clarify its import.
There is another related conceptual incoherence which follows directly from the doctrine of essentialism. To draw out this consequence, we need to distinguish between an essential property and an essence. An essence is the conjunction of all of an object's essential properties. So F is an essential property of object x if and only if x could not exist as x without being F. And F is the essence of object x if and only if x could not exist as x without being F, and there is no other property G such that x could not exist as x without being G. Put intuitively, the idea is that an essential property is a property that an object could not lack, and its essence is all of the properties which it could not lack, that is, the minimum of properties required for that particular object to exist.

As a consequence, it ought to be plain that nothing can have two essences. The logical framework of essentialism precludes this possibility. The reason is simply that the essence of something is what it is and could not fail to be. Nothing can be and not fail to be two different things. The point can be demonstrated through the following reductio. Suppose object x could have two essences, F and G, where F is not the same property as G. Then since F is the essence of x, x need have no other properties apart from F in order to exist as x. So x does not need the property of being G in order to exist as x. But then, by definition, that means that G is not the essence of x, and so x does not have two essences after all.

---

39 This traditional distinction is also made use of by Nagel. See his (1986) *The View from Nowhere*, p. 48.

40 I shall ignore the Scholastic idea of “transcendental properties”, which were supposed to be essential properties that do not enter into the essences of things. Key examples are existence and temporal persistence; these were supposed to be properties objects could not lack, but which are so universally applicable as not to count as part of the essence of any particular substance. But we shall not count existence as a property at all, and will allow temporal persistence to count as part of an essence, along with other similarly general properties. For Descartes' treatment of transcendental properties, see Stephen Schiffer (1976) "Descartes on his Essence", p. 22 and ff.
This is just like the incoherence with speaking of objects whilst denying they have their essential properties. In both cases, we simultaneously presuppose that a property holds of necessity, and deny that a property holds of necessity. In both cases, we explicitly contradict ourselves. If, for some reason, we had an overwhelming reason to believe that both of the elements which generate the contradiction were true, we would have a good right to be confused. We might reasonably express our confusion by asking how both elements could be true, given that logic alone looks to preclude such a possibility. Something like this explains initial intuitions of an explanatory gap, as I shall now explain.

2.d. The Moulding of the Problem by the Framework

The root of the explanatory gap problem is a conception of consciousness as appearance. We shall begin to demonstrate and explore this claim in the next chapter, Chapter 3. Despite the tangle of irreconcilable difficulties associated with this conception of consciousness, we will find widespread commitment to it amongst inflationists and deflationists alike, though deflationist adherence is less overt. And we will also come to see the difficulty of abandoning this conception, since all the principal marks of the mind lead back to it.

The reason for pre-empting all this now is so that we can show how the themes presented so far fit into our analysis of the explanatory gap problem, those themes being scientific essentialism, individuation relative to a conceptual framework, and the idea of the explanatory gap as the need to provide internal justification for physicalism.
Given what we have yet to establish, that inflationists and deflationists alike conceive of consciousness as appearance, then if we impose the essentialist framework upon them, their commitment is to appearance being the essence of consciousness, or an essential property of it.\textsuperscript{41} This is because an attribution of an essential property to something is based on our inability to conceive of it lacking that essential property. In the case of conscious states, essentialist intuitions are at their strongest. For instance, it seems clear that we cannot conceive of a pain that does not appear to be a pain, that is, which does not feel like a pain, or have the “what it is like” character of painfulness.\textsuperscript{42} This idea of what is essential to consciousness might be thought to be implicit within V#2, waiting to be drawn out.

In fact, the intuition might be thought stronger than this, if \textit{all} that is needed for a state to count as a pain is that it feels like, seems to be, or appears to be pain. In that case, appearance would not just be an essential property of pain, but its essence. However, the appeal of this stronger intuition is somewhat mitigated by widespread adherence to V#1, for if appearance were all that is needed for consciousness, then disembodied pain might be a possibility, and the physicalist intuitions expressed by V#1 incline us to dismiss this.\textsuperscript{43} Conscious states must be essentially physical as well.

The intuition that appearance is essential to consciousness is problematic for a physicalist, if this is taken to mean that properties of appearance are partially constitutive of physical essences. This is because appearance is not a concept employed in canonical physical descriptions. In fully describing the intrinsic nature of a brain state, we have no need for the concept of appearance. So it would seem that the

\textsuperscript{41} Of course, some inflationists and deflationists reject essentialism, but the justification for understanding their views within this framework will have been given by the end of this chapter.

\textsuperscript{42} We have yet to make the case for commitment to appearance as the essence of consciousness generally, and so have picked on pain as the least controversial example possible.

\textsuperscript{43} Unless, of course, appearance could be a physical essence.
physicalist cannot identify conscious states with brain states without giving up either the claim that we can adequately describe physical essences without mentioning appearance, or else the requirement that appearance is essential to conscious states. This point needs elaborating.

If physical concepts are adequate to the physical properties they are concepts of, then in grasping the concept, you grasp the essence of the property. Concepts of brain states are not exceptional physical concepts, but are rather ordinary concepts applicable to certain portions of the objective world, not fundamentally unlike our concepts for types of inanimate matter. As such, these concepts make no use of the idea of appearance; our ways of thinking about brain states do not essentially involve us in thinking about appearance. In other words, you can grasp the physical concept without grasping anything about appearance: you can just think about the brain state as a brain state.

Now this does not mean, of course, that physical concepts do not in fact refer to appearances, for brain states might just be appearances. But this could only be so if our physical concepts are inadequate to brain states / states of appearance. For if the physical concept were adequate to the essence of the state, but not a concept of appearance, then appearance could not be part of the essence of the state.

This means that physicalists cannot identify conscious states with physical states, whilst holding onto the essentialist requirement that appearance be an essential property of conscious states, without accepting what we might call an “opaque physicalism”. This sort of physicalism accepts the inadequacy of our present physical

---

44 It will be developed much more fully in subsequent chapters.
45 An appeal to future physics might be made at this point; our subsequent discussion will bear on this idea.
concepts, by endorsing the hypothesis that appearance is part of the essence of certain physical properties. As Galen Strawson argues, if we are to make sense of our belief that physicalism is true, then “we must admit fundamental ignorance of the nature of the physical.”

If appearances are intrinsic properties that form part of the essence of physical states, then physicalism is an opaque doctrine because we have no understanding of appearance as a physical property. There must be something very important missing from our understanding of the physical world. (An alternative to this opaque physicalism would be to allow that both physical and irreducibly mental properties belong to conscious states: property dualism.)

If appearance is the essence of consciousness, then since no one thing can have two essences, we could be forced into an even more opaque physicalism. For then certain physical essences just are appearances, and our physical concepts of these essences must be entirely inadequate to them. (Alternatively, the conviction that conscious states have different essences from anything physical might move you to accept substance dualism.)

So a commitment to appearance as essential to consciousness means that if you formulate physicalism as an identity claim between states or properties, you may have to accept that physicalism is an opaque doctrine. This forces the hand of those who think physicalism ought to be an unproblematic doctrine, towards adopting a functionalist position. For if appearances are functionally defined macroproperties realised by physical states, then appearances can be essential properties of those states without compromising the adequacy of physical concepts, and thereby rendering

---

physicalism opaque. However, making this move will mean our having to qualify the simple definition of an essence as the conjunction of all the essential properties of a thing. We shall come back to this last point once we have further clarified the way in which the framework exerts pressure towards a functionalisation of mind.

Functionalism promises to resolve the unsteady combination of physicalism, a conception of consciousness as appearance, and physical concepts which include nothing of appearance. For if conscious states are individuated within our conceptual scheme by properties of appearance, then physicalism threatens to land us with the sort of incoherence we get in speaking of objects whilst denying they have their essential properties. In speaking of a conscious state, we presuppose that appearance is essential to the state, but in also upholding the identity of conscious and physical states, we seem to preclude the possibility of appearance entering into the essence of the state, since physical essences have no room for appearance if our physical concepts are adequate to them.

This tension is latent within the combination of V#1 and V#2, as they are commonly interpreted. Essentialist intuitions about the mind, deriving from V#2, result in our meaning to talk about appearances when we talk about conscious states. Physicalism, which commits us to something like V#1, makes us hold that conscious states are physical states. But if appearances are intrinsic properties and physical concepts are adequate, then this constitutes an implicit denial that conscious states have appearance-properties as part of their essence. And we cannot just renounce the properties through which conscious states are picked out in the first place, without embroiling ourselves in a logical contradiction within the essentialist framework.
This is a deeper reason than multiple-realisation for why an identity theory is untenable. Type identity claims are untenable because appearance must be essential to a conscious state, but cannot be essential to a physical state, unless there is something crucial that we do not understand about physical states. Multiple-realisation intuitions are parasitic upon essentialist intuitions about consciousness: we grant pain to the Alpha Centurion with D-fibres firing because we take it that he instantiates the appropriate appearance. But given that intuition about what is essential to pain, an identity claim is untenable in any case, because physical concepts do not incorporate appearance.47

We shall discuss the idea of functionalising the mind in detail, in subsequent chapters. But let us note for now that if appearance is conceded to be essential to consciousness, then this approach will only work if appearances are identified with functional properties, and not if functionalism is used as a premise for an identity theory between mental and physical properties. On the latter sort of theory, a physical property counts as mental if it satisfies the appropriate functional characterisation. But it is a contingent matter whether or not the functional characterisation is met, and so the intuition that appearance is essential to consciousness is not sated.

If on the other hand, we identify appearances with functional properties, then a physical property will realise an appearance-property if it satisfies the appropriate functional characterisation. This is a physicalist thesis that can meet the requirement that appearance is essential to consciousness, for the realisation of the functional property will be a requirement of state consciousness.48 However, problems then arise from the conceivability of physical realisation bases existing in the absence of

---

47 Untenable only for those not prepared to acquiesce in opaque physicalism, of course.
48 Or more strictly, it is a thesis compatible with physicalism.
appearances. These intuitions are a major stumbling block to physicalism, as we shall see, because of the special nature of the concept of appearance, which demands that such intuitions are given more credence than they might normally be given.

Further pressure towards functionalisation is exerted by scientific essentialism's presupposition that essences are underlying structural properties, the paradigm of which are physical microproperties. These are the deep properties from which observable properties "flow", as Locke said. Or on Levine's updated rendering of the framework, they are the properties from which observable properties can be explained.

But appearances do not even look like candidates to be underlying structural properties. They are not unobservable properties posited on the basis of a combination of theorising and observation, in order to explain patterns of observation. Rather, the concept of appearance comes from philosophy rather than science, and appearances are if anything supposed to be epistemically closer to us than observable macroproperties, rather than epistemically further away as are unobservable microproperties.

If appearances really were underlying structural properties, then we would have the more opaque physicalism mentioned above: appearances would be physical essences, our physical concepts would perforce be radically inadequate, and we would have to hold out for a conceptual revolution of some sort.

However, appearances could fit naturally into the framework if they were functionally defined macroproperties in need of explanation in terms of underlying structural properties. Then they would fit Levine's Lockean model like any other macroproperty. Let us illustrate this with an example.
For the scientific essentialist, gold is the element with the atomic number 79. This is the underlying structural property of gold: its essence. From this, we can explain its macroproperties. Take for instance the yellowish colour of gold. This is reducible to a property that is instantiated by being realised by the instantiation of the underlying structural property of gold. For this we need a bridge principle connecting the commonplace idea of a yellowish colour to a scientifically specifiable function. To this effect, we might reduce the colour of gold to a specific spectral reflectance function, though we would have to include qualifications about "normal conditions", to rule out the presence of coloured light, for instance. With the bridge principle in place, then if it can be shown that the element with atomic number 79 must satisfy this reflectance function, we will have demonstrated that the instantiation of this microproperty realises the yellowish colour of gold.

This macroproperty is an essential property of gold, since its instantiation can be shown to follow deductively from the instantiation of the essence of gold: we can in principle deduce the instantiation of the macroproperty from physical theory and initial conditions including the instantiation of the microproperty. This amounts to a D-N explanation of the macroproperty in terms of the microproperty.

However, though this macroproperty is an essential property of gold, for it is inconceivable that gold not instantiate this property, it does not seem right to call it part of the essence of gold, exactly because it is a functional property. Hence we will need to qualify the definition of essence we gave earlier. Given the nature of gold, it must satisfy a certain spectral reflectance function, just as given the nature of water, it must boil at a certain temperature in certain conditions. But reflecting light in a certain

---

way, and boiling at a certain temperature are not properties constitutive of what gold and water are.

So an essence cannot just be the conjunction of all of a thing’s essential properties. This could be a huge list, because a given property such as gold would necessarily satisfy any number of functions. Imagine a distant planet where there is no light, but something else called “t-light” which is not a stream of photons, but a stream of t-photons. Given the nature of gold, if we suppose that gold could be attributed something like a reflectance function for t-light, then whatever this function was, gold would necessarily realise it. But though this would be an essential property of gold, for it would be inconceivable that something be gold and yet not realise this function in the appropriate circumstances, we would not want to count it as part of gold’s essence.

So let us say that an essence is the conjunction of all the non-functional essential properties of a thing. Macroproperties like the colour of gold, the boiling point of water, malleability, transparency, and liquidity are all essential functional properties. They are realised by the instantiation of fundamental physical microproperties. Only these microproperties are instantiated in a basic way (and consequently are not multiply-instantiable), and so only they should be counted essences, or constituents of essences, within the scientific essentialist framework.

These microproperties themselves are of course extrinsically defined by the physical theories in which they are posited. The nature of the element Au is given by its place within physical theory. However, microproperties are not multiply instantiable: the theories which define them are detailed enough that nothing else could play the role of a particular microproperty except the microproperty itself. So microproperties are not themselves functional properties; rather the idea is that physical theory gives the
intrinsic nature of microproperties. Consequently microproperties are instantiated in a basic way: they are not realised through the instantiation of some other property.\footnote{Those unhappy with intrinsic properties \textit{per se}, could hold that all properties without exception are relational, whilst still remaining within the framework by counting only properties which are instantiated in a basic way as constitutive of essences.}

We can now see how the combination of essentialist intuitions about consciousness, and physicalism within the scientific essentialist framework, restrain the forms an answer to the explanatory gap problem can take. There are three basic options: opaque physicalism, functionalism, and the one we have not mentioned up until now, eliminativism. A physicalist who thinks appearances must be intrinsic properties will have to accept an opaque physicalism of some sort. This view is common to a number of prominent inflationists.\footnote{Nagel (1986) \textit{The View from Nowhere}, pp. 51-3; Strawson (1994) \textit{Mental Reality}, pp. 104-5.} But if physicalism is to account for appearance with ordinary physical concepts remaining intact, then appearances will have to be functional properties. Inflationists and deflationists alike favour this approach. Alternatively, if consciousness is inseparable from the concept of appearance, and appearance cannot be accommodated within the physicalist framework, then you may take the view that the ontology of mind will have to go, as some deflationists do.

Of course, if you do not think that appearance is essential to consciousness, then you would not necessarily be forced into this corner. However, abandoning the concept of appearance is easier said than done, and the bulk of the central section of this thesis aims to show the difficulties in avoiding this commitment. The other obvious way out would be to reject the framework of scientific essentialism. We shall look at this option in the next section, Section 2.e, only to find that it does not get you very far.
Of the three options above, only a functional account holds out any promise for consciousness to be incorporated into physicalism as an ordinary property like any other, as V#1 requires. The other two are more pessimistic. Both despair of fitting appearance into the physicalist framework we have, and conclude either that the framework must change, or that our concept of mind is irreparably flawed and must therefore be abandoned. Neither approach would be justified if it could be shown that the conception of consciousness as appearance which has been drawn from V#2 is a mistake, and that properly understood, consciousness presents no principled obstacles to physicalistic reduction. So our approach to opaque physicalist and eliminativist positions will not be so much to argue against them, as to undercut their motivation.

Now in our previous chapter, we saw that the explanatory gap problem is best construed as the need to provide internal justification for physicalism. We can now see what form that internal justification would have to take, given that essentialist intuitions about consciousness and the framework of scientific essentialism require that conscious properties be explained as functionally defined macroproperties. We originally said that to internally justify physicalism would require showing why mental and physical concepts are both applicable to the same property, in terms of that property. But now have we reneged property identities, this will need to be adjusted.

What now seems to be required is that we show why certain microproperties, for which we have physical concepts, must realise macroproperties, for which we have mental concepts. If we can show in physical terms why appearances must be instantiated, then we will understand how mental concepts can have application within a physical world. Or to get closer to our original formulation, we might say that if we
can explain why physical properties must instantiate conscious properties, then we will be able see how mental and physical concepts can both apply to the same event.

This, as we said before, is effectively how Levine interprets the situation. He thinks we have an explanatory gap due to our inability to deduce the instantiation of an appearance from the instantiation of a physical property. Levine interprets conceivability arguments about consciousness as a reflection of this. Once a macroproperty has been explained, he says, it becomes inconceivable that the microproperty be instantiated in the absence of the macroproperty. Given a sufficiently rich microphysical story about gold, it would be inconceivable that gold not look yellowish to us in normal light. It would, however, remain conceivable that something other than gold look the same colour: this reflects the fact that the yellowish colour is a multiply instantiable property. 53

But in the case of consciousness, we have no comparable theoretical connections to any microproperties. In fact, the conceptual distance between appearance and anything physical is so great that it is hard to envisage where we might start. A bridge principle between looking golden, and having a certain spectral reflectance property makes ground-level sense: the colour of a thing is to do with how it reflects light. But there are no such intuitive connections between neural functions and feelings, bar the imperative embodied in V#1 that there has to be one. Our ways of thinking about objective and subjective phenomena are so divergent that any proposal to connect them seems arbitrary.

We have seen that the essentialist framework makes something like Levine’s model for closing the explanatory gap inevitable. But when we reflect on the fact that

the root problem is to provide some connection between mental and physical concepts (an internal justification of physicalism), then it becomes apparent that Levine’s model has placed the emphasis for the philosophical problem in the wrong place. Essentialist intuitions are the driving force for the problem. Levine’s focus on the need for a deductive explanation simply reflects his expectation that the conceptual problem will be solved for philosophy from within science.

This is because what does the work in connecting the concepts is not the deduction from the micro to the macro property, but rather the bridge principle between the concepts of these properties. We do not think there is an explanatory gap between gold and its colour, for the reason that we can reduce “having a colour”, with some qualifications about normal conditions, to “having a reflectance function”, and we can do this because we understand the principle that the colour of a thing is to do with how it reflects light.

We need an equivalent principle to connect appearances with brain functioning, if mental and physical concepts are to be connected in such a way as to close the explanatory gap. This would have to be something we could intuitively see. Deducing the instantiation of an appearance from the instantiation of a physical microproperty is a secondary matter: this is something which would have to be possible, in principle, if brain states realise conscious states.

Levine is right that the possibility of deducing the instantiation of the macroproperty constrains conceivability intuitions, but only because conceivability tracks our intuitions about essential properties. We cannot conceive of gold without its colour, because we think this is one of its essential properties, and thinking this, we presume that the colour is a deductive consequence of the underlying nature of gold.
Likewise we can conceive of a physical state without an appearance-property, because we do not think appearances are essential properties of physical states. If we had some principle connecting an appearance with a physical state, then we might come to think it was essential to the state, and then we would expect its presence to be deducible from the instantiation of the state. This would constrain our conceivability intuitions.

So what is really required to close the explanatory gap, given the scientific essentialist framework, is some principle which connects the concept of appearance with our ordinary physical concepts. However, if the position we are going to develop is right, such a principle will never be discovered, for the simple reason that the philosophical conception of consciousness as appearance is a mistake, and we can understand consciousness without it. In fact, we can only understand consciousness without it.

2.e. Rejecting the Framework

We have seen how the framework of scientific essentialism and essentialist intuitions about consciousness mould the explanatory gap problem, and the form to which a solution must take. This suggests a very natural response on behalf of those who think that physicalism about the mind should be an unproblematic doctrine: reject the framework. Physicalism need not be a form of scientific essentialism. From a pragmatist perspective, for instance, we might construe physicalism as the view that our best descriptions of the world are framed in the vocabulary of physical science, where by “best” we mean something like “most useful”, in the broad, pragmatist sense
of “useful”, i.e. the least liable to generate contradictions, the most testable, achieving
the most explanatory coverage.

Pragmatism promises to undercut the explanatory gap problem at a stroke. The
problem is the combination of a conception of consciousness as appearance and
physicalism. But once we abandon essentialism, we are not obliged to conceive of
consciousness in this way by “the true nature of reality”. The difficulties appearance
presents for physicalism are, for a pragmatist, an obstacle to progress generated by an
obsolete way of thinking about the world in need of replacement by more useful modes
of description. A conception of consciousness as appearance is not foisted upon us by
the true nature of reality, but is just a component part of an historically contingent set
of metaphors adopted for talking about people, now no longer serviceable.

However, being a pragmatist about consciousness is easier said than done. This
is because once a conception of consciousness as appearance is abandoned, we lose
our grasp of what consciousness is supposed to be. The lack of conceptual connections
between appearance and the physical world which generates the explanatory gap,
translates into the problem for pragmatists of justifying any claiming to still be talking
about consciousness.

Rorty has correctly pointed out that a pragmatist explanation of X need not
“include statements about X”, for it can just “explain why people think there is such a
thing as X”. So for example, the Copernicans did not explain the movement of the
sun, but rather why people thought the sun moved. So a pragmatist need only explain
why people have thought that there are appearances, rather than explaining
appearances themselves. However, to explain why people have thought that there are

appearances, would be to explain why the concept of appearance has seemed to be applicable to certain physical phenomena. And this simply reiterates the need for an internal justification of physicalism.

So the pragmatist who acknowledges the need to relate the concept of appearance to the physical world, albeit under the apprehension that since the concept is flawed there cannot really be any appearances, faces a variant form of the explanatory gap problem.

Alternatively, the pragmatist may respond to the conceptual tension between physicalism and appearance with the suggestion that we just abandon the concept of appearance, and restrict ourselves to physical descriptions to see how much can be explained in this way, expecting that this will be more useful in the long run, and that interest in the concept of appearance will wane. Given what we have yet to establish, that our idea of consciousness is presently inseparable from the concept of appearance, this constitutes an eliminative approach.

So we have now seen two routes to eliminativism, one from scientific essentialism, and the other from pragmatism. The first takes it that the world is essentially physical, and from this, reaches the conclusion that nothing in the world can answer to the concept of appearance. The second abandons the vocabulary of appearance because it is incommensurable with the vocabulary of physicalism, which is deemed more useful.

But eliminativism does not dispense with the conceptual problem posed by the explanatory gap, it simply realigns it. By definition, eliminative materialism has no gap between the mental and the physical, but in holding that all phenomena can be

---

55 On the assumption that consciousness is inseparable from the concept of appearance, which is something we shall try to establish in subsequent chapters.
completely explained in physicalistic terms, eliminativism does not thereby rule out the possibility of an explanatory gap between sets of those phenomena. And this new gap can amount to much the same as the old one, and be just as problematical. The new gap will be a product of our need for physical theory to provide an explanatory account of the connection between the sort of physical phenomena that used to be loosely grouped together as “mental”, and the sort of physical phenomena that were not. To sustain a deflationary stance, the eliminativist needs the additional claim that there is no unity to mentalistic classifications. But in Chapter 4, we shall see that this claim is untenable: the concept of appearance provides all the prerequisite unity to get the problem of consciousness underway.

So pragmatism and eliminativism cannot solve the problem on their own. And the reason for this is that we have very strong essentialist intuitions about consciousness, embodied in the concept of appearance. No position will be satisfactory unless it relates to the concept of appearance: even if we ultimately do away with this problematic concept, we must first explain it.

The explanatory gap problem is a product of essentialist thinking. On the one hand there is the intuition that everything is essentially physical. On the other hand there is the intuition that consciousness is essentially appearance. The two intuitions conflict, and yield confusion about how “this” could be nothing over and above “that”, because we enter into the thought experiment with a definite view about what “this” and “that” essentially are.

This is why we will analyse the debate within the framework of essentialism. Of course, you can be a physicalist without being an essentialist, and it might ultimately be

---

56 Or if what were thought to be mental phenomena do not exist, then an account is needed of our dispositions to judge that they existed.
best to eschew any such weighty metaphysical commitment. Not only pragmatists, but also empiricists have found good reasons to reject essentialism.\textsuperscript{57} Moreover, the attributions of essences to individuals which are routinely implied in discussions of consciousness imply a particularly strong form of essentialism, for we would not normally think of a state or an event as having an essence.\textsuperscript{58} However, the fact that for most events, like a football match for instance, our intuitions about essence are unsure to say the least, only serves to underline the asymmetry with mental events or states, where we do have strong intuitions about individual essences. Essential to having a pain is how the pain appears, for instance.

In short, there may be all sorts of things wrong with essentialism, but that does not mean it does not provide the best framework for understanding intuitions of an explanatory gap. And in any case, a similar analysis of the explanatory gap to that which we shall provide could be made exclusively in terms of our concepts, with little loss except for simplicity. As Chalmers says in criticism of Kripke's argument against physicalism,

an essentialist metaphysics is inessential, except insofar as the feel of pain is essential to pain as a type - but that is just a fact about what "pain" means...\textsuperscript{59}

Nevertheless, the explanatory gap problem is usually more easily seen out in the open with the referents of our concepts, rather than in the concepts themselves.\textsuperscript{60}

\textsuperscript{57} D.H. Mellor argues against essentialism in his (1977) "Natural Kinds". Classic arguments for pragmatism can be found in Quine (1951) "Two Dogmas of Empiricism", Rorty (1972) "The world well lost", and Putnam (1987) "Why there isn't a ready-made world".

\textsuperscript{58} Lycan notes this, and remarks that, "events as such simply do not have individual essences unless their essences are very rarefied and elusive haecceities": (1987) \textit{Consciousness}, p. 17.

\textsuperscript{59} Chalmers (1996) \textit{The Conscious Mind}, p. 149.

\textsuperscript{60} Generally we shall talk about both concepts and their referents, though in Chapters 7 and 8, the focus will switch more to concepts.
A final consideration in support of our use of the essentialist framework is that abandoning it will not help us with the problem. Even if you reject essentialism, the intuitive gap between the subjective and the objective will remain just as large as it ever was. This is just as true for those who fully take on board a strong form of pragmatism, and realise that there is no “way the world is” which obliges us to employ the language of appearance. We may recognise that difficulties are presented by the concept of appearance, but it remains an indispensable tool for describing a natural way we have of thinking about the world. It may not be an essence, but without an alternative explanation, it retains a use. The freedom to abandon a form of description is unhelpful if we have nowhere else to go.
Chapter 3: Consciousness as Appearance

3.a. Immediate, Phenomenal, and Occurrent

Descartes thought that the essence of mind, or rather of a mind, since he considered every mind to be a particular substance, was thought. On his view, there would be no question of separating a person from his or her mind; Descartes took himself, Descartes, to be a mind, and so he took his own essence to be thought. The passage to this conclusion has three principle moments. First, Descartes finds that he can doubt the existence of everything except for his thought, and concludes that all he can be sure belongs to his essence is thought. Second, he develops his standard of clarity and distinctness, an internal hallmark of truth which is available to diligent introspection. Thirdly, since the mind can be clearly and distinctly apprehended apart from anything else, and God can create anything so apprehended, then a mind can exist without any properties but thought. So thought is the essence of mind.

The chosen Cartesian conception of mind, and the method of argumentation used to establish it are intimately connected. That conception turns on Descartes’ conception of thought, about which he is quite explicit:

Thought. I use this term to include everything that is within us in such a way that we are immediately aware of it. Thus all the operations of the will, the intellect, the imagination and the senses are thoughts.\(^{61}\)

---

\(^{61}\) The Philosophical Writings of Descartes, Volume 2, p. 113.
This conception of thought as that of which we are immediately aware is suggested by the method of doubt. For Descartes is able to doubt that his experiences are veridical, that they are true signs of the world which they seem to reveal to him, but he cannot doubt that he actually has those experiences. He does not know the status of his experiences, whether they are experiences of the world or mere phantasms, but he does know that the experiences exist. There is something of which he is immediately conscious, and this he called “thought”.

As has often been noted, Descartes’ conception of thought was revisionary. It included both thoughts about mathematics, and sensations of red. This was so as to cover anything which turned up within the “stream of consciousness”, anything available to be introspected. As the explanatory gap problem is usually presented today, sensations take centre stage as the paradigm of conscious states. But for Descartes the thought that two plus three equals five also counted as a thought in virtue of its conscious aspects, in virtue of being available to introspect.

So Descartes used consciousness in a broad sense to demarcate the mental. We may gesture towards this broad sense of consciousness as meaning: everything immediately given to the subject in experience, or everything available to be introspected. And here we see that though his conception of consciousness may be broader than some contemporary conceptions of the phenomenal, his conception of the mental is clearly narrower. For conscious states, even on Descartes’ broad conception, are only a subset of the states we ordinary classify as mental these days; we can have mental states that happen not to be conscious, such as beliefs which are not presently

\[62\] See, for instance, Anthony Kenny (1967) “Descartes on Ideas”.

\[63\] This phrase derives from William James. It is introduced in chapter XI of his Psychology (1910).
being thought about, and we can have mental states which are unconscious as a matter of principle, because they are repressed in the Freudian sense.

Rorty, who is a critic of the notion of the mental as a unified subject matter, once distinguished between two types of mental entity: those that are events and those that are not events. This might seem to give us the distinction between Descartes' thoughts, and other mental states. In the first category of events, there are occurrent thoughts and sensations. Occurrent thoughts are episodes of our mental life, of the sort reported when we say something like, "I just thought: we didn't turn off the heater". Sensations are distinct perceptions, visual imaginings, hallucinations, and anything else with a conscious phenomenology. Sensations and occurrent thoughts, Rorty says, "make up the content of the stream of consciousness."^5

Mental entities that are not events are such items as non-occurrent beliefs, moods, emotions, desires and intentions. These are attributed to people as a basis for explaining their actions. Of this latter category, Rorty says,

Not only are they not events, but it strikes one as an odd, peculiarly philosophical, hypostatization to think of them as particulars of any sort.\textsuperscript{66}

Attributing this sort of belief, the sort that is not an event, is attributing a disposition for Rorty. Such dispositions provide normative explanations of human behaviour, by explaining our movements as those of agents, rather than just as physical objects. The centrality of these states to the philosophy of mind in this century reveals a strikingly

\textsuperscript{64} Rorty (1970) "Incorrigibility as the Mark of the Mental".
\textsuperscript{65} Ibid., p. 407.
\textsuperscript{66} Ibid., p. 407.
anti-Cartesian turn: beliefs and desires, occurring or not, have become paradigmatic mental states.

Does that mean that an occurrent mental event is just a conscious mental state, in Descartes' sense? Not unless we can rule out unconscious mental events as incoherent in some manner. But though the unconscious is usually conceived dispositionally, it seems coherent to imagine something happening within a person's unconscious mind. An example would be somebody hearing of the death of their sister, and instantly blaming and forming a hatred for their brother-in-law, without being conscious of anything but grief for the sister. Such an event might not enter into the stream of consciousness. Rorty wants to distinguish between those types of mental entity which generate mind-body perplexities, "the paradigmatically nonphysical"^67^, and those that do not. But the occurrent / non-occurrent distinction, without further argument to disqualify unconscious mental events, does not seem to meet this requirement.

In a later treatment, Rorty grouped together all the occurrent mental entities through their possession of phenomenal properties, with all the non-occurrent dispositions lacking these.^^Rorty (1979) Philosophy and the Mirror of Nature, p. 22 and ff.\(^{68}\)^ The phenomenal / non-phenomenal distinction marks a much more significant division of mind than does the occurrent / non-occurrent distinction. The requirement of being phenomenal brings in its wake the requirement of being occurrent, for a merely dispositional state could not have phenomenal, "felt" properties. This suggests that the plausibility of restricting the "paradigmatically nonphysical" to mental events was not just due to the idea of something mental.

---

^67^ Ibid., p. 409.

happening, but rather something mental happening within consciousness, something of
which the subject is aware.

However, Descartes' criterion of a mental state that it, "is within us in such a
way that we are immediately aware of it", and the notion of a state having phenomenal
character which Rorty appeals to, are not clearly equivalent. For if occurrent thoughts
do have phenomenal character, it is not obvious that they do. There is no distinctive
quality associated with having the thought that two plus three equals five, though there
is such a quality associated with tasting chocolate, or looking at something bright red.
However, the thought that two plus three equals five does exist in the mind of the
person entertaining the thought, in such a way that the person is immediately conscious
of it. This is just as true with propositional thoughts as with sensuous experiences. To
reason from this that propositional thoughts must have phenomenal character is
spurious. Of course, we could stipulate that anything within the stream of
consciousness possesses phenomenal character, or else we should not be aware of it.
But this just seems to strip "phenomenal" of its significance. Phenomenal qualities can
strike us as something particular we are aware of; they are not a condition of
possibility for any such awareness.

We can save the specificity of "phenomenal" by glossing Descartes' "within us
in such a way that we are immediately aware of it" as there being "something it is like"
to be in such a state. There is something it is like to entertain a mathematical thought,
but from that, it does not follow that there is any phenomenal property associated with
thinking that thought. The entailment need not hold the other way either, if we can
make sense of an unconscious sensation, for example a pain that I am not attending
to.69

Let us call a mental state which there is something it is like to be in, a
conscious mental state. So if subject S is in a conscious mental state M at t, then there
is something it is like for S to be in M at t. Then we shall recognise three separate
distinctions: conscious and non-conscious, occurrent and non-occurrent, and
phenomenal and non-phenomenal. Conflating these distinctions serves to both muddy
the waters, and beg interesting questions. If a mental state is conscious, it does not
follow that it is phenomenal (it may be argued that conscious, occurrent thoughts are
not phenomenal). If a mental state is phenomenal, it does not follow that it is conscious
(it may be argued that there can be unconscious pains, for instance). If a mental state is
occurrent, it does not follow that it is conscious (it may be argued that there can be
unconscious mental events). However, if a mental state is conscious, it does follow
that it is occurrent. This is because of a conceptual entailment between the idea of
“something it is like” and the idea of “something happening”. “Something it is like” is
an incomplete locution which is satisfied by the specification of an event or process
which the subject is undergoing. There can be no: what it is like for nothing to occur
within consciousness (“nothing” as opposed to “nothing in particular”).70

The conclusion we have reached is that the central concept in understanding
consciousness, is the one Descartes proposed, immediacy, and this cannot be glossed

69 The need to distinguish states with qualia from other states of which there is something it is like to
be in that state, has been felt by a number of philosophers. See Kim (1996) Philosophy of Mind,
chapter 7, Tim Crane (1998) “Intentionality as the Mark of the Mental”, and Braddon-Mitchell and
70 If the mental state at t was not describable as an event or process, then there would be nothing that
S was undergoing at t, which could be said to be “like something”. There is a similar entailment from
being phenomenal to being occurrent. But if a mental state is occurrent, it does not follow that it is
phenomenal (it may be argued that occurrent thoughts are not phenomenal).
in terms of being occurrent or being phenomenal. The physicalist’s dilemma would be more tractable if it were. If being occurrent were the issue, then some criterion for mentality, manageable within physicalism, could be introduced, with the extra component of occurrence needed for consciousness. And finding a physical account of what makes a mental item count as an event should not pose any principled difficulties.

The same is true for being phenomenal, despite orthodox opinion to the contrary. For explaining why a state might be qualitative is not so antithetical to physicalistic explanation if the qualities can be attributed to external objects. This is the proposal of strong intentionalism, as we shall see in the next chapter.

But the more important concept for picking out consciousness, which these subsidiary mental concepts are related to and derivative from, is something intimated by talk of immediacy, or availability to introspection, or there being something it is like to be in some state. And here there really are difficulties for physicalism.

3.b. The Appearance / Reality distinction

In the last section, we introduced Descartes’ conception of “thought”, which we shall understand as a conception of consciousness, and as a criterion of conscious states. This concept of a conscious state was gestured at in a number of ways. We said, following Descartes, that conscious states exist in such a way that we are immediately conscious of them. Immediate would seem to be the operative term here, the opposite of which is mediate. I spoke in passing of conscious states being immediately “given”. I also used the word immanent, the meaning of which is opposed to transcendent.
Another gloss I used was that of being available to introspection. And finally, I used Nagel’s “what it is like” terminology.

All of these concepts, immediacy, immanence, being “given”, being introspectable, being “like something”: they all centre around one idea, though somewhat indirectly in the case of the Nagelian concept. This one idea is that of an appearance / reality distinction. Conscious states are appearances, through which a subject can be aware of reality. Appearance itself is what mediates the subject’s contact with reality, by offering a fallible representation of the state of reality. On this metaphysic, which springs naturally from Descartes’ method of doubt, reality is transcendent and appearance is immanent. Descartes can doubt transcendent reality, but cannot doubt the appearances which purport to open him to that reality. The fact of appearance itself is beyond doubt, for appearances are immediately “given” to the subject. The subject is aware of these appearances because it is “like something” for the world to appear in consciousness.\(^{71}\)

Of course, this conception of appearance is controversial, but it has its fair share of overt advocates in our times. Well known defences of this idea of consciousness as appearance are those of Kripke and Searle. Kripke’s argument against the identity theory was premised on there being no distinction between an apparent pain and a real pain.\(^ {72}\) If a subject had a mental state which felt like pain, which seemed or appeared to the subject to be a pain, then it just would be a pain. In effect, he said that the essence of the mental state of being in pain, is what it is like to undergo that state, or how being in that state appears to its subject. He takes the

---

\(^{71}\) A more complete account of the connection between lacking an appearance / reality distinction, and there being “something it is like” will be given in the next chapter.

individuation conditions of pains as solely determined by feeling, so that any particular
token pain could fail to have any property except the very feeling of which the subject
is aware. He could as well have said: everything which you find in introspection is
everything there is to that pain.

Searle also holds that conscious states are appearances, and thinks this explains
why they resist physicalist reduction. Standard ontological reductions of, for example, heat, colour and solidity, were based upon a prior causal reduction: a pre-
thoretical notion based upon perception, such as the subjective appearance of heat,
was found to be caused by an underlying microstructure. Then on the basis of this, an ontological reduction was made, which redefined heat in terms of its underlying microstructure, relegating the subjective appearance to an effect. But with consciousness, so goes Searle’s story, we cannot follow this pattern of reduction, because we cannot distinguish between the appearance of the property and the property itself. With heat, we are concerned with the property itself, and not with how the property appears to us. But the whole point of the concept of consciousness is to refer to appearance itself. Appearances are the “epistemic bases” that give us access to properties like heat, which it is the purpose of reduction to leave behind in the interests of attaining greater objectivity, but, Searle says,

reductions that leave out the epistemic bases, the appearances, cannot work for the epistemic bases themselves. In such cases, the appearance is the reality.74

Searle seems to think that appearance being the essence of consciousness is just part of the concept of consciousness. He admits that we might produce a causal

---

74 Ibid., p. 122.
reduction of the mind, correlating types of experiences with types of brain states. But he goes on to say that there would be no point in a further ontological reduction. This is because the concept of a conscious state is the concept of an appearance, for which there is no appearance / reality distinction to be made. However, there is an appearance / reality distinction to be made for brain states. Presumably, if physicalism is true, there would be both how the state appears to the person whose state it is, and a potentially infinite number of ways the state might appear to others. The former would be what the concept of consciousness was designed to pick out. But physicalism would leave this behind in favour of an underlying physical reality: the brain state itself. So appearance itself could not be physical.

Another useful way of thinking about this is in terms of the framework of scientific essentialism: in terms of Lockean real essences, and natural kinds. Conscious states are conceived of as states whose essence is to have no real essence. For real essence is exactly what is supposed to lie behind the multiplicity of possible appearances, the patterns of which lead to our groupings of natural kinds. That conscious appearances can have no real essence can be argued for in Kripke or Searle’s style. With Kripke, we can say that conscious states can be conceived in the absence of any physical microstructure, and so cannot have a real essence. With Searle, we can say that our concepts of conscious states are only meant to pick out how conscious states appear to us, and not any underlying reality they might possess. Both these views relate back to Descartes, who, on the basis of being able to conceive of conscious states in abstraction from anything else, forged a conception of conscious states meant only to pick out appearances.

---

75 We shall take a more careful look at Kripke’s conceivability argument in Chapter 7.
McGinn has noticed another good reason for thinking that conscious states cannot have real essences, which is that mental terms do not have the semantic features characteristic of natural kind terms. These characteristics derive from the tradition of scientific essentialism, as developed by Kripke and Putnam. One is that ordinary, folk criteria for inclusion within a natural kind, based upon perceptual and functional considerations (how an item appears), can be overturned by scientific investigation. So, for instance, we recognise that some substance might meet all the ordinary criteria for being gold, but turn out, on examination of its microstructure, to really be some other substance, "fool’s gold". Correlative to this is a certain deference to experts to decide the extension of our natural kind terms. An ordinary classification of a tree as an elm remains open to be overthrown by someone who knows better; it might turn out to really be a beech. Likewise, empirical investigation can trump ordinary groupings of natural kinds: glass has turned out to be a liquid, not a solid, and tomatoes have turned out to be fruits, not vegetables. In short, the extension of natural kind terms is determined not by their ordinary, ostensible features (their nominal essences, as Locke would say), but by their hidden nature, their real essence.

But mental terms have none of these features, on McGinn’s conception of them as designed for picking out appearances. Natural kind terms have their extensions fixed by real essences, reference to which was secured by surface features, or appearances. But mental terms simply name surface features or appearances. So the extension of mental terms is exclusively determined by how mental states appear. All that glitters is not gold, but everything that hurts is pain. If a mental state meets the ordinary criteria  

---

for being a pain, then it is a pain, and this fact is immune to revision on the basis of any empirical investigation into that state whatsoever.

This discrepancy between mental and natural kind concepts is reinforced by considering Twin Earth examples. In the standard case, speakers from Earth classify the substance XYZ on Twin Earth as water, because it meets their ordinary criteria for being water. But in doing so, they are making a mistake, for the extension of the term “water” is different in the two communities. The case is different with mental terms: speakers from Earth could apply their ordinary criteria for the application of mental terms to people from Twin Earth without the risk of making a mistake. This is because the extension of such terms is fixed by their ordinary criteria of application, and not by an underlying essence. As McGinn puts it, inverting Putnam’s conclusion about the meanings of natural kind terms, the meanings of mental terms are indeed “in the head”.

McGinn thinks mental terms have these semantic peculiarities because their use is for picking out appearances. In this he concurs with Searle, though McGinn goes on to say that because mental concepts are only supposed to apply to appearances, they are superficial concepts. Unlike natural kind concepts, they are not meant to apply to a deep structure about which empirical discoveries could be made. Hence their conditions of application are not susceptible to revision on the basis of such discoveries. Thus McGinn is of the opinion that,

Mental terms are not merely promissory notes hitherto unredeemed; they make no promises and are beyond redemption.

---

77 Putnam (1975) "The Meaning of 'Meaning'".  
79 Ibid., p. 133.
It is easy to see the problems this creates for the thesis that the conscious properties to which these terms refer are physical. For anything physical possesses an appearance/reality distinction, and so if mental terms referred to physical properties, there would have to be room for discoveries about the properties referred to by mental terms.

The way conscious states are conceived of, is deeply tied up with what they are supposed to be. They are conceived of as that which appears, and this conception leaves no room for them to be anything else except that which appears. This is why conceivability arguments, as used by the likes of Kripke, are accorded so much credence. And it is why Searle thinks he can rule out physicalist reduction just because of our concept of consciousness. The link is Descartes' methodology, by which he formed a conception of consciousness as that which he knew immediately.
Chapter 4: The Entrenchment of Appearance

4.a. First and Second-Order Essence

The Cartesian conception of the essence of consciousness that we have been looking at is that the essence of consciousness is its appearance to a subject. So, to take some concrete examples, the essence of pain is how pain appears to a subject, the essence of a particular pain is how that pain appears to its subject, and the essence of an occurrent thought is how that thought appears to its subject.

For Descartes, this essence is the essence of the mental, *simpliciter*. However, if a physicalist were to take this view without qualification, then it would seem to amount to the adoption of an opaque physicalism, as discussed in Chapter 2. This is because, as we said before, you can grasp physical concepts without grasping anything about appearance: you can just think about a brain state as a brain state. And so if physical essences are appearances, then our physical concepts must be radically inadequate to those essences.

This is one sort of inflationist position, but if we are to analyse inflationist and deflationist positions generally as resulting from a commitment to appearance as essential to, or the essence of, consciousness, then there must be some other way of combining this kind of essentialist presupposition with physicalism. One way is to eliminate, the thought being that since appearance would indeed be the essence of consciousness if it existed, then since physicalism is true, consciousness does not exist.
But the mainstream physicalist view, and the only option which presents any foreseeable hope for closing the explanatory gap, as we saw in Chapter 2, is to give a functional account of the mind. We said there that functionalism is in the sole position of being able to combine essentialist intuitions about the mind with physicalism within the framework of scientific essentialism, by treating appearances as functionally defined macroproperties in need of explanation in terms of underlying structural properties. This means treating appearance as a second-order essence.\(^80\)

First-order essences are intrinsic properties, which for the scientific essentialist are microphysical properties. A physicalist who takes appearance as only a candidate to be a first-order essence, either accepts opaque physicalism, or uses this as a premise for elimination. Second-order essences, on the other hand, are extrinsic or relational properties, realised by first-order essences. Examples include macrophysical properties like liquidity, and artefact kind essences, like being a carburettor.

We can now see how functionalism can combine essentialist intuitions about the mind with the essentialist intuitions behind physicalism.\(^81\) The former are accommodated by the idea that appearance is a second-order essence: the essence of conscious phenomena as kinds. The latter are accommodated by the idea that first-order essences are physical, and since second-order essences are realised by first-order essences, then conscious properties must be realised by physical properties.

This also serves to further explain why, in Section 2.d., we were prepared to count functional macroproperties as essential to certain physical states, but were wont to count them as part of the essence of those states. The reason is that such a

\(^80\) This terminology is employed by Michael Tye; it is introduced at pp. 47-8 of his (1995) Ten Problems of Consciousness.

\(^81\) This is not to say that physicalism is necessarily committed to essentialism, as we have already noted at Section 2.e.
macroproperty is not a constituent part of the first-order essence which realises it: it is not an intrinsic, essential property, but rather a functional, second-order essence. But nevertheless, given the instantiation of certain microphysical properties, certain functional properties are necessarily realised, and so functional properties can be essential properties of states that realise them: certain physical states (brain states for instance) cannot exist as the states that they are without realising certain functional properties.

An advantage of this way of looking at things is that it allows us to think of appearance as the essence of a conscious state, rather than just an essential property. This seems to be the intuition we have: all a pain needs in order to be a pain is that it appear to be a pain. Now this intuition (from V#2) is mitigated by physicalist intuitions (from V#1): if pain were an essence, then a disembodied pain is possible, and this does not accord with the physicalist world-view. But this is not so if pain is a second-order essence, for then we can accommodate physicalist intuitions by saying that such an essence requires a physical realiser.

Before moving on, we should note that most physicalists would deny that our conceptions of appearances are adequate to the second-order essences they denote. It is generally thought more plausible to count concepts of appearance as imprecise, folk conceptions of second-order essences, to be replaced by or reduced to more theoretical second-order essences. The replacement or reduction issue is dealt with variously. Functionalists of the empirical theory or commonsensical platitudes variety, may view the situation as one of replacement because they think we are unable to elucidate fully the functional essence of mental states. In general, the empirical theorists gesture towards a fully formed account of brain functioning, and the
common-sense theorists appeal to our implicit knowledge of the various interconnections between mental states. This suggests that whatever the second-order essences of mental states are, they are not something captured by imprecise concepts such as concepts of appearance.

Others interpret the situation as one of reduction. The idea here is that appearance is indeed the essence of conscious states: that property without which they could not exist as the states they are. But the property of being an appearance is discovered to be identical to a certain functional property, and since functional properties are second-order essences, not intrinsic properties, then appearance must be a second-order essence. This is reduction rather than replacement if our mental concepts of appearance are used to fix reference to functional properties, rather than giving or defining second-order essences in and of themselves.

4.b. Is the Mental Unified? A Taxonomy

Many philosophers would count the idea of appearance as the essence of mind too coarse grained, thinking that it glosses over the fundamental intentional / phenomenal distinction between types of mental state. A common view would be that appearance may be an appropriate essence for conscious phenomenal states, or simply for qualia, but not for other mental states. So it has been argued that the mind is not unified, or else that it is unified by some other property. Such views challenge the Cartesian essence. Here are the principal options:
(1) All mental states are “thoughts”, in Descartes’ broad sense. This view categorises phenomenal and intentional phenomena as mental, in virtue of their being conscious states, and excludes them if they are not.

(2) The mental is unified by consciousness. This view is very close to (1), except that it acknowledges the intuitive distinction between phenomenal and intentional states, and draws them together by imposing the requirement that if an intentional state or phenomenal state is to be counted mental, it must be conscious, or else be the sort of state that could become conscious. This final clause lets in the unconscious states which (1) excluded.

(3) The mental is unified by intentionality. This view acknowledges the intuitive distinction between phenomenal and intentional states, but draws them together by holding that all phenomenal states are ipso facto intentional states. Whether a mental state is conscious, on this view, is a matter of what type of intentional state it is.

(4) The mental is not a unified notion. Some mental states have phenomenal essences, and some have intentional essences.

(1) is Descartes’ view. (2) is Searle’s updated Cartesian view, which incorporates non-occurrent intentional states, unconscious states and the like, by means of their meeting the conditions of his “Connection Principle”. 82 (3) is Intentionalism. (4) encompasses views defended by McGinn, Chalmers, and Rorty,

---

amongst others. (3) and (4) jeopardise the contention that appearance is the essence of consciousness. If (3) were right, intentionality could be employed to demarcate all mental states, including conscious ones, and if this could be done without drawing on the idea of appearance, then physicalism would not necessarily have to provide an account of appearance. And if (4) were right, then depending on the degree of disunity admitted, there might be no distinctively mental essences for physicalism to account for at all, or at least all the states we classify as conscious might not have appearance as their essence.

4.c. Weak Intentionalism

Intentionalism, also known as representationalism, is, in its weakest form, the view that all mental states are intentional. Intentionality is claimed to be a necessary property of any mental state. This is the view we shall call weak intentionalism. A motivation for this view is to unify the mental, by finding something in common between an apparently mixed bag of conscious, unconscious, intentional and phenomenal types of state.

The obvious objection which the intentionalist has to overcome is that though some phenomenal states are clearly also intentional, it does not look as though they all are. For instance, an occurrent desire has both intentional and phenomenal properties, since being in such a state involves a feeling, and the state is directed upon something, namely the object of the desire. Or another example is enthusiasm, which is generally about something or another, as well as feeling a certain way. Sensations can also be

83 This terminology derives from the distinction between a “stronger” and a “weaker” thesis of intentionalism, in Crane (1998) “Intentionality as the Mark of the Mental”.

like this, for instance a bodily sensation of heat, which feels a certain way, whilst having heat itself, molecular motion, as its intentional object. But when it comes to the so-called “intransitive sensations”, the paradigm of which is pain, it looks as though intentionality is entirely absent. We cannot distinguish between pain and the sensation of pain, as we can between heat and the sensation of heat, for the familiar reason that there is no appearance / reality distinction to be made here. So the intentionalist must show that contrary to expectations, there really is such a distinction: the “intransitive” sensations must turn out to be transitive after all.

Armstrong’s way of doing this was to construe the apparently intransitive sensations as a type of perceiving. He analyses the pain report “I have a pain in my hand” as equivalent in meaning to “It feels to me that a certain sort of disturbance is occurring in my hand, a perception that evokes in me the peremptory desire that the perception should cease.” This account yields both transitivity, and an appearance / reality distinction, just as we have in the case of the sensation of heat, and heat itself. With pain, its feeling as if there is a disturbance in the hand is the appearance, while the disturbance in the hand, if indeed there is one, is the reality. Of course, there may be no correspondent reality, for this may be a phantom or a sciatic pain, and this is a consequence that bodes well for the intentionalist, since a traditional sign of an

---

86 Brentano made much the same move to intentionalise pain. He notes, in (1973) Psychology From an Empirical Standpoint, p. 84, that “we do not say that we sense this or that phenomenon in the foot with pain; we say that we feel pain in the foot”. This conflation of the sensation with the “phenomenon”, in Brentano’s sense, is just as confused as if we were to conflate a harmonious chord with the pleasure it typically elicits.
intentional state is that the intentional object is not guaranteed to exist. As Descartes would have agreed, feeling pain is no guarantee of bodily damage.

It is important to note where exactly this appearance / reality distinction lies. Though being in pain is now taken as an intentional state, Armstrong still agrees that there is "no distinction between felt pain and real pain". In other words, there is still no appearance / reality distinction to be made for the sensation itself. It is just that the sensation of pain turns out to refer beyond itself, just as the sensation of heat does.

What difference would it make to our position that appearance is the essence of consciousness, if weak intentionalism could be established? It could mean that intentionality would comprise at least a part of the essence of any conscious state. This in turn would mean either accommodating the fact within a developed conception of appearance, or if this could not be done because of some incompatibility arising, then it might mean abandoning appearance as the essence of consciousness.

Reflecting upon how an intentionalist such as Armstrong would like to understand such an apparently paradigmatic intrinsic state as pain as an intentional state, it becomes plain that no incompatibility will arise, and that we can take the former option mentioned above, of accommodating intentionality within a conception of appearance. For counting pain as an intentional state does not involve denying that pain is nothing more than what appears to a subject. All it does involve is affirming that what appears to the subject purports to be revelatory about some state of affairs.

Admittedly, physicalists such as Armstrong will then go on to use their intentionalist analyses to affect a reduction of the mental state to some underlying

---

87 To this effect, Brentano spoke of intentionality as "quasi-relational"; an intentional state only purports to take an object. See the appendix to Psychology from an Empirical Standpoint.
88 op. cit., p. 314.
reality; this “strong” intentionalist tactic will be discussed next. However, the weak intentionalist thesis that all mental states are intentional is entirely compatible with appearance being the essence of the mental. It might turn out that intentional appearances can be reduced to physical states, but that need not affect an account of what mental states essentially are, qua mental, so long as we make the first-order / second-order essence distinction.

In fact, there is more than compatibility here. Rather, the Cartesian conception of appearance actually presupposes weak intentionalism. The framework which Descartes sets up in the Meditations is an essentially relational conception of experience; the meditator reflects upon the world, and seeks to expound his relation to it. Implicit in this picture is what might be called the state-specific subject / object distinction. This distinction parallels such natural linguistic distinctions as those between seer and seen, between knower and known, between experiencer and experienced, and between believer and what is believed. What is common to each case is that subject and object are distinguished as a function of intentionality. The perceiver or thinker is classified as a subject in virtue of having mental states which are intentionally directed upon external phenomena. And that which is thought about or perceived is classified as an intentional object in virtue of having a perception or thought directed upon it. Consequently, whether an item is a subject or an object in the state-specific sense, is determined by its position within an intentional state. The reflecting meditator takes the subject position within all of his reflections.\(^{89}\)

\(^{89}\) There is another commonly used metaphysical subject / object distinction, which derives from the state-specific one. This is the distinction between that which is capable of taking the subject position of an intentional state, and that which is not. According to this distinction, people are subjects, and tables and chairs are objects, though people, tables and chairs can all be state-specific objects. Descartes notoriously included animals with tables and chairs, on this distinction.
Once this relational conception of experience is in place, it is easy to see why appearance should presuppose intentionality. The idea is just that for any appearance, there must be something which appears, and a subject that it appears to. Appearance is of something, to someone. The whole idea of an appearance / reality distinction, of immediate “givens”, of the world as a phenomenon for a subject: this is all part and parcel of a relational conception of experience, which has intentionality at its core. And once at least some mental states are conceived in line with the state-specific subject / object distinction, then any exceptions begin to look unstable. For if a feeling is not given to the subject as an intentional object, or is not itself an awareness of an intentional object, or a property of such an awareness, then we are owed some account of how the subject comes to know of its presence. Even if sensations are qualia, the subject must be related to those qualia in some way: clearly a subject cannot feel a quale if it is entirely unrelated to it.

Apparent counterexamples to weak intentionalism, sensations and emotions, for which it is not obvious what the intentional object is supposed to be, create their own problems if construed as non-intentional states. For if we suppose that a sensation is just a property of a subject, an instantiation of a non-intentional quale, then we still do not know why the subject should be conscious of that particular property. The subject has many properties, not all of which are conscious. The natural answer to why the subject should be conscious of this particular property is that it has appeared within the subject’s consciousness, on account of its being given, or actively attended to. But as soon as we accept that a sensation must be given or actively attended to, we can draw a subject / object distinction: between the person and the sensation which appears
within their consciousness, between the attender and the attended to. And then we have intentionality, defined as the relation between subject and object.

The defender of sensation as entirely non-intentional may reply that a quale just is a conscious property, a feeling, and so needs no connection to intentionality in order to be conscious. This response cannot, of course, be glossed as the subject being aware of that property, for then intentionality would have been introduced. But there is little conceptual distance between “just having a sensation”, and “being aware of a sensation”, and it is hard to see why we should insist on the former. In fact, the latter has a distinct advantage, based on the fact that we have a certain degree of control over our sensations: we can concentrate on them or try to ignore them. To accommodate this control, it is natural and perhaps indispensable to introduce intentionality. Then we say that the sensation itself is an intentional object, which the subject can concentrate on or try to ignore, or else that the sensation must be a component part of an awareness of some other intentional object, such as the state of one’s own body, so that the degree of attention paid to the body would affect the degree of sensation.

Intentionality, then, should be seen as a component part of any conception of appearance as the essence of consciousness, implicit within that conception. Weak intentionalism is a consequence of understanding experience in this Cartesian mould. Once mental states such as beliefs, occurrent thoughts, and perceptions have been understood in terms of a subject / object structure, as they invariably are in contemporary philosophy, then it is a foregone conclusion that similar descriptions will be discoverable for all other mental states.

---

90 See Crane (1998) “Intentionality as the Mark of the Mental”.
In fact, weak intentionalism has been a mainstay of continental philosophy since Kant. The “Refutation of Idealism” can be read as an argument for assigning intentional objects to all mental states. Inner perception presupposes outer perception, because all representations, including those of the self, “require a permanent distinct from them, in relation to which their change, and so my existence in the time wherein they change, may be determined.”\(^{91}\) Within Schopenhauer’s system, there is a heavy reliance upon the idea that a subject presupposes an object, and \textit{vice versa}, a typical expression of which is: “A consciousness without object is no consciousness at all.”\(^{92}\) Brentano went so far as to deny the coherence of experience that fails to exhibit the subject / object structure. He says of Sir William Hamilton’s description of feelings as “subjectively subjective”, that this is “an expression which is actually self-contradictory, for where you cannot speak of an object, you cannot speak of a subject either.”\(^{93}\) And Merleau-Ponty summed up this line of thought by saying, “there is no reason to ask whence I derive these ideas of subject and object, since they are simply the formulation of those conditions without which nothing would exist for anybody.”\(^{94}\)

Weak intentionalism has been resisted in analytic philosophy of mind mainly because of the influence of Chisholm’s linguistic reinterpretation of Brentano, which made the nonextensionality of a sentence or other linguistic context the sole criterion of intentionality.\(^{95}\) Propositional attitude reports create nonextensional contexts, since their truth conditions can turn on the particular concepts under which reference is made within the propositional clauses, and also because such sentences can be true

\footnotesize

^{91} Kant (1933) \textit{Critique of Pure Reason}, B276 (p. 245), as altered in the Preface (p. 34).
^{92} Schopenhauer (1958) \textit{The World as Will and Representation}, Volume 2, p. 15.
^{93} Brentano (1973) \textit{Psychology from an Empirical Standpoint}, pp. 89-90.
^{94} Merleau-Ponty (1962) \textit{Phenomenology of Perception}, p. 219. Merleau-Ponty is here characterising what he calls “reflective thought”.
^{95} Chisholm (1961) \textit{Perceiving: A Philosophical Study}, chapter 11.
even if singular terms in the propositional clause fail to refer. But the same is not obviously true of sensation reports. The truth-conditions of “John has a pain” operate just as if a pain was an object, in that the sentence, if true, will be true no matter how the pain is referred to, and cannot be true if there is no pain. However, Tye has found reasons for thinking that sensation reports contain a hidden nonextensionality. He noticed that more precise specifications of sensation-types, “a pain in the foot” rather than just “a pain”, behave like content specifications, and create nonextensionality in much the same way. An alternative to this approach which some philosophers have adopted, has been to simply drop the linguistic criterion of intentionality altogether.

4.d. Strong Intentionalism

The stronger version of intentionalism says that mental states of all sorts, including conscious experiences, can be completely characterised as intentional states. Intentionality is claimed to be the essence of the mental. An additional motivation for this view, over and above the unificatory motives of weak intentionalism, is that it promises to dispense with the need for intrinsic qualitative properties, or qualia. The phenomenal character of an experiential state is fixed by the type of intentional content it has. And intentionalists who distinguish, as we have done, between a state’s having phenomenal properties, and its being conscious, will also be committed to non-phenomenal, but nevertheless conscious, “what it is like” properties, being characterisable within a specification of intentional content.

---

98 On some versions of intentionalism, the attitude taken to that content is also a contributive factor in determining qualitative character.
Holding that mental states can be entirely characterised, without residue, in terms of their intentionality has distinct advantages for the physicalist. For if the strong intentionalist adheres to a reductive account of intentionality, as most of them do, then all the obstacles presented to a reduction of mind by apparently non-intentional properties, qualia, would have been overcome. Strong intentionalism can then be seen as a variety of functionalism, with intentionality being given a functional analysis in terms of causal covariation, asymmetrical dependence structure, teleological function, or the like.

There are both externalist and internalist versions of strong intentionalism. According to the externalist version, a phenomenal property is the property of representing some external object as having a certain objective property, such as being hot, or coloured. Thus on the externalist version, phenomenal content is world-dependent. According to the internalist version, a phenomenal property is a part of the intentional content of a state, a content which is determined by its role within the cognitive economy of the subject of that state. Thus on the internalist version, phenomenal content is world-independent.

The internalist version just supplements the standard functionalist claim that qualia can be functionally characterised, by assimilating phenomenal to intentional properties, on the assumption that functionalising the latter should prove relatively unproblematic. It is the externalist version which is the more popular, largely due to the fact that most reductive accounts of intentionality are externalist in character. But it also offers a distinct advantage over other versions of functionalism. A persistent problem for functionalism has been that qualia seem to be intrinsic properties, and so stubbornly resist functional characterisation. This is reflected in the conceivability of
systems that meet all the functionalist’s criteria for a conscious mind, and yet entirely lack qualia, and in the conceivability of two functionally identical systems having inverted qualia with respect to each other. These arguments testify to our ordinarily conceiving of qualia, and conscious properties more generally, as having an essence which is independent of any functional roles that they might happen to play.

However, in addition to our intuitions that appearances are intrinsic properties, we also have intuitions that they are multiply realisable. Externalist strong intentionalism promises to accommodate both of these intuitions. As Levine has noted, the appeal of the view is that it,

... seems to combine multiple realizability with intrinsicality, by displacing the intrinsic property onto the external object.\(^9^9\)

According to such a view, the quale itself, a property of a representation, is multiply realisable, because it is a functional property. But the content of the representation is an intrinsic property, because it is an ordinary, physical, objective property, such as heat or colour. This accommodates the intuition that qualia are intrinsic properties, because it posits a necessary link between being a certain quale, and representing a certain intrinsic property.\(^1^0^0\) A visual representation with the phenomenal character of being red, just is a visual representation of objective red, which is an intrinsic property, and so to have that phenomenal character, the state must represent red, though it may be realised in any number of different ways.

---

\(^1^0^0\) I shall ignore the prerequisite epicycles needed to account for misrepresentation, as they are not relevant here.
But another way of explaining the co-existence of multiple realisability and
intrinsicality intuitions, is that conscious states are conceived of as essentially different
from physical states, and yet empirical evidence for conscious states and physical states
always going together has given us supervenience intuitions: we think physical changes
are needed for mental changes. So we think of conscious states as having an intrinsic
nature (they are appearances), and we think they owe their existence to physical states.
But a bare intuition of essence coupled with one of supervenience, will impose no
restrictions on the type of physical state required for mental states to supervene upon.
Thus we have the multiple realisation intuition.

Now it may be conceded that the intrinsicality intuition is at root a dualistic
intuition. That is why many deflationists want to give an entirely extrinsic account of
the mind, with no reference to intrinsic properties. But the present suggestion is that
the multiple realisability intuition is equally dualistic. We have it not because we
conceive of mental states functionally rather than physically, but rather because our
conception of the mind is isolated from the physical world, save for the requirement
that something physical be going on to support the mind. Our intuitions are born of an
uneasy allegiance to both V#1 and V#2.

If this is right, then there is no reason to expect that the externalist strong
intentionalist has an alternative to a conception of consciousness as appearance. And in
fact, as we shall see, it is no alternative, but rather an account of appearance which
must appeal to the conception of appearance we already have.

Strong intentionalism’s claim that intentionality is the essence of conscious
states must rely on an appeal to appearance, because it is not good enough for a mental
state to have a particular intentional object for it to be phenomenal. We cannot say, for
instance, that as long as the content of a state is objective red, the state will have a red
phenomenal character, because we can have all sorts of mental states about red, not all
of which need be phenomenal. So since you can have both phenomenal and non-
phenomenal states about exactly the same thing, the simple fact of having an
intentional state about that thing cannot be enough for being in a conscious state. To
have a conscious experience of red, you do not just have to represent red (you do this
when you think to yourself that red is a dramatic colour), but you have to represent red
in a certain way.

What this “certain way” is cannot be spelled out in terms of the intentionality of
the state, that is, what it is about. Rather, it is to do with how the state represents its
object. If the state represents objective red in the right way, then the representation
will have a certain property of redness. But this property is what we already grasped
before the intentionalist analysis: it is an appearance of red. Consequently, strong
intentionalism (for physicalists) is simply undertaking a reduction of appearance, and
relies on our conception of consciousness as appearance in order to do so.

The point is that for the strong intentionalist, intentionality is the essence of all
mental states, both conscious experiences and non-occurrent beliefs. So the question
arises of what is supposed to make some mental states, and not others, conscious
mental states. The answer the intentionalist will give is: the sort of intentional content
it has. According to Tye, for instance, conscious states are those states with “poised,
abstract, nonconceptual intentional content”. But this just amounts to a functional
difference between conscious and other mental states, and in that case, being
intentional is not enough to make something conscious for the reductive strong

---

101 Tye equates phenomenal character with “what it is like”. For his PANIC theory, see esp. pp. 137-
intentionalist. Rather a conscious state is an intentional state which functionalises in a certain way. And the only means we have for telling whether we have the right functionalisation, is whether or not an appearance is realised. So effectively, the strong intentionalist has given us no alternative to a conception of consciousness as appearance, reference-fixing on certain phenomena which we then try to functionalise.

Another reason that strong intentionalism's claim that intentionality is the essence of conscious states must rely on our conception of appearance is to do with the fact that the mental apparently forms a proper subset of the intentional. Writing, signs, and symbols generally have all the key elements associated with intentionality; they have relational structure, and gesture beyond themselves to something else which they represent. They also pass the linguistic criterion of intentionality; sentences about sentences can be nonextensional. But if intentionality were the essence of the mental, then having intentionality would be enough to make an item mental. So if intentionalists are not to count symbolic items as mental, then they must show that those items are not intentional, or not intentional in the required sense.

There are two ways to go here. The first is to accept that minds, symbols, books, and artefacts of all kinds, such as chess computers, and even thermometers, all have intentionality in the same sense. This view is most closely associated with Dennett. If the strong intentionalist takes Dennett's view, that all intentionality is in the same boat, differing, so to speak, quantitatively rather than qualitatively, then their view is liable to collapse into functionalism. This is because if some intentional states are conscious (states of humans) and some are not (states of chess computers), then being intentional cannot be enough for a state to be conscious. The natural suggestion

---

102 See his (1987) "Evolution, Error, and Intentionality".
for this extra requirement is that the state be the sort of intentional state that
functionalises in a certain way. But then function is the essence of the state, and
intentionality is just an essential property.

Alternatively, a strong intentionalist could make Searle’s distinction between
intrinsic and derived intentionality. On this view, words, for instance, have no
intrinsic meaning; they are just marks on paper, and mean something only as read and
interpreted by a human being whose mind is intrinsically intentional. So their
intentionality, as with all artefacts, is entirely derived. On such a view, it would be
intrinsic intentionality, and not intentionality simpliciter, that is the essence of
conscious states.

The question remains, however, of why certain states, namely human mental
states, should be intrinsically intentional. Dennett thinks that there is no good reason to
mark off some states as intrinsically intentional, only piece-meal intuitions about
artefacts being “our tools”, whose intentionality is a “by-product of our practices”. This is a “persuasive theme”, but “not really an argument”.

However, there is a concealed argument here for marking off intrinsic mental
states from other intentional states, and the argument relies crucially upon an appeal to
consciousness. The argument, ignoring for the moment an important qualification, is
the following. For any conscious occurrent thought, there is no distinction to be made

---

distinction between “intrinsic”, “as-if”, and “derived” intentionality. Sentences have derived
intentionality because they “literally” mean what we say they do, but systems like thermometers and
computers only behave “as-if” they had intentionality, and “as-if” intentionality is not a kind of
intentionality.” (p. 78) Searle never explains why he thinks there is an important difference between,
for instance, a thermometer registering that it is hot, and a sentence saying that it is hot, so we shall
ignore this distinction, and just talk about intrinsic (or “original”) and derived intentionality, as
Dennett does (Dennett 1987) “Evolution, Error, and Intentionality”).

104 Though Searle for one does not think that an artefact could not have intrinsic intentionality. If one
could be built with equivalent causal powers to a human brain, on his view, that would be enough. See

between it seeming as though the thought has a content, and the thought really having a content. With any artefact, on the other hand, whether it is intentional depends upon factors extrinsic to it: it is determined only within a wider context. So for example, the word “papa” means different things in French and Portuguese, and for a people with no written language, may just be a shape bearing no intentionality at all. Hence there is an appearance / reality distinction to be made for the intentionality of artefacts: in reality, they have no particular intentional content, but they do as they appear to us. Conscious thoughts and perceptions, on the other hand, involve appearances of their intentional objects, perhaps experientially, perhaps as a component of a proposition. Since there is no appearance / reality distinction here, they must really have content, and so be intentional states. And since the content is given in appearance, interpretation of such a state is not up for grabs; rather it comes built in.

The important qualification is that the concept under which the intentional object is subsumed may be inappropriate, and in this sense there will indeed be an appearance / reality distinction to be made: between what the intentional object really is, and what the subject takes it to be. For instance, I may take myself to be looking at a blue book, but in fact it is a red book in strange light, or a projected image of a blue book. The intentional object of a mental state may even be fictional. However such cases work by comparing the mental state to the reality it purports to be about. When we restrict our attention to the mental state itself, this discrepancy does not arise. If the perception seems to be of a blue book, then it is a perception “as of” a blue book, and there is no distinction to be made between seeming to be a perception “as of” a blue book, and really being a perception “as of” a blue book. There is no question of its not being an intentional state, only of whether the intentional object is what the subject
takes it to be. Just as a pain in the foot may turn out not to be correctly described as
“in the foot”, but is unqualifiedly a pain as if in the foot, so a perception of a blue book
may turn out not to be correctly described as “of a blue book”, but is unqualifiedly a
perception as of a blue book. This same point could be reformulated for the case of
propositional thoughts.

In short, if it seems to you that you are consciously thinking about something,
then you are consciously thinking about something, even if you are mistaken about
what that something is. Even if the table you are thinking about turns out to be a
mirage, it would remain the case that your thought was intentional. Thoughts are
intrinsically intentional when they are before the mind, for appearances are appearances
of something. An artefact, on the other hand, merely derives its intentionality by having
the capacity to cause conscious intentional thoughts.

This seems to be the thought behind marking off some states as intrinsically
intentional. However, intrinsic intentionality as the essence of consciousness, comes to
the same thing as appearance as the essence of consciousness, qualified by weak
intentionalism. So once more, we see that strong intentionalism does not allow us to
do away with appearance: this problematic conception remains integral to any such
model.

4.e. Disunity of Mind

Intentionalism’s challenge to our interpretation of the explanatory gap debate
as tied to a conception of consciousness as appearance, was basically that we can do
without appearance and demarcate the mind with intentionality instead. The next
position to consider, number (4) on our taxonomy, comes in two varieties. The second of these says that the mind is split between two types of mental state, phenomenal and intentional, and we shall consider it in Section 4.f, the next section. In this section, we will look at the view that even different phenomenal and different intentional states have nothing in common, and that we have simply made a mistake in thinking there is any unity to our mental classifications.

According to this view then, our conception of consciousness as appearance is based on an error. Usually, philosophers who take this sort of view argue along the lines that since our conception of consciousness as appearance is unworkable in the context of contemporary physicalism or cognitive science, in the sense that it resists reduction or that the categories of appearance mark no salient physical or functional distinctions, then as a consequence, the ontology of consciousness must be eliminated. The emphasis is on the failings of conceiving of consciousness as appearance. However, as we said back in section 2.e., there would be no need to follow this eliminative path if we could show that there is another more innocuous way to conceive of consciousness.

Rorty is unusual in trying to give an account of where we went wrong in conceiving of consciousness in the first place, which is much closer to the line we are taking. However, though we may ultimately agree that conceiving of consciousness as appearance is a mistake, the position we will defend is nothing like Rorty’s, and has no such negative consequences for the unity of our concept of mind.

Rorty agrees with us that the feature which marks off the phenomenal from the physical is that “in the case of phenomenal properties there is no appearance-reality

---

distinction". But Rorty takes the import of this to be that states with phenomenal
properties bear the epistemic characteristic of incorrigibility. Mental events, which all
have phenomenal properties within Rorty's picture, are distinctive in that,

certain knowledge claims about them cannot be overridden. We have no
criteria for setting aside as mistaken first-person contemporaneous reports of
thoughts and sensations, whereas we do have criteria for setting aside all
reports about everything else.\footnote{Rorty (1970) "Incorrigibility as the Mark of the Mental", p. 413.}

We are incorrigible about how our own conscious states seem to us. This for Rorty, is
just a linguistic practice, a rule of a language-game marking the predictive advantages
of allowing first person testimony to trump the usual third person behavioural
considerations as a basis for mental state attributions.

As a purely epistemic mark of the phenomenal, this does not tell us why it
should be supposed that consciousness is essentially non-physical. For this
incorrigibility might be a temporary affair, contingent upon our ignorance of
psychophysical laws. In the future, the use of cerebroscopes might become so reliable
that the third-person evidence they provided would trump first-person
contemporaneous reports of thoughts and sensations, leading to the emergence of an
epistemic appearance/reality distinction where there was none before.

This is exactly the basis of Rorty's sceptical position; he thinks that from a
feature of linguistic practice, a fallacious inference was made to an essentially non-
physical conception of consciousness. From the fact that practice rules that people
cannot be mistaken about the mental events they are undergoing, so that there is no
distinguishing how my conscious states seem to me from how they actually are, a

\footnote{Rorty (1979) Philosophy and the Mirror of Nature, p. 29.}
metaphysical conception of a mental event as something which is what it appears to be was born. Since physical events can always be other than they appear to be, it followed that mental events could not be physical unless a substantive theory could be produced which explained away this apparent incompatibility.

If this were right, then thinking of appearance as the essence of consciousness would be a mistake. There would be a metaphysical appearance / reality distinction to be made with conscious states, as with everything else. All that would mark conscious states off from physical states would be a contingent linguistic practice, based upon the epistemological feature of first-person privileged access. In fact, Rorty goes so far as to claim that “the purportedly metaphysical ’problem of consciousness’ is no more and no less than the epistemological ’problem of privileged access’”.

One of Searle’s discussions bears interesting comparison with this stance. Searle explicitly targets Rorty for criticism on the grounds that sincere first-person reports of mental states are not incorrigible; he gives the example of somebody coming to realise that they had not been in love, but had rather only been infatuated, and so mistaken about the contents of their own minds. This is wide of the mark as a criticism of Rorty, who is as deflationary as could be about strict incorrigibility, and employed the concept as part of a critique of the thesis that the mental is a unified notion. But in any case, what is really interesting is Searle’s diagnosis of a mistake concerning incorrigibility, which is the exact opposite of Rorty’s. Rorty says that something epistemological, incorrigibility, led illegitimately to an ontology of entities for which there is no appearance / reality distinction. Searle, on the other hand, says

111 Ibid., p. 145.
112 Besides, in the article Searle cites, Rorty actually makes the point that emotions are only “almost incorrigible”: Rorty (1970) “Incorrigibility as the Mark of the Mental”, p. 420.
that something metaphysical, the subjective ontology of entities for which there is no appearance / reality distinction, led illegitimately to the epistemological notion of first-person incorrigibility about those entities. How should we adjudicate here?

We need to know what exactly is supposed to be the passage from incorrigibility to a lack of an appearance / reality distinction, and what exactly is supposed to be the passage from a lack of an appearance / reality distinction to incorrigibility. Well in the latter case, the basic idea seems to be simple enough. It is that conscious properties being appearances leaves no room for the subject of those properties to make mistakes about them: the appearances cannot be misleading guides to reality, since the appearance is the reality.

However to adopt this line of reasoning, according to Searle, is to make a mistake. We are not incorrigible about our own mental states. But this is not because the appearance misleads us as to the reality; Searle agrees that there is no appearance / reality distinction to be made. But there are other forms of mistake that we can make in self-ascribing mental states. These are all forms of misdescribing one’s own mental state. For instance, in self-deception, an overriding desire not to have a certain mental state, such as hatred, can make the subject ignore this aspect of their mental state, and thus misdescribe it. Or a subject may misinterpret his feelings, classifying them as feelings of love for instance, because he fails to correctly relate them to his other mental states. In such cases, the mistake does not rest on appearance being misleading as to reality. How the state seems constitutes what it is: a state that does not seem like hatred, a state that seems like love. But the subject may be mistaken in how they describe their state: as not involving an element of suppressed hatred, as really being a case of love.
What about Rorty’s converse diagnosis of error: the passage from incorrigibility to the lack of an appearance / reality distinction? The basic idea here seems to be that a subject’s privileged access to their own mental state, leads to the public practice of taking the subject’s reports of how the state seems to them as a description of how the state really is. The mistake, according to Rorty, is to draw a metaphysical conclusion from this practice. The false inference is from the subject of a conscious state having a special mode of epistemic access to mental states (they appear to the subject), to conscious states being nothing other than appearances (they are “pure seemings”).

What makes Rorty so sure that the metaphysical status of consciousness is not the source of the linguistic practice? For if conscious states were just appearances, in the metaphysical sense, then that would provide an intuitive explanation (though the wrong one, according to Searle) of why we have privileged access, and why the linguistic practice of attributing incorrigibility should have arisen. Rorty’s argument is compressed, and so is worth clarifying through the following reconstruction:113

(1) Question: What convinces some philosophers that physical and phenomenal terminology cannot just be different ways of talking about the same thing?

(2) Answer: The fact that there is no appearance / reality distinction for conscious properties.

(3) Explanation: The idea of there being no appearance / reality distinction for conscious properties originates from the practice of attributing incorrigibility, which in turn originates from a special epistemic feature of conscious properties: they are known through their appearing to a subject. On the basis of this feature, a reliable linguistic practice arose of equating sincere first person descriptions of how phenomenal properties appear, with descriptions of how the phenomenal properties really are.

(4) If (2) is supposed to provide an answer to (1), then physical and phenomenal properties must be being defined in terms of the appearance / reality distinction. If physical properties are defined as properties which admit of an appearance / reality distinction, and there is no such distinction to be made for phenomenal properties, then physical and phenomenal terminology cannot be different ways of talking about the same thing.

(5) But according to the explanation given in (3), conscious properties only lack an appearance / reality distinction because of how we know them: they appear to us. Why should an epistemological distinction be made to mark an ontological one, in the manner of (4)?

(6) Because whatever appears to the subject is what the conscious property is. Since there is consequently no appearance / reality distinction for phenomenal properties, and phenomenal terminology is appropriate for describing appearances of conscious
properties, then phenomenal terminology is appropriate for describing the essence of conscious properties.

Rorty thinks that the lack of an appearance / reality distinction for conscious properties is adequately accounted for in (3), as the upshot of incorrigibility. Since we each have an immediate way of knowing of our own conscious properties, we have found it useful to treat first person reports as the best guide to what those properties really are. But that does not mean that those properties really are best described by such reports, however sincere and diligent. So (2) does not obviously answer (1); the definitional move made in (4) is needed, and the justification for this move is provided by (6).

What (6) does is run together the epistemological and the metaphysical senses of “appearance”. In the epistemological sense, a conscious property is an appearance in the sense that we know it through its appearing to us. In the metaphysical sense, a conscious property is an appearance in the sense of being the sort of thing that appears, and whose nature is exhausted by its appearing. The epistemological sense does not support the contention that phenomenal terminology picks out the essence of conscious properties. From the fact that we know a property through its appearing to us, it does not follow that we are presented with its ultimate nature. And even if the practice of incorrigibility means the subject’s phenomenal description of the property will be taken as the definitive phenomenal description, this does not rule out a physical description fitting the reality more exactly. So (6) does not warrant (4), and without support from (4), (2) is not a satisfactory answer to (1). In short, that conscious properties lack an appearance / reality distinction is no good reason for thinking that
physical and phenomenal terminology are not simply different ways of talking about
the same thing.

However, there is a better case to be made for (6). For we can acknowledge
the distinction between an epistemological and metaphysical sense of “appearance”,
but defend the view that conscious properties are appearances in the metaphysical
sense, on the basis that this provides an explanation of their epistemological
peculiarities. On this view, the situation is not so much that an epistemological
distinction is being allowed to mark an ontological one without further argument, but
rather that an ontological distinction is put forwards as an explanation of an
epistemological one. Consciousness is being supposed to have a special type of essence
to make sense of its being immediately known, the corollary of this claim being that
nothing physical could be known in such a way. The proposal that this essence is
appearance itself, may be seen as a neutral label under which to group together the
sorts of properties distinctive of consciousness, such as intentionality and there being
something it is like to undergo a conscious state.

This idea is reinforced when we reflect on the degree of affinity between the
epistemological and metaphysical senses of “appearance”. For appearance naturally
contrasts with reality, the contrast being largely constitutive of the sense of these
terms. But though there is a clear distinction between the epistemological and
metaphysical senses of “appearance”, the same cannot be said of “reality”. For
“appearance”, there is appearing as a way of becoming epistemically available, and
there is being an appearance. But for reality, there is only being a reality, or more
grammatically, being a part or constituent of reality. In other words, it looks as if we
only have a metaphysical conception of reality.
If we wanted an epistemological contrast for appearance, so as to generate a truly epistemological appearance / reality distinction, we might contrast appearing, with being known mediately by means of appearance. But then it is all too easy to slide into a metaphysical conception on the basis of this, which is just to say, all too easy to make move (4). We move from appearing and being known by means of appearance, into that which appears, and that which is known by means of appearance. But this is not just an illegitimate move. Rather, it provides a natural answer to the obvious questions that arise once two fundamental ways of knowing have been distinguished, namely: what exactly is it that appears, and what exactly is it that appearing gives us mediate access to? Built into these questions is the requirement that they be answered differently. To see this, suppose that we give as an answer to both questions: something physical. So something physical, such as a tree, appears to us when we perceive it. But then we cannot also say of this tree which appears, that our knowledge of it is mediated by its appearing. If we are tempted to say both that the tree appears and that the tree’s appearing gives us mediate access to the tree, this is only because we have lapsed into treating the tree’s appearing as a happening, process, event, property or thing, which is different from the tree itself.

The epistemology and metaphysics of consciousness are deeply enmeshed conceptually and historically. How we know consciousness has been a key premise in assigning properties to it, and hence in the development of a philosophical conception of its essence. From the fact that it is known by appearing, its essence has been supposed different from reality, which is not so directly known. Descartes could doubt the existence of anything in the physical world, but could not doubt that something was appearing to him, even if the appearance was of the physical world. He then used
conceivability to establish the metaphysical conclusion that appearance was not a part of the physical world. But more powerful than this argument is the metaphysical conception of consciousness implicit in the method of doubt: that there is a common element to appearances of the physical world, and illusory appearances of the physical world. Consciousness then becomes the common element of that which is known immediately. This metaphysical conception purports to explain the epistemology. For if appearance were essentially akin to reality, why should we not also be able to doubt appearance? The answer given is that consciousness is a different sort of thing, whose properties (immediacy, immanence) make it insusceptible to such doubt.

The slide between epistemology and metaphysics infects the whole problem of consciousness. The explanatory gap is, as we have said, an epistemological problem, that of working out how two different sorts of concept relate to the same thing, but it has metaphysical repercussions if appearance cannot be reduced. But this difficulty may have an epistemological source, if how we know consciousness led to our conception of appearance as its essence.

However, there is no clear reason why we ought to avoid allowing epistemology to determine metaphysics to some extent, though this desideratum is often treated as a given. The principle that areas of inquiry should not encroach upon each other in such a way at least requires defending. In the case of epistemology and metaphysics, the procedure looks *prima facie* to be quite reasonable. Suppose for instance, that I am trying to determine the type of object lying behind the door. The only mode of access I have to the object is that I can hear it. From this, the way that I know the object, I can infer all sorts of things about the properties it has. I know that it can produce sound, and hence I can infer that it has the sort of properties which allow
something to produce sound. The same, it might be argued, goes for consciousness, where all or most of what we have to go on in drawing up a metaphysical conception of it, is the way that we become aware of it.\(^{114}\)

In the end then, it seems that both Rorty and Searle are right about what matters, in different ways. Rorty is right that the metaphysical conception of consciousness as appearance is based upon an epistemological premise, our mode of access to consciousness. But this does not necessarily reflect badly on Searle, who can simply reply that our mode of access to consciousness has been an important formative influence upon our conception of what it is. The remaining issue of whether the practice of incorrigibility led to consciousness as appearance, or whether consciousness as appearance led philosophers to mistakenly assume incorrigibility, drops out of consideration, from our perspective. For either way, it is our way of knowing consciousness which led to the notion of consciousness as appearance, either directly or via the practice of incorrigibility. In itself, this does not force us to reject consciousness as appearance.

Rorty’s argument would present a serious challenge to a conception of consciousness as appearance, if it could be shown that privileged access could be explained without recourse to such a conception. This would be to show that a purely epistemic appearance / reality distinction can be sustained, without the need to slide into a metaphysical distinction by substituting appearance as a mode of access, for appearance as a pure seeming. The idea that this can indeed be done is the basis of a currently popular response to the explanatory gap problem: the modes of access

\(^{114}\) Then we could see the explanatory gap as arising from the imposition of the metaphysical constraint that everything be entirely physical, upon a conception built up when our only lead was epistemological.
response. Appearance as just a way of knowing and not something in itself promises to alleviate the problem if it can be shown that the difference between knowing a conscious state and knowing a physical state can simply be the difference between two ways of knowing the same thing. This approach is the subject of Chapter 7, where we will come to broadly negative conclusions about its prospects. But for now, we may just say that the epistemological basis for the conception of appearance as the essence of consciousness will only undermine that conception and support Rorty’s scepticism about a unified notion of the mind, on the condition that a modes of access response to the explanatory gap can be made to work.

4.f. Duality of Mind

Rorty thinks that we should “see the intentional as having no connection with the phenomenal, and the phenomenal as a matter of how we talk.” His case for seeing the phenomenal as a matter of how we talk, we have already discussed. It is the claim that there is no connection between phenomenal and intentional states which now concerns us.

Phenomenal states are strictly incorrigible, on Rorty’s view, in the sense that there is no available procedure for rationally overriding a subject’s contemporaneous report of having a phenomenal state. For him, these are the paradigmatic mental states. Intentional states, those that are non-occurrent and hence non-phenomenal within Rorty’s picture, these states are only “nearly”, or “almost” incorrigible. This near-incorrigibility explains their assimilation to phenomenal states and subsequent

---

classification as mental. Moreover, intentional states have a tendency to cross the boundary from being mere dispositions into being actual episodes. There is, as Rorty notes, little conceptual distance between believing the table is there now, and having the occurrent thought “the table is there now”, or between having a desire for a peach, and having an occurrent thought about a peach. Rorty stops short at using this tendency to collapse into episodic thought as an explanation of the near-incorrigibility of intentional states, though this seems a natural consequence of his views.\textsuperscript{117}

Rorty’s conclusion is that no interesting necessary and sufficient conditions can be formulated for mentality, since only phenomenal states are strictly incorrigible, and intentional states have been assimilated on the basis of a resemblance. But this position is only negative about the unification of mind in attitude, not content. For Rorty might just as easily have said that a state is mental if it is strictly incorrigible, or is the right sort of state to become strictly incorrigible. This recalls Searle’s connection principle (position (2) in our taxonomy). Granted, Rorty’s strict incorrigibility, the practice of “taking people’s word for what they are feeling”,\textsuperscript{118} is not supposed to be an essence of the mental, but rather an explanation of why anyone ever thought there was one. But since he starts out with the intuitive distinction between phenomenal states and intentional states, and ends up with the exactly parallel distinction between strictly and nearly incorrigible states, he can hardly be thought to have exposed more disunity. Whether a social practice or a distinctive essence makes the phenomenal mental is another issue.

\textsuperscript{117} He really ought to have made this move, otherwise the near-incorrigibility is rendered mysterious, there being no incorrigibility surrounding the intentionality of artefacts.  
\textsuperscript{118} Rorty (1979) \textit{Philosophy and the Mirror of Nature}, p. 32.
There is widespread acceptance of the intuitive distinction between phenomenal and intentional states, and agreement that they are exhaustive of the mental. The question which concerns us is whether this shows that our notion of mind is disunified, split between two categories with "no connection", as Rorty says. If so, we would be forced to revise our contention that the explanatory gap debate is driven by a seemingly unavoidable commitment to appearance as the essence of consciousness, since some conscious states would exist that are essentially intentional, but only contingently appearances.

Rorty's own account of the tendency of intentional states to become episodes of thought does seem to belie the "no connection" response, the obvious connection being that intentional states are of the right sort to become conscious. Those who think that conscious states have a distinctive type of essence, as Rorty does not, could argue for Searle's connection principle in order to give unity to the notion of mind. However, our concern is with consciousness, and not with the wider project of finding a unifying principle of mind. Consequently, dispositional beliefs, and non-conscious mental states generally are not of interest to us here. Our interest is in whether we can readily conceive of conscious mental states which are essentially intentional, and only contingently count as appearances.

The idea that the mind, including consciousness, is essentially split between the intentional and the phenomenal, is pervasive even amongst those without such polemical concerns as Rorty. McGinn thinks that sensations and thoughts have different essences. The essence of a sensation is its phenomenal quality, how it feels, and the essence of a thought is a psychological relation to a propositional content.

---

McGinn adds that any phenomenology accruing to the having of a thought, would only be a contingent property of that thought.

A more complex, but essentially congruent stance is adopted by Chalmers, who makes much of the distinction between what he calls the phenomenal concept of mind, and the psychological concept of mind. The former concerns conscious experience, and the latter is to do with causally explaining behaviour. For occurrent intentional states, Chalmers tentatively sides with McGinn; he doubts that conscious feel makes a state into a belief, and doubts that it endows it with content, at least in the case of most beliefs. But many mental concepts lead a “double life”, referring to both a type of experience and a psychological notion. A good example is pain, where the feeling can be conceptually separated from averse reactions in response to bodily damage, as well as less ostensible roles played by such states within our cognitive economies. Such concepts have both a psychological and phenomenal aspect, weighted according to the concept in question, though they can be used to refer to only one of these aspects, as philosophers generally use the concept of pain to refer only to the phenomenal aspect of a pain state.

For McGinn, conscious occurrent thoughts have intentional essences, or for Chalmers, just psychological, functional essences. The phenomenology of such a state would not be essential to its being a thought for either of them, as it would not be for

---

121 Ibid., p. 20.
122 Ibid., p. 16 & ff.
123 Chalmers’s claim that “Both of these aspects are central to the commonsense notion of pain” (op. cit. p. 17) is disputable. Strawson, for instance, argues that “pain” is just a word for a type of appearance. Any connections with behaviour are contingent, and have to do only with learning the meaning of the word. In learning the language of pain, he argues, “it is important that we are observably similar in the way in which we typically react to pain, and again in respect of the things that typically occasion pain in us. But things can be important without being essential.” Strawson (1994) *Mental Reality*, p. 244.
many other philosophers who take the intentional / phenomenal distinction at face value.

We shall argue on the contrary, that the standard conception of occurrent thoughts employed by Rorty, McGinn, and Chalmers actually presupposes that the essence of occurrent thoughts is appearance, but that phenomenology and intentionality are necessary components of appearance. Appearance is the fundamental essence according to this way of conceiving occurrent thoughts, and phenomenology and intentionality are essential properties. This is because appearance is the concept which unifies all conscious phenomena.

There are two lines of argument in defence of this position. The first is that there clearly are cases where intentionality and phenomenology do form part of a unified essence of a mental state: they will not come apart. Given that this is so, then we have at least some cases where appearance in our sense is the essence of a mental state, which makes is more plausible that this essence might extend to all conscious states. Moreover, given that there are such states, then we have a counter-example to the view that the mind is split between two sorts of state: there are at least some states which are both essentially intentional and essentially phenomenal.

The most convincing examples of mental states with essential intentional and phenomenal aspects, come from the emotions. For example, suppose I arrive at a party with a friend who then proceeds to boast and spill his drinks all night. This makes me embarrassed. My mental state could not be the type of mental state that it is, if characteristic types of feeling were absent: cringing, shivers, frissons, and so on. But neither could it be that type of state without its particular intentional content, which involves my friend sitting over there, boasting and spilling his drinks. This can be seen
from the fact that if I suddenly started undergoing the characteristic feelings without
my mental state being directed upon any particular state of affairs, I would not be
embarrassed, but would rather wonder what was going on. I might presume that I was
sick, or cold for instance. And similarly, if I had a state with the same content but no
phenomenology, then that would not be embarrassment either. Somebody impressed
by my friend’s boasting, or just listening politely, could just as well have a state with
that content.

The second line of argument concerns the claim on the part of the defender of
two essences, that for linguistic thoughts, phenomenology does not determine content,
and so cannot be the essence of such thoughts. To an extent, the idea that
phenomenology does not determine content is obviously true, and this must be a key
motivation for the two essence view. Take Chalmers’ example:

when I think that Don Bradman is the greatest cricketer of all time, it seems
plausible that I would have had the same belief even if I had had a very
different conscious experience associated with it. The phenomenology of the
belief is relatively faint, and it is hard to see how it could be this phenomenal
quality that makes the belief a belief about Bradman.\(^\text{124}\)

The point can be made more generally for any sort of intentional state. Perceptions are
not always veridical, and so how a perceptual experience seems to its subject does not
determine what the experience is of. And a basic lesson of externalism is that attention
to detail will not reveal the content of a mental state.

But such points should not blind us to the fact that phenomenology is at least
partially determinate of content. At the very least, how a state seems to its subject
places constraints on what the content could be. If it seems to me that I am thinking

\(^{124}\) Chalmers, op. cit., p. 20.
about the taste of lemon sorbet, then it just cannot be the case that the content of my thought is that Don Bradman is the greatest cricketer of all time. Thoughts about Bradman might be possible with a number of different phenomenologies, but they must all fall within certain parameters. It must seem a little like thinking about Bradman and cricket.

If intentionality were the essence of such a thought, and phenomenology a contingent property, then this constraint would be lacking. One consequence of this constraint being absent, would be that though I think I have been thinking about this topic, I may have been thinking about sorbet. That seems absurd. But if, on the contrary, we hold that phenomenology and intentionality are joint constituents of the essence of conscious states, then it should be possible to make sense of their mutual dependence. For content does apparently determine phenomenology (what it is like to think about Bradman and about sorbet are different) and phenomenology does apparently constrain content (you cannot be thinking about sorbet if it seems you are thinking about Bradman, though it may turn out that you are actually thinking about Taylor or Clarke rather than Bradman).

We have now worked through our taxonomy to see if any of the current methods of moving away from conceiving of consciousness as appearance are viable. They are not. All roads lead back to appearance. When thinking about conscious phenomena, it seems that we always end up leaning on the concept of appearance to some extent, if only to demarcate the terrain to which physicalist reductions are directed. Of course, this does not support the much stronger claim that appearance is the essence of mind generally, or that it can be used to unify mental phenomena with
the addition of a device such as Searle’s connection principle.\textsuperscript{125} But if conscious phenomena are the paradigm of our concept of mind, which they surely are, we can at least see the motivation behind such an agenda.

4.g. Appearance explains other Characteristics of Mind

We shall look at some other features which have often been considered distinctive of the mind, in order to show their essential connection to the concept of appearance. This will reinforce our claim for the centrality of the concept of appearance to philosophical thinking about consciousness.

Nagel’s notion of there being “something it is like” to be in a conscious state, has been the dominant essence of consciousness used within the explanatory gap debate, with inflationists going out of their way to draw attention to it through thought experiments and the like, and deflationists concentrating on either reductively explaining it, or discrediting it. It is also said, in expression of the same idea, that conscious states are subjective, or perspectival. Up until now, we have helped ourselves to these ideas in explication of appearance. We shall now see that they are bound up in the idea of appearance as the essence of consciousness.

The notion of appearance presupposes a subject. An appearance is an appearance to somebody or another. To say that something is a state of appearance (which could be a state of a person rather than a mental entity), is to presuppose the existence of a subject who has something appear to them. This in turn brings with it

\textsuperscript{125} For a discussion sympathetic to the view that appearance is the essence of mind generally, such that “the only distinctively mental phenomena are experiential phenomena” (p. 174), see Strawson (1994) \textit{Mental Reality}, chapter 6. Strawson disagrees with Searle that dispositional and unconscious beliefs can be counted as mental (p. 168), but leaves open the possibility of there being occurrent but nonexperiential mental phenomena (pp. 168-175).
the idea that conscious states are subjective, or perspectival. The idea here is that such states are only available to be known from a certain perspective. Subjective states contrast with objective states, which are such that in principle, anybody can know them from any number of different perspectives, though this is only in principle, because there may be empirical, contingent barriers. Objective states are “just there”, existing independently of any subject’s awareness of them. But subjective states are “there for someone”, and so their existence depends upon a particular subject’s awareness of them.

Is this subjectivity epistemological or metaphysical? Is it that some states can only be known from a certain perspective (they appear), or is it that they exist in a subject-dependent way (they are appearances). This gets us back to our discussion of Section 4.e., about whether a purely epistemic appearance / reality distinction can be maintained, full assessment of which awaits in Chapter 7.

For present purposes, it is enough to note the pulls from both directions. Favouring the epistemological reading, is the fact that the explanatory gap is a debate amongst physicalists, and physicalism does not prima facie look conducive to the thesis that some states are states of appearance. But favouring the metaphysical reading is the fact that appearances only exist as attended to. A state that does not appear to a subject, that is not available to be known from a particular perspective, is not a state of appearance. This suggests that its epistemic availability is wrapped up in what it is. This in turn promulgates the idea that the nature of appearance might explain our mode of access to it.

Reinforcement for this idea comes from the framework of scientific essentialism. For reality can always be other than it appears to be (for example, a stick
that appears bent is really straight), but the same cannot be said of appearance itself, at least at some level (allowing for misinterpretation of appearance). If appearances were ontologically akin to reality, then there would be a part of the physical world which could not but be the way it appeared to be. Yet we are inclined to think that reality can always turn out to differ from how it appears to us. Making appearances realities is tantamount to admitting that real essences can be given in appearance.

Vacillation between these epistemological and metaphysical readings of appearance is reflected in the ideas of subjectivity and perspective. To say of a state that it is subjective or perspectival may mean nothing more than that we have privileged access to it (it is only known from a certain perspective), or it may mean that the state belongs to a "subjective ontology" (having the property, unshared by anything objective, of only existing from a certain perspective). The property of there being "something it is like" is similarly infected. It can just describe epistemic access, there only being something it is like for the person undergoing the state, a property giving that person special access to the state. Or it may mark a metaphysical distinction between states it is like something to be in, and states that it is not like anything to be in.

Why should there be something it is like to be in certain states. Because those states are appearances to their subject, and it is like something to have something appear to you. Why should there not be something it is like to be in some states? Because such states are not instances of an appearance to a subject; they are states of reality which exist just the same whether or not anyone attends to them. There being something it is like to be in a state is an inheritance of the idea that some states are appearances.
The next characteristic often attributed to the mind which we shall consider is incorrigibility. Given that conscious states appear to their subject but not to anybody else, then it is natural to think that the subject’s opinions about their own conscious states would not be liable to be overturned by others. This is because the subject has privileged access to their own conscious states; they are in the exclusive position of being subject to the appearing. However, this is a tendentious essential property for conscious states, because incorrigibility requires that the subject’s opinions be honest. Others have every right to overturn a person’s testimony about their mental states if they think they cannot trust them; it is only honest reports that are candidates for being beyond reproach. But then we wonder why such reports should be treated as incorrigible, since we can never be sure that they are honest. And even the honest reports may be misinterpretations, or believed honest but made in bad faith. The problem with incorrigibility is that it does not so much describe the nature of conscious states, as describe a consequence of their nature. This consequence is the upshot of discourse about appearance rather than reality.

Another closely connected property supposed to be definitive of consciousness is privacy. This idea falls out of the concept of appearance, since appearances appear to a subject, and hence by their nature are not open to public scrutiny in the way of entities whose nature is independent of any mode of access to them. The supposed privacy of conscious phenomena is what inclines us to treat reports of such phenomena as incorrigible.

Non-spatiality is another feature that has often been attributed to the mind. Here the idea seems to be that spatiality is exclusive to physical reality, appearances

\[126\] This once popular attribute has recently been resurrected in McGinn (1995) “Consciousness and Space.”
cannot be equated with reality (they are how we come to know reality), and so appearances cannot be spatial. The idea leans to some extent on introspective access to appearance. For the intentional objects of many appearances are spatial objects, which might lead you to believe that when in introspection you focus on the appearance itself, that cannot also be spatial. Of course, this does not follow, but the compelling thought remains that once you have reasoned to the effect that all of the apparent spatiality of your conscious state belongs to the object of the conscious state, rather than the conscious state itself which you are presently focusing upon, then there is at least nothing about your experience to make you think it is spatial. Add in an epistemological premise, like your ability to doubt the existence of anything spatial, and we arrive at non-spatiality, albeit haphazardly since the conscious states you cannot doubt might still be spatial for all that has been said.

Another feature sometimes attributed to mental states is fine-grainedness, or, as Searle has called it, "aspectual shape". This is also explained by appearance. Fine-grainedness reflects the fact that we always think about things in some way or another, under some aspect to the exclusion of other aspects. This makes mental states fine-grained in the sense that in individuating them, we cannot just mention their intentional object, but must also mention the sense attached to the intentional object. Two thoughts may differ only in that one is about Hesperus, the other about Phosphorus. This fine-grainedness requires reflection in the language used to describe mental states, and subsequently leads to nonextensionality, which was once used as a test for intentionality.

---

Fine-grainedness exists because appearance is perspectival. Appearances are always appearances to a subject, and present reality from the perspective of that subject. Reality exists in abstraction from any particular perspective that is taken upon it, but appearances just are perspectives taken upon reality. To say they are perspectives upon reality rather than reality itself, is just to say that they present reality under some aspect, in one way rather than another. Moreover, the other key feature leading to nonextensionality, namely the possible non-existence of intentional objects, is absolutely central to the idea of appearance. For appearances are neutral to the question of their own veridicality; they may or may not reveal some part of reality to their subject, irrespective of whether or not they purport to do so.

Appearance brings with it all the characteristic properties associated with mind: intentionality, phenomenology, what it is like, subjectivity, perspective, privileged access, incorrigibility, and non-spatiality. Some of these are part of the concept of appearance, some follow from it (or have been thought to follow from it), and some are developments of it. These are very close and inter-related ideas. They all come from the Cartesian framework for understanding the mind, which explains the strong interplay of epistemology and metaphysics. Sometimes metaphysical conceptions arise in response to epistemological facts (like privileged access), sometimes there are both metaphysical and epistemological readings to be had of a concept (like subjectivity).

---

128 Phenomenal properties, or qualia, are a subset of phenomenological properties. Having phenomenal properties is being “like something”, which in turn is just being an appearance. Describing phenomenology is describing what it is like to undergo the state, which is describing the appearance (in a certain way).
4.h. So what is appearance?

Given that this whole thesis is about appearance, it is somewhat embarrassing that we cannot provide anything like a precise answer to this question. But that is just par for the course; one side of the obstinacy of appearance. What we can at least say is that it is supposed to be obvious what appearance is. Searle suggests that those with any doubt about what the subject matter of the problem of consciousness is, need only pinch themselves to be presented with an appearance of pain.\textsuperscript{129} McGinn is content to assume our familiarity with the “technicolour phenomenology” of which he speaks.\textsuperscript{130} Levine, in response to eliminative strategies, feels the urge to reply “‘I literally don’t understand what it means to deny this’ (pointing somewhere vaguely in the direction of my head).”\textsuperscript{131} Examples are enough: the taste of coriander, a feeling of dizziness, etc. For appearance is supposed to be something incredibly close to us, impossible to miss. Our intuitive grasp of the concept of appearance is unassailable, if inarticulate.

The point we have been trying to press home is the apparently anodyne one that this is indeed the subject matter of the explanatory gap problem. It is appearances that all the participants to the debate are talking about, no matter what they go on to say about them. It is hard to see what could possibly be wrong with this. For if you want a theory of anything, you must focus your attention on that thing. A theory of music must be about music, and a theory of consciousness must be about consciousness. And what we apparently mean by “consciousness” is what appears to

us: pains, visual experiences, moods, tastes, and all the other conscious phenomena living, waking creatures can have.

However, it is hard to separate the concept of appearance from the core framework of Cartesian philosophy, in particular from the idea of introspection, and the idea of epistemic givens. The ability to introspect is the ability to separate off appearance from reality; to examine, perceive or just think about mental states themselves, rather than, so to speak, "looking through them" to the world beyond. Some contemporary philosophers are suspicious of the idea of a faculty through which we inspect the contents of our own minds, but it is enough to allow that we can think about our mental states themselves, not what they are about (just what they purport to be about). This is enough, in the sense that it leaves the Cartesian apparatus intact, for what we are concentrating on in introspection is appearance as opposed to reality. This different mode of access, or at least different focus of attention, underlines the distinction between appearance as an epistemic given, and reality as something known via appearance.

"Experience is a fallible guide to reality": this thought motivates the whole idea of an appearance / reality distinction, and enshrines a classic dictation of metaphysics by epistemology. The idea that what we are immediately aware of may or may not be a true guide to reality exerts a powerful influence to this day. This remains true, despite the suspicion cast upon introspection by philosophers such as Dennett, McGinn, and Armstrong. Armstrong argues for the fallibility of introspection, McGinn that it is limited in scope (it may be infallible, but "omniscience does not follow from

---

inerrancy"¹³⁴, and Dennett thinks it is so fallible and limited as to be close to useless. Nevertheless, excepting Dennett who is very unusual in this regard, it remains the case that though philosophers will stomach superficially anti-Cartesian views of introspection, they retain the conviction that something is epistemically given. We might be able to make all sorts of mistakes about our own mental states, but there must be at least some aspect of them that is beyond doubt. And that can only be appearance. Experience is a fallible guide to reality, but could experience be a fallible guide to experience? Yes, if our interpretations of experience are fallible. But that still leaves the interpreted: that which we interpret.

We see an interesting manifestation of this continued commitment to the Cartesian picture juxtaposed against self-consciously anti-Cartesian proposals, in a paper by Lycan.¹³⁵ Lycan sees that his higher-order monitoring theory of consciousness (neatly summarised by Braddon-Mitchell and Jackson as the view that “the essence of consciousness is the ability to reflect on one’s mental states”¹³⁶) has a certain consequence, which is that it “implies an appearance / reality distinction for subjectivity”.¹³⁷ The monitoring of one’s own mental states is a physical process, and hence is not immune to error. So Lycan admits that on his theory, you might introspect a pain even though no such pain exists. In such a case, appearance misleads about reality, and certainly cannot be equated with reality.

Is this really a view on which nothing whatsoever is given, and there is an appearance / reality distinction for everything. Clearly not, just in virtue of how Lycan

¹³⁴ McGinn, op. cit., p. 63.
¹³⁵ Lycan (1997) “Consciousness as Internal Monitoring”.
¹³⁶ Braddon-Mitchell and Jackson (1996) Philosophy of Mind and Cognition, p. 136. We shall construe the higher-order monitoring theory as reducing or replacing appearance with a functional property.
¹³⁷ Lycan, op. cit., p. 756.
sets up the case of registering a “false positive”\textsuperscript{138}: he says that you introspect a pain, though it turns out that there is no pain at all. But clearly you introspect something (this point goes through just the same if introspecting a pain is instantiating a property; there is no need to invoke an act-object account, which Lycan would reject). Lycan attempts to soften up opposition to his introduction of an appearance / reality distinction, by pointing out the functional requirements of real pain that might be lacking, and by pointing out how rare and pathological would be such a false positive. But then he comes around to admitting that there is a sense of “pain” which means just “the introspective awareness (as) of pain”.\textsuperscript{139} Well in that case, this awareness is appearance, and there is a level at which there is no appearance / reality distinction. Lycan’s thesis is not as radical as he thinks; he is just drawing attention to a more extreme way in which we can misdescribe our own mental states, and thus builds on Armstrong’s picture of the fallibility of introspection. He cannot turn his back on the bare bones of the Cartesian framework.

Dennett attempts a more radical withdrawal from the Cartesian terrain. Interestingly, he does not insist on an appearance / reality distinction for conscious states, which is the standard deflationist move. Rather he portrays this as his opponent’s commitment.\textsuperscript{140} This is very odd, since it is conventional for inflationists to premise their arguments on the (essentialist) claim that there is no appearance / reality distinction for conscious states. However, Dennett is not saying that conscious states are appearances (as opposed to realities), but is rather saying that how we judge appearances to be cannot diverge from how they actually are. This is because there are

\textsuperscript{138} Ibid., p. 757.
\textsuperscript{139} Ibid., p. 757.
\textsuperscript{140} Dennett (1991) \textit{Consciousness Explained}, p. 132.
only judgements about how things appear, and no separable appearances upon which those judgements are based. Dennett equates an appearance with a judgement that something has appeared, and so his denial of an appearance / reality distinction is a denial that there is anything over and above sincere judgements that might make those judgements false, or that could occur in the absence of such judgements. He is using "appearance" in the sense of "appears to be the case".

This is a neat subversion of the standard terminology, and gives novel sense to the idea of appearance as the essence of consciousness, the sense being that consciousness is "the one arena" where verificationism "makes manifest good sense". However, does this mean that in giving his account of consciousness, he is no longer talking about appearance? It is hard to see how it could. It is often said that Dennett "denies the existence of phenomenal consciousness". But to accuse a philosophical theory of denying the manifest data towards which the theory is directed is rarely if ever credible. Dennett recognises the data to be accounted for in a theory of consciousness as well as his opponents do, but is trying to give a new account of that data. Of course, Dennett does deny the existence of something for which no metaphysical appearance / reality distinction can be made, but he only thereby denies the existence of consciousness if that is what consciousness is, and since he obviously does not think that is what consciousness is, to accuse him of denying the existence of consciousness is to beg the question against him.

Dennett is attempting to take a radical route out of the Cartesian framework, for he thinks that the retention of "pure seemings" within physicalism is "a hopeless

---

143 The classic example of this sort of gesturing is Dr Johnson's "refutation" of Berkley, by kicking a rock. Obviously, Berkley thought his account could accommodate the solidity of rocks.
metaphysical dodge, a way of trying to have your cake and eat it too." What Dennett rejects is a certain conception of consciousness, according to which "appearances" are something separate from our judgements about what is appearing, and upon which such judgements are based. Nevertheless, he is still talking about appearances, like the taste of coriander and the feeling of pain, because that is how he conceives of consciousness, just as his opponents do. What he is offering is a novel analysis of appearance: a particular appearance of the taste of coriander is in fact something like (to simplify enormously to make the point) a judgement about the taste of coriander attributed to a fictional subject (a centre of narrative gravity). This in turn reduces to a functional state.

So Dennett is also engaged in reducing appearances; it is just that he refuses to take them at face value, as particulars or properties. And he is not at all candid about this: the central argument of *Consciousness Explained* begins with Dennett introspecting a particularly vivid experience involving sound, colour, warmth, pleasure etc., and wondering how "all that" could be a state of his brain, and near to the end of the book, he returns to this same description of his experience in order to show how his account has indeed explained consciousness, i.e. the appearances.

So it may actually be possible to talk about appearance outside of the Cartesian framework. Nevertheless, the concept of appearance remains apparently indispensable in talking about consciousness. No alternative will do the job, as we have seen. But if it

---

144 Dennett, op. cit., p. 134.
145 Dennett, op. cit., p. 26 ff.
146 Op. cit., p. 406 & ff. On p. 410, Dennett suggests answering his earlier question of how "all this" could be a "combination of electrochemical happenings" in his brain, with the rhetorical: "Well, what do you think it would seem like if it were just a combination of electrochemical happenings in your brain?" This makes it clear enough that he sees himself as having undertaken a successful reduction of appearance, or rather, of having cleared away the intuitive opposition to a reduction of appearance.
is just a datum that appearance is that towards which any theory of consciousness is directed, as seems to be the case, then physicalism is in an awkward situation.
Chapter 5: Why Appearance will not Functionalise

5.a. Appearances as Intrinsic Properties

In trying to give topic-neutral translations of expressions used to refer to mental states, physicalists were trying to show that our ways of talking about the mind are not such as to prejudge ontological questions about what the mind actually is. The idea that mental talk is topic-neutral has a certain intuitive appeal, expressed by the thought that you might internally demonstrate “all this”, meaning to refer to your own experience, without having any idea what experience is, just as you might externally demonstrate “all this”, meaning to refer to a novelty you have just encountered in the world, without having any idea of what you have encountered. It certainly makes sense to construe talk about natural phenomena such as lightning as topic-neutral, making “lightning” just stand for something which looks a certain way, and plays a certain role. You do not need to know the nature of that something in order to know what “lightning” means. And there might reasonably be thought no reason to treat discourse about mental phenomena as any the less ontologically neutral. For physical theories and mental theories alike deal with phenomena which were talked about before their inceptions.

Topic-neutral analyses were not just contrived attempts to analyse away anything distinctively mental in mental discourse, motivated by nothing more than

147 The difference between internal and external demonstration, if such notions are legitimate, may just be a difference in intention, as in the case of perception vs. introspection, but with some mental phenomena, like thought, only internal demonstration would be possible.
physicalism, as is often suggested. Rather, they pick up on the need to make
discourses about mind and world as metaphysically innocent as we would expect them
to be, and this is why Smart considered it a merit of his analyses that they would
explain, "why the ancient Greek peasant’s reports about his sensations can be neutral
between dualistic metaphysics or my [Smart’s] materialistic metaphysics." For it is
natural to assume that the ancient Greek could talk about his mind without the benefit
of a theory of what minds are, of the sort later to be devised by the likes of Descartes
and Smart.

That said, Smart’s physicalism certainly gave him a vested interest in topic-
neutral analysis, for unless mental expressions could be shown to be topic-neutral, his
ontology would have been compromised by irreducible mental properties. This is
because if mind-brain identities are discovered \( a \ posteriori \), then it must be possible to
give distinct sets of identifying properties for the mental and physical expressions used
in the identity statements. The mental and physical terms need different meanings, to
explain how it might be a non-trivial discovery that they both refer to the same thing,
just as “Hesperus is Phosphorus” is saved from triviality by the fact that the name
“Hesperus” is associated with the property of appearing in the evening, and
“Phosphorus” with the property of appearing in the morning.

But when we try to elucidate the sense of mental expressions, we find them to
be associated with identifying properties which are alien to the physicalist’s
vocabulary. For instance, we pick out states of pain through their having the property
of feeling a certain way, but we cannot then say that it has been empirically discovered

\(^{149}\) Smart (1959) “Sensations and Brain Processes”, p. 150.
\(^{150}\) Natural, but perhaps anachronistic; see Wallace Matson (1966) “Why is the mind-body problem
not ancient?”
that something physical, a brain, is the bearer of this property, in addition to its more ordinary physical properties of extension, location and chemical composition, with which we might normally associate the word “brain”. Neither do we want to say that just because we pick out a yellowish-orange after-image through its colour, then mind-brain identity entails that brains have been found to possess this sort of phenomenal property. To do so, and say that mind-brain identities hold in virtue of the brain having been found to possess mental properties, just as we might say that the Hesperus-Phosphorus identity holds in virtue of Hesperus having been found to possess the property of appearing in the morning, would be tantamount to accepting the existence of emergent mental properties if mental properties are intrinsic properties. For mental properties are not among the normal structural properties of physical objects; they would have to be the sort that emerge from a certain level of organisation within a physical system. To avoid attributing emergent and irreducible mental properties to physical systems, the physicalist needs it to be the case that in giving the identifying properties associated with mental terms, no departure from normal physical properties is required.

Smart did this by translating mental descriptions into descriptions which only reported the occurrence of a neutrally characterised effect which had resulted from specified environmental circumstances. The famous example is his rendering of having a yellowish-orange after-image as having something go on which is like what goes on when seeing an orange in normal conditions. Armstrong developed topic-neutral translations by making the mind a “mediator between stimulus and response”, the effect of certain stimuli as well as the cause of certain responses. All the subsequent

151 Smart, op. cit., p. 149.
versions of functionalism were variations on this theme. There were two principal

types of approach, one of which was to take ordinary mental concepts, and undertake

conceptual analysis upon them in order to yield relational descriptions which

mentioned only physical intrinsic properties, and the other was to allow the relational
descriptions to take the form of scientific hypotheses about brain functioning, which

were meant to capture the referent of ordinary mental concepts. Either way, the result

was that mental states could be understood as functional states defined by the

relational descriptions, or else identified with the physical properties which satisfied the

relational descriptions.\textsuperscript{153}

Since the only intrinsic properties physicalism can allow are physical properties,

then any phenomenon to be reduced must first be relationally construed, allowing

intrinsic properties of the sort sanctioned by physicalism to meet the relational
description, and thereby be the referent of the mental concept.\textsuperscript{154} But attempts to

functionalise mental states, particularly qualitative mental states, have met with a

standard response: that it is entirely conceivable that the functional description of a

conscious state be satisfied in the absence of consciousness. Block's Chinese Nation

equation, in which interacting people meet the functional characterisation of a brain

state without achieving collective consciousness, is the classic example.\textsuperscript{155} The

intuitions even persist if restrictions are placed upon possible realisers, for many find it

conceivable that a complete physical duplicate of a human being might nevertheless be

unconscious (a zombie).\textsuperscript{156}

\textsuperscript{153} This choice depends on whether the topic-neutral characterisation is taken to be essential to the

type or not. If so, then we have functionalism, if not, then we are just using functionalist premises for

type identity.

\textsuperscript{154} As Kim has noted: "the reducibility of a property critically depends on its \textit{functionalizability}": Kim

(1998) \textit{Mind in a Physical World}, p. 27.

\textsuperscript{155} Block (1980) "Troubles with Functionalism".

\textsuperscript{156} Chalmers (1996) \textit{The Conscious Mind}. 
Now thought experiments may not be able to establish that conscious states are beyond functional characterisation, but they show our lack of understanding of how meeting a functional description might be enough for phenomenal feel. The reason for this is simple. It is that we conceive of conscious states as individuated by intrinsic mental properties. Mental properties are placed on a par with microphysical properties. Microphysical properties are conceived of as the building blocks of reality, and mental properties are conceived of as the building blocks of appearance. And as we shall see, if appearance is anything at all, it has to be an intrinsic property. The simple consequence of this fact is that appearance will not functionalise.

Consider what would happen if we tried to functionalise an undoubted intrinsic property of the sort that is acceptable to the physicalist, a microphysical property. Let us suppose H$_2$O to be such a property, for the sake of this example. To functionalise H$_2$O would be to find a functional construal of it, that mentioned only its external relations to other properties. Then if the functional description were detailed enough, we might be prepared to claim that anything which could meet this functional description in its entirety, would have to actually be H$_2$O. Alternatively, we might weaken the functional description, so that it stopped short of functional characterisation in the microchemical domain, and used only functions pertaining to the role of water in human life. Now we would no longer want to claim that anything meeting the relational description would be H$_2$O. Rather, we would now allow that other properties like XYZ might meet this description, for this coarser grained functional description would allow for variable realisation.

---

157 As Levine argues in his (1993) "On leaving out what it's like".
Compare this to what happens if we try to functionalise appearances. In the case of the weaker functional description, the aim would be to describe an appearance in accordance with the role it plays in our lives, so in the case of appearances of red, for instance, we might look to roles such as being what happens when normal subjects are suitably positioned in front of red objects in good light. Supposing we got as good a characterisation as we could in explicating this folk functional role, then presumably, anything which fitted our functional description would appear to be an appearance. This has two strange consequences that are lacking from the H$_2$O case. First, something might play this role, and so appear to be appearance, and yet not really be an appearance. We might cast our eyes towards red, and be disposed to report a red sensation in the absence of an appearance. And we would not be able to tell the difference between real appearances, and mere appearance-role players. The parallel consequences are all right for H$_2$O, since it is not strange that something might appear to be H$_2$O, but not be. The second consequence is that there could be a case of appearance which failed to exhibit all the characteristic signs of appearance incorporated into its folk role. A few molecules of H$_2$O alone in the universe would also come into this category, but this poses no conceptual issues. Yet an isolated appearance is a dubious notion.

The natural way to respond to these sorts of consideration is to move to the more detailed functional description. It could be thought that the above difficulties just show that appearance could not be an intrinsic property which is reference-fixed upon by a folk role, as H$_2$O is. But still, an appearance might be a theoretical entity, defined by its place in a theory, as H$_2$O also is. So we might proffer a detailed theory of brain functioning, and say that anything meeting a certain complex functional description
would have to be an appearance, just as anything meeting a sufficiently complete theoretical description would have to be H$_2$O. However, this strategy fails for the reason that a sufficiently complete theoretical description would define a physical property. For if the functional description of brain functioning were detailed enough, then the only thing that could fulfil it would be a specific neuronal pattern. If theoretical functional descriptions can be detailed enough to tie down specific physical properties, as they must be to support scientific discourse, then there is no reason why they should not pick out selections from the standard fare of microphysical and microchemical properties when applied to human brains. In short, these descriptions will not pick out appearances, they will pick out neurons.

Of course, it might then be said that in virtue of picking out the neurons, appearances have thereby been picked out. But if functionalising something physical just is functionalising appearance, there must something about those physical properties which have been uniquely characterised by the functional description, that makes them a case of appearance. It cannot be the function, for that just serves to pick out the intrinsic physical property in question. This would put us in the situation of needing an \textit{a priori} conceptual link between conscious and physical properties, of the sort opaque physicalism holds out for. If we had that, then we could with justification say that a theory might be so detailed as to pick out a physical property uniquely, and that since that property was appearance, in virtue of conceptual links between the physical and the phenomenological, then appearance had thereby been functionalised.

The dilemma with trying to functionalise appearance is that if we make the functional description so specific that appearance is treated as a theoretical entity, defined by its place in theory, then the entity we end up picking out will just be a
normal physical one, and if we weaken the functional description so that it only reference-fixes upon appearances, then we allow for the possibility of something playing the role of an appearance without actually being one. Basically, if the functional description tells us what appearance is, then it will tell us it is something physical, and we are none the wiser without an *a priori* conceptual link between physical and phenomenological properties, and if it only tells us how to recognise it, then we allow room for error where there does not seem to be any. There is an obvious incoherence in the idea of something appearing to be an appearance, but not really being one. And what is more, even if we are only told how to recognise appearances, then given physicalism, we must be being told how to recognise something physical, and again we are none the wiser without an *a priori* conceptual link between physical and phenomenal properties.\footnote{It should not be thought that we are imposing epistemic, explanatory constraints upon reduction here. The point is rather that without any conceptual links between physical properties and appearance, then the property which is defined or just picked out by the functional description will be nothing more than a physical property, as it would normally be.}

5.b. Appearances as Functional Properties

The problem of functionalising appearance is partially the problem of functionalising any property of which you have an independent conviction that it is an intrinsic property. In the case of H\textsubscript{2}O, a loose functionalisation will reference fix on the intrinsic property, and a theoretically complete one will define the intrinsic property. But with appearance, since physicalism only allows for physical intrinsic properties, a loose functionalisation will reference fix on a physical intrinsic property, and a theoretically complete one will define a physical intrinsic property. If you start off
thinking that conscious properties are intrinsic, then you have no room to functionalise them in the framework of physicalism. They are bound to get "missed out", which is of course the inflationist's perennial complaint.

The natural way out is to deny that appearance is an intrinsic property that we might functionally construe, and say instead that it is a functional property. This sort of proposal would have us identifying being an appearance with performing a certain function. The intuitive problem with this is that unlike the situation with a functional property like being a carburettor, where we can make conceptual connections between particular realisers and the realised property, in the case of appearance, our concepts of what are supposed to be the realisers (brain states) and what are supposed to be the realised properties (appearances), are largely independent and unconnected.\(^{159}\) Though we could tell a story connecting the intrinsic properties of a particular carburettor with its capacity to mix fuel and air, there are no similar intuitive links between appearances and physical properties. And without a conceptual link between performing functions defined over a neurological domain, and appearing, it seems that all we are really doing is naming a certain functional property, and stipulating that it is the same as a property of appearing. Without a conceptual link, there is no reason to think a functional characterisation of a physical system should be anything other than that: a functional characterisation of a physical system.

The lack of a conceptual link does not afflict this proposal as severely as it did the last, however, for then we had a functional description picking out a physical property which we had no reason to think was also a conscious property, whereas in this case, meeting a functional description is supposed to be what makes something an

\(^{159}\) Except for the constraint that there be a connection, imposed by V#1.
appearance; it is just that we do not know why this should be so. If appearances are intrinsic properties, then physicalism can only be true if it is opaque (if our present conceptions of physical essences fail to capture their essential connection to appearance). If appearances can be functional properties, then we simply lack an explanatory framework proving some principle to explain to us why the instantiation of certain properties necessitate the realisation of appearances. However, there are reasons to think that the contention that appearances are functional properties is not a stable resting place, even if an explanatory gap of this sort is admitted.

For a start, there is something odd about being able to introspect a relational property. It is tempting to say that when we encounter a functionally defined item like a carburettor, we do not see its functional property of being something that mixes fuel and air, but rather only see its intrinsic properties, and associate them with the function, or even perhaps infer the function on the basis of its intrinsic properties. But this is contrived: we do not encounter pieces of wood and metal, and infer their use as hammers. Rather, we always encounter objects under some description, with disinterested contemplation of something as a physical object with certain intrinsic properties being no exception. Yet this phenomenological point does not make sense of the idea that we might directly experience functional properties. Rather, it simply suggests that when we experience things (in virtue of their intrinsic properties), we automatically subsume them under some functional description or another.

The idea that we are aware of things, and not their relations to other things, is hard to shake, and if we are wont to grant ourselves experiential awareness of functional properties in the world, there is at least no obvious reason why we should want to say that we are aware of them in consciousness.
Perhaps it would be better to eschew talk of functional properties altogether, as Kim argues we should, and just talk about functional descriptions, and the intrinsic physical properties which can meet those descriptions. But if it is the intrinsic properties we are aware of, their functional roles just making them count as appearances, then we are back with all the problems enumerated earlier.

Most of the plausibility attaching to the idea of conscious properties as functional properties, as when McGinn likens mental properties to artefact kinds, comes from thinking of being an appearance as nothing more than playing a certain role. But then there is a tendency to slip back into thinking of appearances as intrinsic properties, taking the role as akin to a theoretical definition which nothing could fit unless it were an appearance. However, the only roles that are this specific pick out intrinsic physical properties. The functionalist requires a looser functional role that can pick out any number of different physical intrinsic properties.

If we can make sense of awareness of functional properties, as we can clearly make sense of awareness of appearances, and if we can avoid slipping into thinking of appearances as intrinsic properties tied down by a certain role, there still remains another encumbrance. This is that we have no reason to think that awareness of these particular functional properties (appearances) should not be just like experiencing the functional properties of any other physical system. For if appearing and having a certain function are being said to be the very same property, and we can be aware of functional properties, then the question arises of why instantiating an appearance should differ so radically from other ways of being aware of the same functional property, by looking at a brain through a scanner, for instance.

---

5.c. Appearance and the Framework

With any reduction, we start off thinking of the property to be reduced as intrinsic, but then we relationally construe it in order to effect a reduction. So it turns out that terms which were thought to define intrinsic properties (for instance, lightning, or heat), in fact refer to phenomena characterisable in terms of physical properties (electrical discharge or molecular motion). So it is no surprise that we should think of appearances as intrinsic properties. However, the case of appearance is special, making intrinsicality intuitions particularly resilient. This is because appearances have a special role to play in making sense of scientific essentialism and our model of reduction.

Consciousness is one of the two areas in which essentialist intuitions are most entrenched. When we are concerned with conscious appearances, we feel that we know what the property is without which the appearance could not exist, and this property is not functional. For instance, it seems entirely obvious what the essence of an appearance of pain is, namely the feeling of pain which appears. This just seems to be a matter of what we mean by "pain". Just as with microphysical properties, the other place where essentialist intuitions are at their strongest, we feel that appearances are something in themselves, in abstraction from any external relations that they might enter into.

Scientific essentialism dictates that we think of a property like $\text{H}_2\text{O}$ as something in itself, something with an intrinsic nature. We also think of appearance this way. This conviction is constantly restated, whenever philosophers say that there is no (metaphysical) appearance / reality distinction to be made for conscious properties. To
say this is to say that conscious properties are what they appear to be, which is to say that they are appearances, that appearance is their essence.

The idea of real essences from which the observable features of the world flow, and of appearances whose essence is their appearing, which provide us with epistemic access to the world, complement each other perfectly because they were designed to: they are features of an inherited conceptual framework. The standard pattern of reduction for non-intrinsic properties is to give them a functional construal, and then find an intrinsic property that plays the functional role. But if you have an intrinsic property like H₂O, though you can still functionally construe it, there really is no point. For if you slot anything except H₂O into the functional role, then you just change the property.

Exactly the same is true of appearance, though for different reasons. If you do functionally construe it, you will change the property if you put anything apart from appearance back in, since we have an independent grasp of appearance. Hence Chalmers’s intuition that if you wanted to provide a functional analysis of a green sensation, for instance, you could not just say “the kind of state caused by grass etc.”. Rather you would have to say,

“the kind of phenomenal state that is caused by grass, trees, and so on.” The phenomenal element in the concept prevents an analysis in purely functional terms.

In other words, in virtue of what we mean in talking of phenomenal properties, any functionalisation of them presupposes that the intrinsic property playing the role be an

---

162 The reason for microproperties is the fine-grained theoretical roles which define them. For physicalism, nothing else can be so defined.
appearance, for we already have a grasp on the intrinsic property prior to the process of functionalising it. This is like the case of functionalising physical properties, in which we presuppose that the intrinsic property playing the role is physical (a reality), though scientific investigation would be needed to find out what sort of physical property it is.

We think of conscious properties as intrinsic, because we have special and direct epistemic access to them: they appear to us. We think of physical properties as intrinsic, because they have a special role within realist understandings of the world: they are constitutive of the reality which underlies and causes the appearances we experience. So the independent grasp we have of these intrinsic properties comes through theory and the framework of scientific essentialism on the one hand, and through our way of knowing appearance on the other.

The standard model of reduction requires these two sorts of intrinsic properties; the epistemic bases for drawing up functional descriptions, and the properties that meet those descriptions. Essences of physical phenomena are not given in appearances of physical phenomena, but rather need to be inferred on the basis of those appearances. But if appearances themselves have essences, they cannot also be inferred on the basis of appearances. For beyond the resources appearances provide for discovering the essences of physical phenomena, they do not provide anything extra that might be used to infer their own nature. So the essences of appearances must be manifest. We know the essences of physical things indirectly, via the way they appear to us, but we cannot also know the essences of appearances indirectly, for there is nothing else for us to know them via. So we must know them directly; the properties which we know directly, which are made manifest in appearance, must be what appearances really are.
The idea that we have a direct grasp of the essence of consciousness, which is independent of the deliverances of any theory, is often relied upon within the literature. Take Searle's argument for the irreducibility of appearance, for instance. He says that appearances are the "epistemic bases" through which we know the underlying physical realities that cause them. Reduction involves "carving off" the epistemic bases to get at the underlying realities, says Searle, but this would be pointless for consciousness where it is the epistemic bases themselves which interest us. There would be no point in redefining consciousness as the patterns of neuronal activity which cause appearances (the epistemic bases), because our concept of consciousness was coined for the express purpose of picking out the appearances.

Nobody would argue this way unless they thought they already knew what consciousness essentially is. Once you accept that appearance is one thing and physical reality another, then of course you cannot put appearance to one side to find out what it really is. Searle dismisses the possibility that appearances might turn out to be something physical, because he thinks he already knows what they essentially are. He accepts as much when he says that,

> once the existence of (subjective, qualitative) consciousness is granted (and no sane person can deny its existence, though many pretend to do so), then there is nothing strange, wonderful, or mysterious about its irreducibility.

Granting the existence of consciousness comes down to acknowledging that its essence is appearance. Any physical or functional account of consciousness is a denial

---

165 Ibid., p. 124.
of consciousness in this sense, if the whole idea of appearance is that it is not a physical reality, but our epistemic means of finding out about physical reality.

As we have said, conceptions of the nature of appearance derive from our way of knowing it. Some intrinsic properties are known directly, some indirectly. Without both of these, the framework falls apart. For suppose there were only physical intrinsic properties. Physical intrinsic properties are known indirectly on the basis of the appearances they cause in us. But if appearance itself were such a property, then that would be known indirectly as well. Yet it does not make sense for appearance to be known indirectly, for then there would have to be something else which we knew them via, but there is not. Consequently, since appearances have to be known directly, their essences have to be on the surface. If the essence of appearance could not be known directly, then we would have to find our via appearance. This requirement of something which is known directly is needed to support scientific essentialism and our model of reduction. It rules out the possibility that appearance might be a functional property. The only way out of this impasse is an epistemic appearance / reality distinction; some way in which certain portions of reality might be known in a more direct way than others. We shall examine this proposal, as we have said before, in Chapter 7.

The apparent topic-neutrality of mental discourse was a misconstrual. The sense in which talking of mental states does not prejudge their nature, is only relative to an ontological framework. The framework dictates that whatever the conscious property is, it is an appearance, just as it tells us that whatever a physical property like lightning is, it is a part of physical reality, and so in some sense constituted from physical intrinsic properties. Knowing that a particular conscious property is an
appearance tells us a lot of general things about it: that it is subjective, intentional, available to be introspected but not perceived, revelatory about properties which it cannot be assimilated to, and so on. We also make general prejudgements about properties in the physical world, such as lightning. When we ostensively define such a phenomenon by its folk functional role, we do so against the presupposition that whatever its essence turns out to be, it will be a physical real essence (or the phenomenon will be realised by a physical real essence).

We can ostensively pick out a phenomenon in the world, and in the absence of any particular theory, not know its real nature. And it seems that we do something parallel when we ostensively pick out mental phenomena. But the level of topic-neutrality attaching to physical or mental concepts was misconstrued: when we pick out a phenomenon, it is not the case that we have a completely topic-neutral concept of it, not knowing what sort of thing it is at all. Rather, we are only ignorant of what sort of physical thing it is, or what sort of appearance it is. Moreover, as regards appearance, a second level of misconstrual of topic-neutrality comes to the fore. For since we know such a phenomenon to be the sort of thing that appears, rather than the sort of thing of which there are appearances, there can be no room for discoveries about its essence. As we have just argued, the essence of an appearance must be made manifest to its subject in its very appearing: there is no conceptual room for anything except a direct grasp of essence when it comes to appearances. So despite the apparent analogy between folk conceptions of physical and mental phenomena, it turns out that concepts of appearances cannot be topic-neutral at all. Armstrong was wrong that, "in speaking of mental states we do not in any way judge the question of their intrinsic
nature”, for we presuppose a judgement about their intrinsic nature just in marking them off as mental, and we know their essence so long as we know them at all.

The idea that you only acquire the definitive concept of appearances by having ostensibly identified them, and thus that you can only know what appearances are once you have had them, is a standard inflationist objection to the functionalisation of appearance. For example, Kripke has objected to topic-neutral analyses of pain, that you could learn them without learning what pain is. For that, you need to feel pain. And Jackson’s Knowledge Argument can be read as relying on the idea that only by having the conscious experience of red can you know the essence of that experience.

These points rely on the idea that consciousness is appearance, for if a conscious property like pain is essentially what it appears to be, then the only way to know its essence is to be subject to such an appearance. Obviously you cannot know what it appears to be if it does not appear at all.

By analysing our concepts of conscious states as simply the concepts of nodes within a functional description, the states become “explanation hungry”, as Rorty has put it. Intrinsic properties need to be introduced to explain why the state in question is “apt” to cause this and be caused by that; a categorical basis is needed to explain the disposition in question. But given the direct way we have of knowing appearances, we could never find out anything new about their essences which might explain the roles they play. Physical intrinsic properties do not stand in need of further explanation: they just exist. And given our conception of appearances as intrinsic properties, the same would have to be true of them.

---

So if consciousness is appearance, i.e. what is before my eyes right now, "this", the "technicolour phenomenology", then there can be no physicalistic account of consciousness. And if appearances are only properties of consciousness, then there exist properties of which no physicalistic account can be given.
Chapter 6: Strategies in a No-Win Situation

6.a. The Stand-off between Inflationism and Deflationism

Inflationists usually believe that appearances are intrinsic properties, and hence could not possibly be functionalised. But they have to balance this with their belief in physicalism, according to which there are only physical intrinsic properties. This does not leave much room for manoeuvre, and explains the popularity of the idea that a resolution of the explanatory gap problem could only come about as the result of some sort of scientific paradigm shift, in which we come to understand appearances to actually be physical intrinsic properties, either through changing our conception of matter so that we can see the link between mind and matter \textit{a priori}, or perhaps through discovering new physical properties. Dennett has noticed this tendency in his opponents:

A curious anachronism found in many but not all of these reactionaries [i.e. inflationists] is that to the extent that they hold out any hope at all of a solution to the problem (or problems) of consciousness, they speculate that it will come not from biology or cognitive science, but from—of all things!—physics!

Dennett then proceeds to list the views of Nagel, Chalmers, Noam Chomsky and Galen Strawson, all to the effect that consciousness could turn out to be a new intrinsic

---

\textsuperscript{170} A deflationist could think this too, but it would be an unusual stance for a deflationist physicalist, since functionalisation of consciousness holds with it the only foreseeable hope of explaining consciousness (or holding that it has already been explained). But a deflationist non-physicalist could readily take this stance.

\textsuperscript{171} Dennett (Unpublished) “The Zombie Hunch: Extinction of an Intuition?”.
property recognised in physics, or else constitutive of the intrinsic nature of the physical properties already known to us.

The motivation for this sort of view (opaque physicalism, as we have called it) is that it reconciles the need for conscious properties to be intrinsic, and for the only intrinsic properties to come from physics. Dennett considers such views absurd because he thinks that any physicalistic account of the mind would have to be in terms of the functional organisation of its parts in order to explain the complex performances of the mind which allow us to discriminate, make inferences and judgements, remember things, and so on. Even if brain cells were individually conscious in some sense, this would not explain how they combined to make a mind, as Dennett sees it.

Here we see the classic stand-off between inflationism and deflationism. The inflationist quite rightly insists that consciousness, conceived of as appearance, must be an intrinsic property. The deflationist quite rightly insists that any physicalistic account of consciousness must functionalise it. Both of these are the intuitive, default positions, but inflationism and deflationism can only lay claim to one each.

The inflationist can embrace a conception of consciousness as appearance, but then they are forced to reject any functional account of consciousness out of hand. Yet functionalisation is the normal way of understanding physical phenomena, and in the case of a complex biological process like consciousness, looks overwhelmingly applicable. On the other side of the fence, the deflationist can embrace a functional account of consciousness, but then they are forced to reject appearance as the essence of consciousness. This is easier said than done. Deflationists may be happy to reject this archaic, metaphysical conception, but then they need to justify their claim to still be talking about consciousness. For if consciousness is not "all this", something of
which an ostensive definition might be in the offing, then it is not clear what we mean by "consciousness", or how we could have ever come to talk about it. The deflationist must then rebut the charge of having just provided an account of brain functioning, and having missed out the mind.

Both sides of the debate have their own Achilles' Heel, and the vast majority of the famous inflationist and deflationist papers either draw attention to this weakness in the other side, or try to mitigate this weakness in themselves. The reason is clear: it is that philosophical common-sense is confused about consciousness, and the two sides in the explanatory gap debate have taken up the intuitive strands of the two opposing elements of this common-sense, V#1 and V#2. Inflationism has taken V#2 on board, conceiving of consciousness as "what it is like", and then is forced to reject the apparently obvious fact embodied in V#1: that consciousness is a physical process (which is most evidently true when we look at others). Deflationism has taken V#1 on board, conceiving of consciousness as a physical process like any other, and then is forced to reject the apparently obvious fact embodied in V#2: that what we mean by "consciousness" is "this" (which is most evidently true when we reflect on our own experience).

How this works out in practice is of course a lot more convoluted and complex, but the pattern is certainly there. The most central discussions in the explanatory gap debate themselves demonstrate this.
6.b. Offensive and Defensive Strategies

First, take inflationist offensives. These work by simply underlining the V\#2 conception of consciousness. So there is the topic neutral problem for the identity theory, the idea that we identify mental states through mental properties that are left unreduced by the identity theorist. This points out that it is through appearance that we get a grasp on the mental, and that if topic-neutral analyses of mental concepts were correct, then the mental / physical distinction would be obviated. Then there is Kripke’s conceivability argument about pain. This points out that appearance is essential to consciousness, but not to a physical or functional property, and so they cannot be identified. Then there is Nagel’s idea of “what it is like” to be a bat. This points out that our conception of consciousness is of appearance, and appearances always appear to someone. Hence our access to them is subjective, in contrast to objective, functional and physical accounts.

Moving on, there is Jackson’s Knowledge Argument, which points out that you can only come to know the essence of consciousness through its appearance to you (because it is appearance: its essence is embroiled with our epistemic access to it). So functional accounts that can be learned from books cannot capture this essence. There are also the Chinese Nation and the Chinese Room examples. These point out that our ordinary conception of consciousness as appearance has nothing to do with

---

172 Of my four categories, inflationist offensives and defences, and deflationist offensives and defences, it is under the category of inflationist offensives that the most famous and influential examples in this area are to be found. This is because they threaten to stem the tide of scientific explanation.


175 Nagel (1974) “What is it like to be a bat?”


fulfilling a functional role, so that we can imagine any function being performed in the absence of consciousness (or connectedly, intrinsic intentionality). Then there is Levine’s statement of the explanatory gap. This points out that physical and functional accounts do not provide any conceptual connection to appearance.

All of these famous cases rely to some extent upon conceivability. They point out, emphasise and illustrate, as evocatively as possible, the V#2 conception of consciousness, and then show that this way of conceiving of consciousness has nothing to do with anything physical or functional. This lack of a conceptual link has particular resonance because of the nature of the V#2 conception. For conceiving of consciousness as appearance means that there is a constitutive link between what appearances are and how we know them: appearances just are what they appear to be. So imagining something that appears to be consciousness is imagining consciousness as it really is, and this invests our imaginative capacities with an authority they would not usually have.

Even for those who are suspicious of drawing metaphysical conclusions from our imaginative capacities, as Nagel and Levine are, inflationist offences can still be employed to point out that our conception of consciousness is not of the sort of thing that functional and physical accounts deal with. The conceivability of zombies is simply a distillation of intuitions of the conceptual isolation of appearance from reality, but here conceivability becomes suspect, for though it is easy to imagine functional duplicates being zombies (like the Chinese examples), many inflationists are not so happy with saying that a physical duplicate of a human could be a zombie, for this offends their physicalism. So here we find Nagel, who accepts this line, saying that his

---

denial of physically identical zombies is “a conclusion from the evidence, rather than just a belief in the evidence.” In other words, it is not that we understand the connection between certain physical natures and consciousness, it is just that we have good empirical evidence to expect that there is one. What we mean by consciousness, however, is just appearance.

Searle also rejects physically identical zombies, and also thinks we just mean appearance by “consciousness”. Hence for Searle, all the inflationist examples I have just mentioned, such as those of Nagel, Kripke and Jackson, are “ludicrously simple and quite decisive”, for all they do is remind us of what we mean by consciousness, of what our conception of it is, namely appearance rather than anything physical or functional. Functionalising consciousness is just redefining it, for Searle.

Second, take the inflationist defences. The inflationist defences are more esoteric than their offences, because they are no longer bolstering an aspect of philosophical common-sense (V#2), but rather using their ingenuity to mitigate the counterintuitive aspect of their view (contra V#1). One inflationist defence, already mentioned, is to hold out hope for a conceptual revolution that allows appearances to be thought of as intrinsic physical properties, or at least necessarily connected with intrinsic physical properties. This rebuts the charge of having abandoned physicalistic explanation. It lessens the weight of the charge that their rejection of a functional explanation of consciousness is implausible, by suggesting that they do not thereby reject the V#1 intuition that consciousness is a physical process. It is just that it must be the sort of physical process that, unlike life or digestion, cannot be given a

---

179 Nagel (1991) “What we have in mind when we say we’re thinking”.
182 Ibid., p. 55.
functional account whilst our concepts are in their present state. Rather, a special sort of physical account will have to be given which allows us to see conscious properties as intrinsic physical properties. If there is to be a functional account, it will have to range over physical items conceived of as appearances. There are also inflationists who have no such programmatic defence, and just concentrate on pointing out the shortcomings of deflationism.

Another position is offered by Colin McGinn, with his idea that we are “cognitively closed” to the solution to the explanatory gap problem. This once more allows inflationism to take on board the central tenet of V#1, that consciousness is just an ordinary physical process, whilst still rejecting any functional account of consciousness. For McGinn, consciousness must be an ordinary part of the physical world, but since our access to the physical world is perceptual, and our access to consciousness is through introspection, and since we have no cognitive capacity to straddle the two, then we will never understand how they fit together. So this time we get an even stronger conclusion; not just that present functional accounts fail, but that future ones are bound to as well. Since there is a reason for this, and it does not involve denying what is obvious in V#1, then the counterintuitive consequences of inflationism are bypassed.

A different sort of inflationist defence is to insist that the irreducibility of consciousness is actually compatible with V#1. Consciousness can just be a physical process, even if it cannot be functionalised. Thus we find Searle saying that the “subjective ontology” of appearance is irreducibly different from any objective physical or functional state, whilst nevertheless refusing to deny that it is physical. Rather, he

---

184 Searle op. cit., chapter 4.
says that it is a higher-order physical property of the brain, caused by the brain’s objective physical properties. This is the attempt to accept both V#1 and V#2 by fiat, and parallels a less self-conscious move made by deflationists, to be discussed shortly. A similar move is made by J.J. Valberg, who also thinks that consciousness is caused by the brain, but denies that it is a physical process. This denial does not contravene physicalism, because of his conception of consciousness as “that within which the world appears”. Consciousness is not a phenomenon in the world, but a presupposition of appearance, and, “If consciousness is not part of the world it is not a mysterious part of the world”. So in this case, instead of insisting that V#1 can somehow be true even if we do not understand it, or insisting that V#1 is unproblematically true, the tactic is to reject V#1 as unnecessary for physicalism.

Third, take the deflationist offences. These work by simply underlining the V#1 conception of consciousness. The most fundamental deflationist tools for affirming the conception of consciousness as something physical are, of course, the physicalistic theories of mind: the identity theory, functionalism, and all their variants. These theories automatically put inflationists on the back-foot, because they too want to uphold something like them, but are constrained by their commitment to appearance as the essence of consciousness.

---

185 The element of fiat enters with the insistence that something caused by and irreducible to something physical, can itself be physical. Causation, on a Humean view, relates “distinct existences”. The liquidity of a sample of water is not “distinct” from its molecular structure, but Searle cannot just say that it is realised, because he denies that consciousness can be functionalised, and wants to assimilate consciousness to other physical macroproperties. So physical macroproperties become effects; as if heat were the “effect” of molecular motion, an idea that just serves to introduce pointless empirical questions (see Paul Churchland (1994) “Betty Crocker’s Theory of the Mind: A Review of John Searle’s The Rediscovery of the Mind”). Inflationists do of course want to say appearance is physical, but Searle’s rhetoric about rejecting “Cartesian categories” does nothing to explain how this might be so.

186 Valberg (Unpublished) “Why there could not but be an explanatory gap”.
Then the deflationists can point to the explanatory successes of functional research programmes, which have elucidated features of consciousness, thus reinforcing and to some extent confirming the V#1 conception. The deflationist can appeal to the plausible thought that to bridge the explanatory gap, we will need a detailed and structured picture of the brain, and a detailed and structured picture of phenomenology, so that explanatory rather than brute correlations can be made between the physical and the phenomenological. Scientific progress waits on our ceasing to think of the brain and consciousness as black boxes whose nature is known to intuition. Instead of asserting that any experience is realised by some state of the brain, we should explain why all sorts of different specific experiences are the way they are in terms of their individual physical realisers. The deflationist can then recommend this approach with some results.

To take an example of Van Gulick’s, a principal advocate of this sort of deflationist offensive, neuroscience has provided an explanation of why it is that we cannot experience a colour as both red and green, in the way that we can experience a colour that is red and blue, i.e. purple. The reason is that human brains have two colour channels, one for red / green discrimination and another for yellow / blue discrimination, and binary colours (colours like purple, which combine two colours) are only possible if each of the component colours of the binary colour belong to different colour channels. Red and green are at opposite ends of the same channel, so they cannot combine to make a binary colour. Now the inflationist might dismiss, reinterpret or just downplay this sort of thing, but given that we are already inclined to think of consciousness as a physical process (as reflected in V#1), then such quibbles

---

will have little weight. The sort of parallels exampled above are readily taken as explanations of consciousness by neuroscience, and the deflationist claim that the gap will close when we have more of the same will have been made considerably more attractive.

Deflationists also build on V#1 by alluding to scientifically discovered functions, which serves to cast doubt on the inflationist’s imaginings of functions in the absence of consciousness: inflationists are just not imagining the prerequisite functions in all their complexity, perhaps because they have not even been discovered yet. A powerful deflationist tactic is to make analogies with computers. Here we have physical objects that are remarkably mind-like when we analyse their performances at the software level, but when we restrict our attention to the hardware, we find only physical processes of high complexity. In the case of computers, the mind-like behaviour is nothing more than the physical processes, and we are invited to expect consciousness to be something similar: a program running on the hardware of the brain.

More generally still, deflationists can just appeal to our belief in the scientific method, and the physicalism which this engenders. Whilst inflationists seem to be trying to hold up progress, the deflationists take the much the more appealing line that we should just get on and explain consciousness, as we have explained other once mysterious natural phenomena, like magnetism. Here we find the intuitive backbone of deflationism: whilst inflationist offensive strategies rest on intuitions about how we conceive of consciousness, deflationist offensive strategies rest on faith in the explanatory power of science: of how we ought to conceive of consciousness, given the scientific world-view.
Fourth, take the deflationist defences. The deflationist defences are more esoteric than their offences, because they are no longer bolstering an aspect of philosophical common-sense (V#1), but rather using their ingenuity to mitigate the counterintuitive aspect of their view (contra V#2). The main aim here is to cast aspersions on the conception of consciousness as appearance which creates difficulties for the sort of functional or physical account offered by the deflationist. One of the key moves in this direction was the physicalist’s insistence on talking about mental states or mental events, making it clear that appearances could not be thought of as mental objects that we ostensively identify, akin to sense data. This was supposed to loosen up essentialist intuitions, for once we stop thinking of appearances like pains as objects, it becomes less obvious what their essential properties are supposed to be. Lycan has even argued that events have no essences, or at least none that we would have strong intuitions about.\footnote{Lycan (1987) \textit{Consciousness}, p. 17. According to Lycan, essentialism about the mind is so entrenched because of a natural tendency to slip on what he calls "the Banana Peel", by thinking of states and events as objects. He may be right that we lack essentialist intuitions about states and events in the world (the sort of example he presents), but when it comes to the mind, we do have strong intuitions about the essences of states, and events. Consider the event of having a shooting pain, or of being in a pain state. We have these intuitions simply because we think of consciousness itself as appearance, and so unlike events and states in the world, we assume the essence is manifest in the state, ready to read off.}

Of course, this still left the properties by which mental states or events are identified (the topic neutral problem). Against this, deflationism has to either insist that such properties are physical or functional properties of the states or events, or deny that there are any such properties. The proposal that they can be identified with physical or functional properties faces resistance on the basis of intuitions that appearances are intrinsic properties, and so deflationists put forward theories such as adverbialism, and its successor strong intentionalism, which try to give an explanation
of how introspectable properties might actually be properties of broadly physical states of perceiving or representing.

Another tactic is to do away with properties of appearance altogether, as Dennett attempted to do in his “Quining Qualia”,\textsuperscript{189} by showing the incoherence of philosophical conceptions of qualia. There are no qualia, just judgements that there are. Or else the deflationist may want to go even further, and eliminate consciousness altogether, arguing that the whole folk conception of mind embodies a false theory, which is incapable of smooth reduction to the neurophysiology or cognitive science we have today.\textsuperscript{190}

In general, deflationist defences look to discredit intuitions based on V#2, which say that we know exactly what we mean by “consciousness”, prior to any philosophical theory. The deflationist instead wants to say that we only know the sort of thing consciousness must be, as set out in V#1. So they attack over-reliance on scientifically uninformed intuition, and dismiss inflationist objections as armchair speculation and dogmatism. They attack arguments based on conceivability, which are important to the inflationist who thinks that the epistemology and metaphysics of consciousness are importantly related. And they compare the objections of inflationists to long outdated debates, in which there were reactionaries who clung to some old conception of a phenomenon, but who were eventually discredited in the face of overwhelming new evidence which mounted up in favour of a successor concept. The choice between inflationism and deflationism is portrayed as the choice between a folk conception and a scientifically respectable one.

\textsuperscript{189} Dennett (1988) “Quining Qualia”.
\textsuperscript{190} Paul Churchland (1981) “Eliminative Materialism and the Propositional Attitudes”.
Another standard tactic is to attempt to dismantle the inflationist’s apparatus, from which they derive the notion of consciousness as appearance. So deflationists question the idea of introspection, and they argue for an appearance/reality distinction in the realm of consciousness, as we discussed in Section 4.c. And some, such as Rorty, will even reject essentialism altogether, a move we discussed in Section 2.e.

Many deflationists do not undertake such programmatic defences, and are content to just answer the problems posed by the inflationists. So they will say that with enough physical information, we could know what it is like to be a bat, or that there is nothing it is like to be a bat, or that of course knowing all about a bat will not allow you to actually be a bat, for only a bat can know what it is like to be a bat, but being a bat is simply a matter of spatio-temporal location. They will say that Mary might know what it is like to see red if she knew everything physical. And so on.

But when deflationists do not take on a type of programmatic defence, then by default, they are in the position of trying to combine V#1 and V#2 by fiat. This is the less self-conscious version of a parallel inflationist proposal which I mentioned earlier, in connection with Searle’s suggestion that consciousness could be both appearance and a physical process, if we dropped what he calls “Cartesian categories”. But what deflationists do when they simply rebut the inflationist arguments, is assume that we all have a grasp on what consciousness is, and purport to be giving their functional or physical account of consciousness in that sense. In other words, if they do not say otherwise, it is natural to assume that they are thinking of consciousness as appearance, and insisting that their account can reduce appearances.
6.c. Conclusion

We have now taken a look at inflationist offences and defences, and deflationist offences and defences. The pattern of argumentation we have found is exactly what would be expected, given the motivations for inflationism and deflationism with which we introduced the discussion. To briefly recap, the inflationist is motivated by their adherence to appearance as the essence of consciousness, the key aspect (or conclusion) from V#2. But this means they have serious difficulties with the idea of V#1, that consciousness is a physical process, even though they are inclined to accept this. The deflationist is motivated by their adherence to consciousness as a physical process, the key idea of V#1. But this means they have serious difficulties with the idea of V#2, that appearance is the essence of consciousness, even though they are inclined to think of consciousness this way, and alternatives are not readily apparent.

The sort of arguments relied on by the two sides bear out our characterisation of the dispute. Moreover, they also demonstrate the strengths and weaknesses of inflationism and deflationism. For where they are strong, they are relying on an element of one of the views. And where they are weak, they are arguing against one of the views. Inflationism is at its most convincing when it is drawing our attention to V#2, and least convincing when it renders V#1 problematic. Deflationism is at its most convincing when it is defending V#1, and least convincing when it renders V#2 problematic.

So when we look at inflationism on the offensive, we find that Nagel, Jackson, and Levine’s examples are much discussed, and widely considered to make a very persuasive case by those without a vested interest in deflationism. And even some
deflationists will want to grant an inaccessible inner life to the bat, or say that Mary learns something. And then on the other side of the debate, when we look at deflationism on the offensive, we find widespread adherence to some sort of physical view of the mind, and optimism that the explanatory gap will prove scientifically assailable along lines of investigation already underway. We find a wealth of experimental results from the likes of Dennett, which seem to illuminate consciousness to us, alongside a body of enlightening similes and competing philosophical theories with which to aid our understanding of the mind.

But then look at the inflationist defences. The inflationists have to pass over all ordinary means of scientific investigation, because that cannot allow consciousness to be functionalised, and yet they do not want to deny that it just is a physical process, as V#1 says it is. So they are forced into an *ad hoc* appeal to a conceptual revolution; the hope that something unforeseen will occur to reconcile the inflationist’s commitments to both appearance and physicalism. But is this expectation a rational conclusion from the evidence? An alternative would be to keep the problem intra-philosophical, and reconsider the inflationist commitment to appearance, or appearance’s resistance to functionalisation.

McGinn’s contention that the problem is beyond us, similar to a position tentatively endorsed by Jackson at one time, is less defensible, and no less *ad hoc* in its motivation. To mention just one problem with it, if the explanatory gap is supposed to be beyond our intellectual powers for some principled reason, then it is not clear why that reason should not also prevent us from grasping the problem. In the sort of examples McGinn relies on, such as armadillos being incapable of doing arithmetic, it is

---

fairly obvious that there is no grasp of or engagement with the problem in the first place.\textsuperscript{192}

Similarly, when deflationism is on the defensive, it starts saying implausible things. The problem here is that if consciousness is not appearance, then it is unclear why we ever started talking about consciousness in the first place. Yet if consciousness is appearance, then it cannot be reduced. So the deflationist is forced to deny that consciousness is appearance, usually doing so implicitly by picking away at the inflationist's resources, without putting forward a viable alternative conception. Otherwise they make manifestly implausible denials of the existence of appearance, thus falling into the hands of their inflationist opponents who will portray them as subject to a kind of bad faith in denying something demonstrably undeniable.

In short, neither inflationism or deflationism can make any progress, because they have not tried to resolve the two views. Both factions defend either V\#1 or V\#2, and then face an uphill struggle when they argue against the other one, or render it problematic.

One philosopher who seems to have grasped this to some extent, is Chalmers, for he has adopted a position which hives off the best aspects of both inflationism and deflationism: those which are conducive to either V\#1 or V\#2. Thus through an original manoeuvre, and one that is quite telling so far as our present analysis of the debate goes, Chalmers tries to combine the best of both worlds by splitting the problem up into the "hard" and the "easy" problems.\textsuperscript{193} The hard problem is to account for appearance. The easy problem is to work out how the functional organisation of

\textsuperscript{192} For a full assessment and critique of McGinn, see Owen Flanagan (1992) \textit{Consciousness Reconsidered}, chapter 6.

the parts of a brain explain the complex performances of the mind, which allow us to process environmental stimulation, integrate information, discriminate, make inferences and judgements, remember things, and so on. By dividing the problem in two like this, Chalmers can accept the deflationist’s claim that any plausible account of the mind would have to be functional. “Of course”, he can say, “understanding consciousness as a physical process is like understanding any other physical process”, and so the deflationist approach to the easy problems is the right one, as we are already inclined to believe. But then, he can also accept the inflationist’s conception of consciousness as appearance, and their plausible thought experiments which draw attention to this conception. “Of course”, he can say, “this is what we ordinarily mean by ‘consciousness’, and it is something that we could never provide a functional account of.”

Chalmers’s position inherits the intuitive momentum behind both inflationism and deflationism. Moreover, it avoids the demerits of deflationism, because he does not have to argue against consciousness as appearance. It also avoids some of the demerits of inflationism, for it does not need to deny that ordinary investigation can help to explain consciousness. But because Chalmers does not face up to the problem of resolving the dispute between V#1 and V#2, all the old problems resurface. In effect, all he does is grant the deflationists their own conception of consciousness, agrees with them on that head, and then reaffirms inflationism at another level. For an explanation of how appearance fits into the physical world is still required, and so he needs to appeal to a conceptual revolution just as the other inflationists do. In fact,

194 On ibid., p. 126, where he speculates that fundamental physics will one day incorporate consciousness. In this regard, Strawson holds a similar view: see his (1994) *Mental Reality*, pp. 104-5.
the situation is rather worse for Chalmers, for in his reluctance to reject the plausible face of deflationism, he ends up allowing their functional accounts to use up all the causal efficacy of consciousness: its information processing, control of action, etc. So appearance ends up epiphenomenal.

We have reached a point of conclusion. The explanatory gap debate is deadlocked because all the participants either base their position on V#1 or V#2, instead of trying to resolve the conflict between the two. Inflationists and deflationists work with different conceptions of consciousness, and so end up talking at cross purposes,195 instead of taking both conceptions on board and trying to work out how they might both relate to the same phenomenon. Inflationists think deflationists “miss out the mind”, and deflationists think inflationists are working with an obsolete, scientifically impenetrably conception which should, on that very basis, be abandoned. No progress will be made because both conceptions represent either V#1 or V#2, and so whenever either side tries to attack one of those conceptions, they will sound implausible. We want to accept both conceptions, but do not know how to. That is the problem. Attacking one from the basis of the other is not the way forward.

Of course, we could accept either inflationism or deflationism, but it would not do any good. If we accept inflationism, we are still left wondering how consciousness could be a physical process. This now becomes a mystery that we cannot solve, and inflationists often do not even pretend to solve it; they just point to the problem and hope for a breakthrough sometime in the future. But we still have our ordinary, conflicting conceptions of consciousness, and so are not at all satisfied. And besides, as

---

195 We saw a classic example of this in a debate between Searle and Dennett, in chapter 1.
we said at the outset, there are good reasons to think that we are dealing here with a philosophical problem, born of philosophical background assumptions.

So what about just accepting deflationism? Well, this will not do us any good either, for we still have our conception of consciousness as appearance, and want to know how such a thing fits into the physical world, or how we formed this sort of conception of something physical. On these, the important questions, deflationism is quiet. The deflationists just talk about physical processes and functions, and assume that this is the same thing as appearance. Or else, they deny that appearance exists. The former tactic fails to address the problem, and the latter is an uphill struggle, with the odds firmly stacked against it because of our continued adherence to V#2, which gave rise to the problem in the first place.

In short, inflationism puts the problem of consciousness on a pedestal, and deflationism brushes it under the carpet. Inflationists tend to accept V#1 and V#2 at face value, and thus end up with an explanatory gap. Deflationists tend to accept V#1 and reject V#2, thus side-stepping the explanatory gap, but leaving us with incompatible conceptions of consciousness which we do not know how to turn our backs on: which is what puzzled us in the first place. Dualists do something similar, also side-stepping the explanatory gap, by accepting V#2 and rejecting V#1. But while dualists might not have been moved by V#1, deflationists are still exercised by V#2. Officially, there is no explanatory gap on their view. But they still go on to explain why there seems to be one. And this is uncomfortably close to an admission that it remains natural to conceive of consciousness as appearance, and the real difficulty has simply been ruled out of court. There is only one way forward on the problem, and that is to
resolve the conflict between V#1 and V#2. The ideas they represent are not going to
go away: they are deeply entrenched ways of thinking about consciousness.

However, we have still not shown that the deadlock cannot be broken in favour
of deflationism, since there is a deflationist proposal which we have yet to consider,
but have mentioned in a number of places as a promising approach to doing just that.
This is what we shall call the Modes of Access Response, and in terms of our analysis,
we can see it as the attempt to sustain a purely epistemic appearance / reality
distinction, that is, to show that to be an appearance is just to be a normal physical
property that is known in a particularly intimate way. This will be the subject of our
next chapter.
Chapter 7: The Modes of Access Response

7.a. Introduction

The Modes of Access Response (MAR)\(^{196}\) is first and foremost a response to the explanatory gap problem, though its scope might even be thought wide enough to describe it as a new form of physicalism.\(^{197}\) The core idea of MAR is that the apparently vast difference between conscious and physical properties can be entirely accounted for in terms of our possession of different modes of epistemic access to those properties, from which we form different sorts of concept. MAR holds that it is possible to account for this conceptual disparity without invoking either intrinsic mental properties associated with mental concepts, as dualism does, or functional properties associated with mental concepts, as type physicalism and functionalism do.

The starting point for MAR is that we each have a mode of epistemic access to our own conscious mind that is distinct from the observational modes we employ to

---

\(^{196}\) MAR is defended by Michael Tye, Brian Loar, William Lycan, Christopher Hill, Brian McLaughlin, David Papineau, Ned Block, Robert Stalnaker, and Scott Sturgeon. I took the label from the following summary of the view by Tye: "In the world itself, there is no gap, no yawning chasm, between two radically different sorts of things. There is just good, old physical stuff that we conceptualize in different ways, depending on our mode of access to it." Tye (1995) *Ten Problems of Consciousness*, pp. 181-2. It seemed appropriate because the distinction between different epistemic modes of access is the key idea holding the various elements of the position together, and yet "mode of access" is such a neutral term that it does not prejudice the formulation of the position. Levine's label, "the modes of presentation strategy" (Levine (1997) "Recent Work on Consciousness") lacks this last virtue, for some MAR theorists place great emphasis on making an analogy between how phenomenal concepts refer, and how demonstratives and / or proper names refer, and talking about modes of presentation suggests a Fregean position which is inessential to the view, and might be rejected by some of its advocates.

\(^{197}\) It is offhandishly described this way, in Hill and McLaughlin (1999) "There are fewer things in reality than are dreamt of in Chalmers's philosophy", p. 451. MAR could be called a new form of physicalism on the grounds that it offers a new take on the way in which mental concepts refer to parts of the physical world. This is arguably what makes the identity theory, functionalism, and anomalous monism all count as forms of physicalism.
access other parts of the world. Corresponding to these different modes of access, we form different types of concept, such as phenomenal concepts for states of our own minds and physical concepts for observed properties. If a conscious property were identical to a property of the brain, there would be two modes of access to that property, resulting in our possession of two sorts of concept for the same thing. According to MAR, our use of these disparate concepts explains why there seems to be a vast difference between consciousness and anything physical, such as brain activity.

In and of itself, there is little new in this; the main idea is just that the gap is conceptual, without metaphysical consequence. But the innovation of MAR is to call a principled halt to the explanation at this point. The basis of this injunction is that to expect an explanation of physicalism which makes conceptual connections between the phenomenal and physical is to misunderstand the way phenomenal concepts refer.

For Levine, the conceptual gap is a problem. His model for solving it is the sort of explanation of macroproperties in terms of microproperties, from which the molecular structure of water can provide insight into its observable behaviour. By producing a D-N explanation which shows microproperties to necessitate the realisation of the macroproperties associated \textit{a priori} with our concept \textit{water}, conceptual connections are made that render the reduction of water to H$_2$O an explanatory reduction. But according to Levine, our inability to connect our concepts of conscious properties and physical properties in this way shows that physicalism about the mind contains an explanatory gap.

MAR denies the applicability of this model to the mental / physical case, and thereby deflates the significance of the conceptual gap. To understand this denial, we
need to make a distinction between two ways a concept can refer, which is crucial to MAR's position, and which will be employed throughout our discussion (as it is in all the literature on MAR). The ideas behind this distinction are familiar from standard models of mind-body reduction, and are employed in Kripke’s argument against the identity theory. MAR’s use of these ideas is distinctive in focusing on concepts.

The distinction is made variously by different MAR theorists, but in the terminology we shall adopt, it is between concepts that refer by connoting functional descriptions which are satisfied by their referent, and concepts that essentially conceive their referents. Concepts of the former type are associated \textit{a priori} with functional descriptions which fix their reference. Concepts of the latter type are not. Rather, they are said to refer "directly", in a way that is "unmediated". It is the central claim of MAR that phenomenal concepts refer in this direct fashion, and not through associated descriptions.

Ordinary natural kind concepts like \textit{water} belong to the former category of concepts. The concept \textit{water} refers to water itself, i.e. H\textsubscript{2}O, but it does not do so "directly". This is because the concept of water is associated \textit{a priori} with functional

\begin{footnotesize}
\begin{enumerate}
\item Loar (1997) "Phenomenal States", p. 600: "Phenomenal concepts and theoretical expressions of physical properties both conceive their references essentially"; ibid., p. 602: "recognition concepts other than phenomenal concepts connote contingent modes of presentation that are metaphysically independent of the natural kinds they pick out".
\item Loar, op. cit., p. 600: "a phenomenal concept can pick out a physical property directly or essentially"; Tye (1999) "Phenomenal Consciousness: The Explanatory Gap as a Cognitive Illusion", p. 713: "[phenomenal concepts] enable us to discriminate phenomenal qualities and states directly on the basis of introspection."
\item Papineau (1998) "Mind the Gap", p. 381: "there are no descriptions associated \textit{a priori} with our everyday concepts of conscious states."; Tye, op. cit., p. 713: "In having the phenomenal concept PAIN, for example, I have a simple way of classifying pain that enables me to recognize it via introspection without the use of any descriptive, reference-fixing intermediaries."
\item Hence directness is not to be equated with rigid designation; the paradigm of "indirectly" referring concepts, natural kind terms, are rigid designators. (MAR is committed to phenomenal concepts rigidly designating their referents, however (see Loar op. cit., p. 603)). Neither do MAR theorists mean "direct reference", as this idea is used in the philosophy of language (see Salmon & Soames (eds.) (1988) \textit{Propositions and Attitudes}, introduction, p. 4), since MAR’s concern is with concepts.
\end{enumerate}
\end{footnotesize}
descriptions which fix the reference of the concept onto whatever it is that satisfies the
description. In the case of *water*, the functional descriptions are being “the stuff that
falls from the skies, fills the oceans and rivers, etc.,” in short, being “the watery stuff”,
as Chalmers puts it.\(^{202}\) Since water itself, i.e. \(H_2O\), is “the watery stuff”, the stuff that
satisfies the descriptions associated with the concept *water*, then *water* refers to \(H_2O\).
So the concept *water* denotes \(H_2O\) by connoting functional descriptions which are
satisfied by \(H_2O\).

Essentially referring concepts do not refer through association with reference-
fixing properties. Rather, they refer directly, by picking out the property which the
concept itself “expresses”.\(^{203}\) Phenomenal concepts express the same properties that
they refer to, where the property expressed by a concept is the property which fixes its
reference. In other words, the property which fixes the reference of such a concept is
the referent itself.\(^{204}\)

Loar describes essentially referring concepts as providing a direct grasp of
essence, by conceiving properties “as they are in themselves”.\(^{205}\) The standard example
is of a theoretical microphysical concept like \(H_2O\). The idea is that \(H_2O\) is conceptually
equivalent to a theoretical, functional description. Rather than this functional
description fixing the reference of the concept, as with an everyday concept like *water*,


the very properties they pick out”; ibid., p. 467: “phenomenal concepts *express* (and refer to)
properties that are identical with properties expressed by certain physical concepts.”

\(^{204}\) Sturgeon (2000) *Matters of Mind*, p. 48: “When introspecting our mental states, we do not take
canonical evidence [for applying a phenomenal concept] to be an intermediary between properties
introspected and our conception of them. We take evidence to be properties introspected.”; Hill and
McLaughlin (1999) “There are fewer things in reality than are dreamt of in Chalmers’s philosophy”,
p. 448: “When one uses a sensory concept to classify one’s own current experiences, the experiences
that guide and justify one in applying the concept are always identical with the experiences to which
the concept is applied.”

\(^{205}\) Loar (1997) “Phenomenal States”, p. 608: “It is natural to regard our conceptions of phenomenal
qualities as conceiving them as they are in themselves, that is, to suppose we have a direct grasp of
their essence.”
in this case the concept gives or expresses the referent itself, since the property is
defined as whatever satisfies the fine-grained functional description. So grasping the
concept $H_2O$ is grasping an essence.\(^{206}\)

Levine’s model of explanatory reduction works only for pairs of co-referring
concepts where one concept refers essentially, and the other is a recognitional concept
that refers by connoting functional descriptions of the referent. The tacit assumption of
the explanatory gap problem is that a reduction of mind would fit this mould, with
physical concepts essentially conceiving their referents, and mental concepts referring
by connoting functional descriptions, as do ordinary recognitional concepts like water.
If this were so, then physicalism would incur the explanatory obligation of saying what
it is about certain physical or functional states which makes the functional descriptions
associated with mental concepts applicable to them. But MAR rejects this tacit
assumption, and holds that mental and physical concepts both essentially conceive their
referents.

So MAR claims that we have two different types of concept that both
essentially conceive the same referent, corresponding to our different modes of access
to that referent. This is enough to address the explanatory gap problem, since given the
way these concepts refer, there is no possibility of devising an explanatory framework
that connects them. The conceptual distance between phenomenal and physical
concepts is just a consequence of our psychological make-ups, and the intuition that

\(^{206}\) Not all MAR theorists hold that microphysical concepts essentially conceive their referents. Tye
that “the concept HYDROGEN refers via a description, for example, “Whatever passes the chemical
tests for the stuff called ‘hydrogen’” (or something to that effect)”. However, so long as phenomenal
concepts do not refer via associated descriptions, the upshot of the view is the same.
attributing a physical nature to consciousness presents us with an enigma that we are as yet unable to unravel, is revealed by MAR as a “cognitive illusion”.  

Our procedure in this chapter will be the following. First, in Section 7.b., we will look at the significance of MAR from three angles: as an original response to the problem of conceptual difference, as a form of scepticism about modal intuitions, and as an acceptable way for physicalists to concur with the intuition that there is no appearance / reality distinction for consciousness. In Section 7.c., we will consider MAR’s take on the import of conceivability, by contrasting it with two other influential views. Then in Section 7.d., our critique of MAR will begin. First, we will show that MAR theorists have provided no positive reasons for thinking that phenomenal concepts do in fact essentially conceive their referents. At most, they have uncovered a possible alternative to a problematic explanatory gap. In Section 7.e., we see that the connection MAR makes between phenomenal concepts and physical or functional properties is problematically strong, and that consequently the alternative MAR purports to offer is not an attractive one. Finally in Section 7.f., we find that despite initial impressions of MAR as a highly deflationary and tough-minded physicalist approach, MAR actually bears close analogy with Nagel’s and McGinn’s views on consciousness.

207 Tye (1999) “Phenomenal Consciousness: The Explanatory Gap as a Cognitive Illusion”. MAR theorists generally, if they do not use the word “illusion” (they usually do), use words to the same effect. Papineau, for instance, says: “you might feel the need for something more, even after you embrace the identity of pain with some material property, if you mistake the phenomenal concept of pain for a concept that refers in virtue of its association with descriptions.” Papineau (1998) “Mind the Gap”, p. 383.
7.b. The Significance of MAR

MAR offers a new way of responding to the problem of conceptual difference, a problem which early physicalists like Smart pioneered a standard response to. This problem is the subject of our next chapter, Chapter 8, by which time we will be better placed to fully understand it, though we can at least broach it here. The problem is basically that since mental concepts cannot be equated with physical concepts, then even if they refer to a physical state, they must still introduce a mental property (by connoting a description that could only be satisfied by something mental). For suppose a mental concept M and a physical concept P both refer to a state S. If, as a physicalist might contend, P essentially conceives S, then since M is a different concept to P, and no one thing can have two essences (as we saw in Section 2.c.), then M cannot also essentially conceive the essence of S. So if M does indeed refer to S, it must do so in virtue of S satisfying functional descriptions connoted by M. Opponents of physicalism assumed that the properties introduced by mental concepts would be intrinsically mental, since these connoted properties would be the properties essentially conceived by mental concepts. Physicalists like Smart countered that they were in fact functional or quasi-functional properties.

Instead of responding to the problem by arguing that mental concepts connote functional descriptions, MAR denies that mental concepts connote any descriptions at

---

208 Smart (1959) “Sensations and Brain Processes”, deals with this problem. Smart’s Objection 3 addresses the problem as Max Black apparently formulated it (Lycan (1987) Consciousness, chapter 2, calls it “Black’s Objection”), and Objection 5 deals with the problem as C.D. Broad formulated it (see Broad (1925) Mind and its Place in Nature, chapter XIV).

209 If M and P essentially conceived essential properties of one unified essence, then the presumptive consequence (excepting the ambitions of MAR) would be property dualism, or some form of opaque physicalism.

210 “Quasi-functional” because it is presupposed that the realiser is “something going on”, an event or state.
all. MAR makes this option available by rejecting the inference from the fact that P essentially conceives S, and M ≠ P, to the conclusion that M does not essentially conceive S. According to MAR, it is possible for two concepts to be distinct, and yet for them both to essentially conceive the same referent. The key to making this option available is MAR’s rejection of the assumption that all concepts essentially conceive their referents in the same way. This makes room for an appeal to the difference between the ways in which mental and physical concepts essentially conceive, to argue against there being any incompatibility in two different concepts essentially conceiving the same referent. So with MAR, there is no need to compromise physicalism with mental properties, and no need to make fulfilment of a functional description criterial for mentality, thereby letting in conceivability arguments that draw on the conceptual distance between functional and mental descriptions.

MAR’s rejection of connoted functional descriptions for mental concepts brings in its wake some serious implications for the significance of conceivability arguments. We can see this clearly by employing David Chalmers’s terminology of logically possible and metaphysically possible worlds. Logically possible worlds reflect conceptual possibilities, so that there is a logically possible world for every world that is coherently conceivable under ideal reflection, whilst metaphysically possible worlds are just worlds that actually could exist. MAR’s claim is that the space of logically possible worlds does not coincide with the space of metaphysically possible worlds; the former exceeds the latter.

Now the sort of rationalism which warrants an unqualified inference from conceivability to metaphysical possibility has long been discredited, but a qualified

version of the inference remains current. This is the idea that any logically possible world corresponds to some metaphysically possible world, but the reflecting subject cannot know which one without empirical information. In short, if your ideas are clear and distinct, they will correspond to something possible, although you may not be able to tell what exactly. Chalmers explicitly defends this position, which derives from Kripke’s views about explaining the apparent conceivability of worlds in which necessary a posteriori identities fail to hold. Stephen Yablo has called the position “textbook kripkeanism” (TK), and we shall make use of this label.\footnote{Yablo (1999) “Concepts and Consciousness” As both Yablo and Chalmers have observed, it is not altogether clear that Kripke would agree with TK. He certainly never states the view that everything ideally conceivable corresponds to something possible, and though he appeals to principles closely akin to TK in his argument against the identity theory, the argument concerns rigidly designating expressions rather than concepts. For considerations relevant to interpreting Kripke’s views in this regard, see McGinn (1975) “A note on the essence of natural kinds”}

According to TK, our ability to coherently conceive of water not being $\text{H}_2\text{O}$, though we have discovered a posteriori that water and $\text{H}_2\text{O}$ are necessarily identical, does not show that we can conceive of the impossible, thereby discrediting conceivability intuitions altogether. Rather, it shows that we must be careful how we describe the possibilities we conceive of. Evaluated exclusively on basis of the intensions of our concepts available to a priori reflection, the claim is that we are not really conceiving of a world where water is not $\text{H}_2\text{O}$, but only a world in which the liquid occupying the water-role is not $\text{H}_2\text{O}$. And there are metaphysically possible worlds corresponding to this latter conception.

Kripke’s argument against the identity theory turned on there being no conceptual room for this sort of misdescription in the mental / physical case. He ruled out the possibility of explaining the sort of mindless logically possible worlds that challenge physicalism as misdescriptions of metaphysically possible worlds compatible
with physicalism. The basis for this was that phenomenal expressions are rigid
designators that do not have their references fixed by accidental properties of their
referents, and so it could not be the case that you are conceiving of being in a
qualitatively analogous epistemic situation, as Kripke might say, where the phenomenal
expression was being used to designate something else. Kripke’s view that phenomenal
terms are rigid designators which reference-fix on essential properties of their referents
follows from his denial of an appearance / reality distinction for consciousness.

MAR can agree with Kripke about all this, but will draw an entirely different
conclusion. For Kripke, the conceivability of pain without C-fibres firing must
correspond to a metaphysical possibility, because it cannot be explained away in the
usual manner. So the non-identity of pain and C-fibres firing is established. But for
MAR, pairs of phenomenal and physical concepts essentially conceive the very same
referent, and so a logically possible world in which those referents came apart would
not be metaphysically possible. So MAR renders such logically possible worlds as
insignificant excrescences resulting from our possession of two sorts of psychologically
unconnected concepts for essentially conceiving the same referents. Thus MAR rejects
the universal applicability of TK for explaining conceivability.

Now we come to a third and more programmatic way to look at the
significance of MAR, relating to our earlier analyses. This is that MAR proposes a
purely epistemic interpretation of the appearance / reality distinction. When
philosophers like Kripke, Searle or Chalmers deny that there is an appearance / reality
distinction to be made for conscious properties, they mean something metaphysical:
that conscious properties are what they appear to be, that they are appearances. MAR,

213 See Section 4.e.
on the other hand, interprets appearance as a way in which some properties are known. It says that one and the same property can count as an appearance or not, depending upon our mode of access to it.

Now though it is perhaps true that any version of physicalism must interpret appearance epistemologically rather than metaphysically, so as not to introduce a new genus with properties defined \textit{a priori}, MAR is distinctive in taking this observation to its logical conclusion. For if appearing is not having a distinctive feature, then there ought not to be anything novel for concepts of appearing properties to pick up on. MAR agrees that there is not; phenomenal concepts essentially conceive the physical or functional properties that physicalists want to identify with mental properties.

7.c. Three Takes on Conceivability

Modal rationalism is the view that so long as something is coherently conceivable, then there can be no principled reason why it should not exist: it is possible. This view is jeopardised by the conceivability of the falsity of quantitative identities discovered \textit{a posteriori}. TK is a form of modal rationalism that caters for this. It explains such cases of conceivability as due to a hidden contingency; not attaching to the identity, of course, but rather to our manner of thinking about the identity. According to TK, the two concepts employed to think about the identity must have their references fixed by different properties, at least one of which must be a contingent property of the referent. Then TK explains the apparent conceivability of a necessary identity being false, as the conceivability of a world where the contingent property which fixes the reference of one of the terms of the identity, has a different
bearer to its actual one. This really is possible, so modal rationalism is vindicated.

Every genuine act of conceivability is explained with a real possibility.

Chalmers expresses TK using his two-dimensional semantics.\textsuperscript{214} He distinguishes two intensions associated with any concept, that may or may not coincide. These are different functions for determining the extension of the concept. The primary intension of a concept determines how the extension of the concept is fixed in the actual world, and the secondary intension determines what the extension of the concept is in counterfactual worlds, on the basis of its extension in the actual world. Determining the primary intension of a concept is an \textit{a priori} matter; Chalmers approximates the primary intension of \textit{water} as picking out the “watery stuff”. But the secondary intension is determined empirically by finding out the extension of the primary intension at the actual world. So there are both \textit{a priori} and \textit{a posteriori} elements to the meaning of a concept.\textsuperscript{215}

Conceivability is an \textit{a priori} matter, and so our access to the space of logically possible worlds through conceivability is according to primary intensions only. That means that if something is conceivable, it only establishes that there is a logically possible world in which the primary intension of that conception is true. Which metaphysically possible world this corresponds to depends upon the secondary intensions employed, which are not accessible to reflection alone. This explains the combination of the conceivable falsity of \textit{a posteriori} identities (evaluated according to their primary intensions) and their necessary truth (evaluated according to their secondary intensions). TK claims that such explanations are always possible, and consequently that for any true \textit{a posteriori} identity, it is possible to distinguish between

\textsuperscript{214} Chalmers (1996) \textit{The Conscious Mind}, p. 56 \& ff.
\textsuperscript{215} On the assumption that meaning is intension.
the primary and secondary intensions associated with at least one of the terms of the identity. If this distinction could not be made, then conceiving of the identity not holding would either establish that the identity was false, or that modal rationalism was false.

Kripke argued in effect that the distinction could not be made for putative identities between mental and physical properties, and so the identity theory was false. But Chalmers uses TK to argue against physicalism without relying on the claim that for mental concepts primary and secondary intensions coincide.\(^{216}\) The argument is that since we can imagine a physical duplicate of our world lacking consciousness (a "zombie world"), then even if this is just a world where the manifest properties of consciousness associated with the primary intensions of our concepts are lacking, it is still a world where different facts hold from our own world although nothing is physically different. So physicalism, construed as a supervenience thesis about facts, is proved false. As Chalmers points out, physicalism would receive the same indictment if we could conceive of a physically identical world to our own where the manifest properties of water associated with the primary intension of our concept \textit{water} are lacking. But he denies that we can.\(^{217}\)

Levine, who offers our second interpretation of the significance of conceivability, agrees that we cannot conceive of a physical duplicate of our world where H\(_2\)O does not have the normal macroproperties of water associated with the primary intension of our concept. But for Levine, conceivability is a reflection of the relations between our representations, and does not yield modal knowledge. On his understanding, our inability to imagine the macrobehaviour of water altering whilst

\(^{216}\) Nevertheless, he does accept this claim: see Chalmers op. cit., p. 133.

\(^{217}\) Op. Cit., p. 133.
everything microphysical is held constant, shows that we can explain the macroproperties of water with microphysical theory. Explanatory frameworks constrain conceivability, for they impose necessary connections between our representations; given a certain microphysical structure, certain macrobehavioural traits could be shown to follow within a D-N explanation.

Levine also agrees with Chalmers that a physical duplicate of our world where facts about consciousness cease to hold is conceivable. But on his picture this demonstrates an independence between our representations of consciousness and of physical properties that is incompatible with an adequate understanding of physicalism. Our explanations contain gaps, and so there is no psychological compulsion for us to move from a representation of a physical condition to a conscious property. The conceivability of a zombie-world shows only that we have not made explanatory connections between the primary intensions of our mental and physical concepts.

So we have here two different explanations of the conceivability of consciousness in the absence of any particular physical or functional property. TK holds that a metaphysically possible world explains the conceivability. Levine holds that our inability to explain consciousness in terms of physical theory explains the conceivability. In each case, conceivability tells us something important about consciousness. For TK, it tells us something metaphysical: that it cannot be identified with anything physical. For Levine, it tells us something epistemological: that we do not understand it in physical terms.

MAR’s view of conceivability is different again. According to MAR, conceivability intuitions may not tell us anything metaphysical or epistemological; they may be entirely delusory, and simply a product of our psychological make-ups. This
view derives from the work of Nagel, as a number of MAR theorists acknowledge.\(^{218}\)

But that is not to say that Nagel draws the same conclusions about conceivability as MAR does, as we shall see.

Nagel rejects TK, saying that,

Kripke’s view about how we must explain away an illusion of conceivability is too restrictive.\(^{219}\)

For Nagel, illusions of conceivability need not always be explained as the misdescription of a possibility. The conceived scenario might not be possible at all because there may be necessary connections in the world which we are oblivious to, our ignorance allowing us to conceive in isolation properties that are necessarily connected. In particular, Nagel thinks mental and physical properties may both turn out to be essential properties of experiences, in which case the present conceivability of disembodied experience counts for nothing. To back up this scepticism about modal intuitions, based on the possibility of necessary connections we do not know about, Nagel introduced the idea of a cognitive defeater of imaginability.\(^{220}\) This is supposed to explain why the mind-brain relation might appear to be contingent even if it is necessary, and turns on a distinction between types of imagination, “perceptual” and “sympathetic”.


\(^{219}\) Nagel (1986) *The View from Nowhere*, p. 47.

\(^{220}\) Nagel seems to regard imaginability as the source of conceivability intuitions. Christopher Hill (op. cit.) emphasises the distinction between imaginability and conceivability, though his psychological explanations of feats of imagination and conception that seem to challenge physicalism come down to much the same thing: the separation between our imaginative faculties on the one hand, and the separation between our concept-forming capacities on the other. David Papineau (op. cit.) is also careful with the distinction, but most MAR theorists talk mainly about one or the other; conceivability in the cases of Tye and Loar, and imaginability in the case of Lycan.
The account proceeds as follows. When we imagine a conscious state in the absence of a physical state, we sympathetically imagine the conscious state by entering a state that resembles it, and then perceptually imagine the non-occurrence of the physical state, by entering a state resembling the one you would be in whilst perceiving the brain in some other state. In virtue of performing these two acts of imagination, we think we can imagine any mental state without its associated brain state and vice versa. But because the different types of imagination are independent, the acts of imaging the conscious and physical state involve entering into cognitively unconnected states. Because these states are unconnected, we can easily imagine one without the other. So mind-brain connections would seem contingent to us, even if they were necessary, because our ways of thinking about first-person and third-person phenomena are so divergent.

The moral MAR theorists saw in this was that if psychology alone could explain away dualistic conceivability intuitions, they would be freed of the obligation to invoke functional descriptions associated with the primary intensions of mental concepts to do the job. A psychological explanation could be invoked to block the inference from the fact that two co-referring concepts with no a priori connections between them must have had their references fixed differently, to the conclusion that they must be associated with different reference-fixing properties. The apparent contingency of mental-physical relations could be accounted for without any contingency creeping in between the concepts employed and their referents. This leaves MAR free to claim that both mental and physical concepts essentially conceive

---

221 Nagel (1974) “What is it like to be a bat?”, footnote 11.
222 Or apparent contingency, since it might be necessary for a physical property to satisfy a functional description connoted by a mental concept, without our understanding why it is necessary.
their referents, which rules out the possibility of conceptual connections being made between them, given that there simply are no reference-fixing properties whose realisation could be deduced from physical theory and the microphysical referent of the concepts.

MAR’s psychological explanation of conceivability intuitions has the following outline. Physical and phenomenal concept-forming capacities are guided by different sorts of experiences, and so have independent and unconnected cognitive roles. If we take one particular mental state, we can have both a physical concept for classifying the state as observed, and a phenomenal concept for classifying the state as the state that we are in. In addition, these different sorts of concepts conceive their referents differently, the physical concepts being theoretical concepts, and the phenomenal concepts being recognitional concepts. So physical and phenomenal concepts can play entirely different roles within our cognitive economies, and conceive their referents differently. MAR contends that this is explanation enough of the lack of inferential connections between physical and phenomenal concepts. If we conceptualise the same thing in two such different ways, on the basis of entirely different experiences, then we ought not expect to be able to tell that we have picked up on the same referent twice.

Using the psychological independence of different types of concept-forming capacities to explain away conceivability intuitions facilitates MAR’s claim that both mental and physical concepts essentially conceive their referents, which if true, would mean that physicalism had discharged its explanatory obligations. But the problem with appealing to psychology in this way is that it threatens a general scepticism about the

223 Keith Gunderson (1970) ‘Asymmetries and mind-body perplexities’, sought to explain anti-physicalist intuitions with just this point. Not all physical and phenomenal concepts are like this however; we can have physical concepts about states we are in (standing up), and third-person phenomenal concepts of other people’s states. Neither are essentially conceiving.
import of conceivability. As Chalmers points out, in response to a typical version of MAR’s psychological explanation by Christopher Hill and Brian McLaughlin,

If this account tells us that Cartesian modal intuitions are unreliable, the same will go for all modal intuitions.\(^{224}\)

Chalmers suggests we could as easily give a psychological explanation of why red squares are conceivable, since the concepts of colour and shape are presumably governed by separate psychological processes.\(^{225}\)

Some might embrace this result, preferring to explain the conceivability of red squares with psychological role rather than correspondence to real possibility. But the emphasis MAR theorists place on their accounts of the different roles played by physical and phenomenal concepts suggests that they consider this an exceptional case. Some even allude to empirical evidence that applying these different concepts involves activity in different areas of the brain, which would be plain overkill if they meant psychology to indiscriminately defeat modal intuitions.\(^{226}\)

In fact, the mind-brain case is an exceptional one, to which Nagel’s psychological account of conceivability intuitions was specifically tailored, and MAR has picked up on this. The possibility of *a posteriori* identities shows that empirical discoveries can trump modal intuitions. This is because modal intuitions only employ primary intensions which determine how secondary intensions are fixed, but reductions are truth-evaluated on the basis of what secondary intensions reference-fix on, which

\(^{225}\) However, Hill does explicitly disassociate MAR from a “radical modal scepticism”. He intends his modal scepticism to apply only to, “*a posteriori* modal intuitions that are formed independently of information concerning the relevant empirical matters.” Hill, op. cit., p. 82. Cartesian modal intuitions form a sub-set of these, for which a psychological explanation is needed in order to reaffirm the priority of empirical evidence, as we shall see.
\(^{226}\) E.g. Papineau (1998) “Mind the Gap”.
must be empirically determined. But the mind-brain case offers an exception to this if both the two concepts employed have primary and secondary intensions that coincide. In such a case as this, empirical evidence that the secondary intensions of the two concepts are the same or otherwise necessarily connected would not automatically trump conceivability intuitions that seemed to show otherwise. If there were a necessary connection between their referents, there would also have to be a necessary connection between their primary intensions. So the empirical evidence of necessary connection and the \textit{a priori} evidence of separability here stand in problematic conflict with each other.

By invoking a psychological explanation for the separateness of the primary intensions, empirical evidence can be given the upper hand in evaluating \textit{a posteriori} identities across the board. It makes room for the possibility of a necessary connection between the primary intensions of two concepts that purportedly co-refer, even if there are no \textit{a priori} connections between those concepts and consequently their referents are conceivable in isolation from each other. There may be necessary connections in the world that we are constitutionally barred from “just seeing”.

So both Nagel and MAR see our psychological make-ups as an obstacle to grasping a necessary connection between the physical and the phenomenal that we might have expected to be able to see \textit{a priori}, given that the concepts in question have primary and secondary intensions that coincide. Given our disparate ways of thinking about the physical and phenomenal, their relations would seem to be contingent even if they were necessary.
7.d. Doing without Reference-Fixing Descriptions

By invoking the dimension of psychological role, MAR is able to explain without reliance on functional descriptions associated with the primary intensions of phenomenal concepts, how it might be conceivable that a phenomenal and a physical concept have distinct referents, when in fact they necessarily share a referent. But this does not give us any positive reason to believe that phenomenal and physical concepts necessarily share a referent, it just shows that conceivability intuitions cannot prove otherwise. This would be fine if MAR could compartmentalise itself as simply a response to conceivability intuitions, but since the response involves rejecting connoted functional descriptions, there are repercussions which extend far beyond this narrow remit. For with this rejection, MAR undermines a standard argument for physicalism.

The standard way to argue for physicalism since Smart, is to proceed from a functional analysis of the mental, to the empirical claim that all mental functional roles are realised by physical states. But according to MAR there are no functional descriptions associated with mental concepts for a functional analysis to pick up on. Mental concepts essentially conceive their referents. So MAR must provide alternative motivation for physicalism.

One way for MAR to motivate physicalism is on the grounds of the necessity of accounting for mental efficacy. Such an argument has been developed by David Papineau, in recognition of MAR’s need for an alternative argument for physicalism. The premises are that some mental causes have physical effects, all physical effects
have sufficient physical causes, and overdetermination is not widespread. The conclusion is that mental causes are physical causes.\textsuperscript{227}

Whether this argument can be made to work is not our present concern.\textsuperscript{228}

What does interest us in our assessment of MAR is the fact that once connoted functional descriptions are disavowed, then our mental concepts cannot be given a role in arguing for physicalism. There can be no analysis of mental concepts showing how they relate to the physical world, once it is held that they refer essentially. Accordingly, MAR’s physicalism about the mind, even if true, cannot count as an “explanatory reduction” in Levine’s sense. Explanatory reductions have two stages, one conceptual and one empirical:

Stage 1 involves the (relatively? quasi?) a priori process of working the concept of the property to be reduced “into shape” for reduction by identifying the causal role for which we are seeking the underlying mechanisms. Stage 2 involves the empirical work of discovering just what those underlying mechanisms are.\textsuperscript{229}

According to Levine’s model, explanatory reduction requires a D-N explanation that derives the realisation of the causal role discovered in Stage 1 from


\textsuperscript{228} The main difficulty for the conclusion of the argument, if it concerns properties or types, will be accommodating variable realisation. If mental causes are multiply realisable, then it is undesirable for the argument to establish their identity with physical causes. Papineau (op. cit.) points out that the argument would also establish the identity of mental causes with functional causes that are physically realised, and in defence of MAR, adds that even if variable realisation shows that mental properties are functional properties, it does not follow that our concepts of mental properties are functional concepts. The problem with this is that intuitions of variable realisation would seem to reflect the nature of our mental concepts. For example, we see a dog exhibiting pain-behaviour and judge that it is in pain, regardless of whether the animal has c-fibres, and this would seem to suggest that our concept is tied to function. MAR theorists can respond by making Chalmers’ distinction between our phenomenal and psychological concepts of mind, and insist that their concern is only with the phenomenal concept. But as we shall see over the course of this chapter, this would be another case of MAR appropriating apparatus designed to lead inexorably to positions they want to reject, just like Nagel’s psychological explanations of conceivability.

the underlying mechanisms mentioned in Stage 2. For example, in the case of the reduction of water to H\textsubscript{2}O, the concepts are connected in something like the following way. We analyse the concept of water as in part the concept of a colourless liquid. We can give an account of being a colourless liquid in microphysical terms; very roughly, groups of molecules are colourless if they do not filter light, and are liquid if they are loosely bonded together. Then our theory of H\textsubscript{2}O tells us that it could be loosely bonded together without filtering light in those normal circumstances under which we apply our concept of water. Once these conceptual connections are in place, then if we empirically discover that water is H\textsubscript{2}O, our microtheory of the behaviour of H\textsubscript{2}O will make sense of the behaviour of water.\footnote{We have already seen reasons to think that this model imposes perhaps overly strong constraints on explanatory reduction, in Section 2.d. However, the point we are making about the need for an account of the applicability of phenomenal concepts to physical states goes through just as well on the stronger or weaker models.}

There is no reason to think that both stages of the reduction should be accomplished simultaneously. In practice, it is inevitable that property reductions will be made before explanatory conceptual connections are in place. There may be empirical evidence that the properties are always found together and play the same causal roles, it may be possible to run tests on a sample of a natural kind to discover its underlying structure, or there may be theoretical reasons for distrusting some property, and thinking that it must really be some other more scientifically respectable property. Yet none of this presupposes that we can derive from underlying microproperties the instantiation of the macroproperties from which the everyday concept has its basis.

David Papineau actually notes, in support of his case that the causal argument is good enough for MAR, that,
while we can nowadays sketch quantum mechanical explanations of why H$_2$O satisfies the requirements of colourlessness, tastelessness, and so on, scientists became convinced that water is H$_2$O long before they could do this, in the first half of the nineteenth century, on grounds that were more like our argument from causation than any argument from realization.$^{231}$

This is an example of a reduction where the empirical stage preceded the conceptual stage. First there was overwhelming empirical evidence that water was H$_2$O, but no conceptual connections had been made between the concepts of water and of H$_2$O. Then later the reduction became an explanatory theory by facilitating explanations of the macroproperties of water, thereby forging conceptual links between concepts which were already thought to co-refer. So it might be said that before this, the theory that water was H$_2$O contained an explanatory gap.

Of course, MAR theorists will reject any analogy with physicalism about the mind, because they hold that phenomenal concepts essentially conceive their referents. But the point is that pending their argument that phenomenal concepts must essentially conceive their referents, Papineau's analogy looks like a good one. We do seem to be in the position of having good empirical evidence that mental and physical properties are identical, whilst remaining unable to connect our mental and physical concepts. We do seem to have a sound basis for a reduction, but lack the necessary understanding to make it into an explanatory reduction. MAR theorists must agree with the half of this that says we have a sound basis for a reduction from causal evidence and theoretical considerations, otherwise their physicalism is unmotivated. But they must go on to deny that the situation is akin to all other cases where a natural phenomenon is reduced, for these reductions facilitate or promise to facilitate conceptual insight into the reduced property.

The rather indirect way MAR must argue for physicalism, which makes no conceptual connections that might provide epistemic leverage on what we are being asked to believe, suggests that physicalism about the mind contains an explanatory gap. If MAR wants to instead hold that in this unique instance, we can never go beyond the indirect arguments so as to make conceptual connections, all on account of peculiarities concerning our concepts, then it must provide an argument that phenomenal and physical concepts both essentially conceive their referents.

MAR does not discharge this obligation. Rather, MAR theorists place all their emphasis on establishing the possibility that two concepts might essentially conceive one and the same essence. Their strategy is to show the availability of an option previously overlooked. MAR’s case rests almost entirely on its ability to solve the problems facing physicalism on the assumption that we possess two different sorts of concept for essentially conceiving the same referent which are not inferentially connected. If our concepts were like this, then we could dismiss the worry that there is something we do not understand about physicalism: we would be expecting a conceptual link when, due to the nature of the concepts in question, there could not be one.

The problem with this strategy is that even if MAR has uncovered a previously unnoticed option which would expose the explanatory gap as illusory, that does not mean it is illusory. It may be that mental concepts connote functional descriptions, and our inability to relate them to physical concepts shows that there is an explanatory gap which we are presently incapable of closing. The lack of conceptual connections between our mental and physical concepts, resulting in the conceivability of mental states without physical states and vice versa, is no more evidence for MAR than
evidence for a problematic explanatory gap. The empirical case for physicalism only
gives us good reason to believe that both mental and physical concepts refer to
physical states, but it is neutral to the question of whether or not those concepts both
refer essentially.

MAR’s argumentative strategy makes more sense if we add in the assumption
that there is not a problematic explanatory gap. Then the theory that we have two
psychologically distinct sorts of concepts for essentially conceiving the same properties
would adequately explain explanatory gap intuitions away. But MAR and a
problematic explanatory gap are competing accounts for those who accept causal
arguments for physicalism.

It is typical for MAR theorists to conclude their accounts by saying that they
have exposed an illusion. For instance, Tye says the explanatory gap is,

>a cognitive illusion, induced by a failure to recognise the special character of
phenomenal concepts.\(^{232}\)

And in the same vein, Loar concludes that,

>At the root of almost all weird positions in the philosophy of mind lies this
rather elementary and unremarkable conceptual fact [that phenomenal concepts
essentially conceive their referents] ... it is a mistake to think that, if physicalism
is true, consciousness as we conceive it at first hand needs explaining in the
way that liquidity as we ordinarily conceive it gets explained.\(^{233}\)

Both Tye and Loar miss out the conditional, for they have only made the case that if phenomenal concepts essentially conceive, and physicalism is true, then we are in the grip of a cognitive illusion.

We should consider MAR for what it is: an alternative to the view that there is a problematic explanatory gap in our theories of consciousness. Evaluating it on this basis, the evidence is ambivalent. We have seen that though there is good causal evidence for a reduction of consciousness, we are unable to make explanatory conceptual links on the basis of that reduction.\(^{234}\) This predicament is neutral between the two theories: perhaps we have not learnt how to make the conceptual connections yet, and perhaps our cognitive constitutions prevent us from making the conceptual connections. The former option, that there is a problematic explanatory gap, is supported by the many conventional examples of reductions that have operated in this way, whereas if the latter option obtains, this would be entirely novel and unprecedented.

However, this last thought should probably not hold too much sway, since the mind-brain case is a particularly important and exceptional case. Even if there is a problematic explanatory gap, it will be a severe one, since *prima facie* at least, it seems the fundamental conceptual difference between subjectivity and objectivity must be bridged.

As we keep looking, though, the evidence against MAR and in favour of a problematic explanatory gap mounts up. Let us now move on to consider another consequence of MAR’s rejection of functional descriptions associated with the primary

\(^{234}\) Though we can even now provide some functional mappings from neural structures to relations between phenomenological qualities; see Van Gulick (1993) “Understanding the Phenomenal Mind: Are we all just armadillos?” (discussed in Section 6.b).
intensions of phenomenal concepts. MAR denies that connoted descriptions are needed to explain the *a posteriority* of mind-brain reductions, because psychology can do that work instead. MAR also denies that functional descriptions are needed to argue for physicalism, because a causal argument can be provided instead. But there is a third role that connoted descriptions play.

The functionalisation of the mind was not just a premise for physicalism. It also provided an explanation of how it could be that if mental and physical concepts refer to the same state, the conceptual difference between them would not ensure that they did so via different properties corresponding to the different concepts. The functionalisation of mental concepts *explained* the difference between mental and physical concepts of the same thing. Conceptual difference was explained by the functional descriptions which mental concepts connoted, but physical properties did not. In rejecting this move, MAR reinstates the problem of conceptual difference.

The problem of conceptual difference, in order to be a problem, presupposes that a very basic principle that could easily be overlooked or taken as axiomatic. Connoted functional descriptions kept physicalism in conformity with this principle. In order to allay themselves of the problems caused by using functional descriptions, MAR theorists have denied this principle. They do so quite self-consciously; Loar names it “The Semantic Premise”, and states it as follows:

A statement of property identity that links conceptually independent concepts is true only if at least one concept picks out the property it refers to by connoting a contingent [functional] property of that property.\(^{235}\)

---

\(^{235}\) Loar, op. cit., p. 600.
So if two different concepts both refer to the same property \( p \), the Semantic Premise requires that in order to account for the conceptual difference, at least one of the concepts must refer to \( p \) by essentially conceiving a distinct property \( q \) that reference-fixes upon \( p \). A straightforward denial of this principle provides as good a short statement of MAR as any.\(^{236}\)

But the Semantic Premise is hardly a spurious assumption that has only to be exposed to the light of day in order to be put into doubt. On the contrary, the upshot of the Semantic Premise looks simply to be that there is only one way to grasp any one property as it is in itself. In other words, no more than one concept can present the definitive conception of any one property. This sounds no more onerous than a commitment to realism. For a physicalist, the basic properties of the world are definitively conceptualised by physical concepts. It is hard to see how there could be logical space for other sorts of concepts to essentially conceive these properties as well, without claiming inadequacy in our physical concepts (opaque physicalism), or contravening the rule that nothing can have two essences.

However, MAR makes a denial of the Semantic Premise plausible by rejecting its implicit assumption: that there is only one way for a concept to essentially conceive

---

\(^{236}\) The falsity of the identity theory follows directly from mental essentialism and the Semantic Premise. If phenomenal concepts essentially conceive their referents just as physical concepts do, then the Semantic Premise requires that those referents are different. This means that in Kripke's argument for the identity theory, the appeal to conceivability is entirely superfluous. Mental essentialism is an explicit premise of that argument, and the Semantic Premise must be an implicit premise, otherwise the argument would quite evidently fail.

In fact, with the Semantic Premise in place, mental essentialism looks to lead inexorably to dualism, which is exactly why the identity theorists had to reject mental essentialism. Nagel found another alternative; mental concepts could essentially conceive appearances without the need to abandon physicalism if appearances were essential properties rather than essences. Then physicalism could be true but opaque, since physical concepts would not give us the full nature of their referents.

In general, though, the combination of the Semantic Premise and mental essentialism is anathema to physicalism. The obvious way for the physicalist to go is to deny mental essentialism, but as we have seen, no alternative conception of consciousness is forthcoming. MAR is an experiment with the alternative of rejecting the Semantic Premise instead.
its referent. Without this assumption, we cannot substitute talk of "essentially conceiving the referent of" for talk of "presenting the definitive conception of", as we did above, and much of the intuitive force of the Semantic Premise is dissipated. Just because a concept essentially conceives its referent, instead of referring by connoting functional descriptions, it does not follow that the concept presents the definitive concept of its referent, in the sense of defining its nature. This model of essential conception may be the right one for physical concepts, so that if you fully grasp the concept of H₂O as it is defined by physical theory, then you know the real nature of H₂O. But MAR denies that direct conception works like this for phenomenal concepts.

MAR theorists say differing things about how phenomenal concepts do refer, but within narrow parameters. They all want to position them somewhere between standard recognitional concepts and theoretical concepts. They do not want them to conceptualise or capture an intrinsic nature, and to this extent they want to align them with recognitional concepts, but they do want them to essentially conceive their referents without the need for reference-fixing descriptions, and to this extent they want to align them with theoretical concepts.

MAR’s synthesis is that they are recognitional concepts, but non-standard ones. They are non-standard, because unlike other recognitional concepts, they are not a priori associated with functional descriptions. But they still count as recognitional because they are applicable simply on the basis of recognition, without presupposing knowledge of the referent’s intrinsic nature. Some MAR theorists say they are demonstrative concepts,²³⁷ or indexical concepts,²³⁸ some that they refer through a

²³⁷ Loar, op. cit.
species of simulation,239 and some say that direct recognitional concepts are in a class of their own.240

The ideas of phenomenal concepts as demonstrative and as indexical fit naturally together for MAR’s purposes. If phenomenal concepts are demonstrative concepts of experiences, they could only be applied to the experiences of the subject using the demonstrative, and so must be indexical as well. If introspective demonstration is a legitimate notion, then it would only be our own experiences that could be picked out as “this”.241 The demonstrative aspect affiliates phenomenal concepts to recognitional concepts, for the ability to conceptualise something as of “this” or “that” sort does not presuppose knowledge of its intrinsic nature. And the indexical aspect promises to account for physicalism’s problem in accounting for the supposed subjectivity of phenomenal states, in terms of the restricted conditions of application of indexical concepts.

One reason for MAR’s appeal to indexical and demonstrative concepts is that their references are contextually determined. If phenomenal concepts are like this, it would explain why we cannot derive the satisfaction of a phenomenal concept, which depends upon the context of application, from the satisfaction of a physical concept, which does not. This is because indexical and non-indexical descriptions are never interderivable, for familiar reasons.242 So there can be no obligation to conceptually connect phenomenal and physical concepts, and thus there can be no problematic explanatory gap.

241 Introspective demonstratives are a common device in setting up the intuitive case for an explanatory gap; there are innumerable examples in MAR theorists and other deflationists, but especially in inflationists.
Levine has responded to this, arguing that if phenomenal concepts were bare demonstrative concepts, explanatory gap worries could not have arisen in the first place. The question of how a property which satisfies a physical concept could also satisfy the concept of being “this” could never have arisen, because like demonstrative concepts themselves, the question “doesn’t itself really have any cognitive content”. Only if phenomenal concepts refer via association with properties like greenness, or painfulness, which might yield to functional description, can we account for the intelligibility of asking how the same property could satisfy both physical and phenomenal concepts.

This certainly would be a problem if phenomenal concepts were construed as bare token demonstratives essentially conceiving their referents, “demonstrative arrows shot blindly that refer to whatever they hit”, as Levine puts it. But such a proposal ought not to get off the ground. For if phenomenal concepts were demonstratives, they would at least have to be type-demonstratives, since they are our means of re-identifying experiences. A generic demonstrative concept of “that” under which we indiscriminately subsumed particular experiences could not do this. If the proposal is to be plausible, we will need a variety of demonstrative concepts for classifying different experiences as of “this” or “that” type.

However if phenomenal concepts are type-demonstratives, we have no reason to think that they essentially conceive their referents. Type-demonstrative concepts are formed on the basis of our dealings with the manifest properties of their referents.

---

244 Levine (2001) Purple Haze: The Puzzle of Consciousness, p. 84.
245 Loar (1996) says they are type-demonstratives; Tye (1999) disagrees.
246 We may also need a generic concept, to allow for our forming a conception of fleeting or novel experiences while we are undergoing them. Tye accommodates this in his (1995) Ten Problems of Consciousness, p. 168.
Having conceptualised water as "that stuff" clearly presupposes no knowledge of its essence. And this seems to be how type-demonstratives work generally; they group together items according to a noticed similarity between them. We cannot form type-demonstrative concepts for essences, since according to physicalism, these are the unobservable microstructures and theoretical properties detailed by fundamental physical science.

So the suggestion that phenomenal concepts are type-demonstratives provides absolutely no reason to think that they essentially conceive their referents, and consequently no reason to favour MAR over the explanatory gap hypothesis. And adding in an indexical component does not help either, since recognitional concepts like water, that refer by connoting functional descriptions, tend to have indexical components, but theoretical concepts like \( H_2O \) never do.\(^{247}\)

MAR has a very fine balance to strike between phenomenal concepts as recognitional and as essentially conceiving: too fine. Phenomenal concepts have to be recognitional concepts so as to keep them vague enough for us not to be able to ascertain physical structures in introspection. But as direct grasps of essence, we have to be recognising the property itself. The problem is that according to physicalism, the property itself is the physical structure. Loar tells us that,

odd though it may sound, the properties [phenomenal concepts] *phenomenologically reveal* are physical-functional properties - but not of course under physical-functional descriptions.\(^{248}\)

Then later in his paper, he adds that,

---

\(^{247}\) If the indexical component means that a Levine-style deduction cannot be obtained, then this may only mean that this requirement on conceptual connection should be loosened.

The illusion is of expected transparency: a direct grasp of a property ought to reveal how it is internally constituted, and if it is not revealed as physically constituted, then it is not so.\textsuperscript{249}

The problem, once more, is that the property which is being essentially grasped is the internal constitution. Loar has no room for this move.

In denying the Semantic Premise, MAR must maintain that phenomenal and physical concepts are different concepts, whilst denying that there is anything different whatsoever in what they conceive. Since any semantic difference has been ruled out, the difference between the concepts becomes psychological difference. The only difference between the concepts that MAR can allow is that they essentially conceive the same property, but in \textit{different ways}. Then the problem arises of how to gloss “different ways”. MAR cannot say that we conceptualise the same thing as physical and as phenomenological, since we are supposed to be conceptualising the referents as they are in themselves. But then it becomes entirely obscure how there can be two ways of conceiving of a physical property as the physical property that it is.

One proposal that at least addresses the issue, is that phenomenal concepts refer by simulating their referents.\textsuperscript{250} Inspired by Lewis’s ability hypothesis, the idea is that forming a phenomenal concept is forming a “template” of a brain state that can be re-activated at later times. But even if being able to simulate a type of experience constitutes “knowing what it is like” to undergo an experience, possessing this ability could at most represent a necessary condition on the possession of a phenomenal concept. Employing phenomenal concepts does not presuppose that we actually

\textsuperscript{249} Ibid., p. 609.

exercise this ability. In thinking about pain, we do not have to recreate a faint pain, as Hume might have insisted. We can just think about pains as sensations, as bodily, as sharp or dull, etc. The simulation proposal makes the mistake of equating our ability to imagine experiences, with our possession of concepts of experiences. But experiences are no different from tables and chairs in this respect; we can imagine tables and chairs, but we can just conceptually think about them as well.

Most of what MAR wants to say about phenomenal concepts is plausible; except their central thesis that phenomenal concepts essentially conceive. It is plausible that phenomenal concepts are recognitional concepts, because we recognise experiences or properties of experiences without knowing their true natures. For the same reason, it is plausible that they are type-demonstratives, and it also makes sense to attribute an indexical component to them. It is even plausible that possessing a phenomenal concept requires possession of Lewis-style abilities, since you cannot apply a phenomenal concept until you have had the experience. But absolutely none of this provides any reason whatsoever to think that phenomenal concepts essentially conceive their referents. Quite the opposite, since all acknowledged recognitional concepts have functional descriptions associated with their primary intensions, and MAR has not explained what "recognitional" could otherwise mean.

There are no other concepts which do what MAR wants phenomenal concepts to do, namely be applicable at the level of recognition whilst conceiving essentially. So MAR cannot soften the counterintuitive consequence of their theory, that introspective concepts essentially conceive physical or functional states, through analogy with a familiar type of concept. MAR's best tactic is to insist that phenomenal concepts belong to a unique class of direct recognitional concepts.
But the only reason for thinking that phenomenal concepts are direct recognitional concepts is that this would explain away the explanatory gap problem. To achieve this aim, MAR makes a series of *ad hoc* assumptions that are not independently motivated. The reduction of consciousness becomes a unique sort of reduction, the Semantic Premise is uniquely inapplicable to phenomenal concepts, and phenomenal concepts form a unique class of direct recognitional concepts.

If we held onto connoted functional descriptions, however, the reduction of consciousness can follow the standard pattern of reduction, the Semantic Premise can remain in place, phenomenal concepts can be recognitional concepts like any other, and essential conception can remain the sole domain of our theoretical concepts. There are however two downsides. The first is that we cannot accept at face value the intuition that for consciousness, the appearance is the reality. However, this is a rationalistic and dualistic intuition that physicalists should mistrust in any case. It is rationalistic because it tells us that reflection alone reveals the essence of consciousness, that it *is* what it appears to be. Since it does not appear to be physical (or non-physical for that matter) MAR embroils itself in obscurities by insisting that we only grasp it recognitionally. And it is a dualistic intuition, because appearance is *prima facie a sui generis* type of essence defined *a priori*.

The second downside is that there would seem to be a problematic explanatory gap. However, it might just be the case that there actually is such a gap. Possible facts should not be avoided at all costs. The evidence, if consciousness is appearance, certainly points to there being a gap, since we have good empirical evidence for a reduction of consciousness, and yet we are unable to conceptually link up our recognitional concepts with our theoretical concepts of brain properties. A similar
situation once prevailed for other reductions, as MAR theorists admit: it is worth noting that none of the MAR theorists reject Levine’s explanatory gap apparatus for other reductions.\textsuperscript{251} And yet they are prepared to make a series of \textit{ad hoc} philosophical moves that rule out the possibility of future neuroscientific findings making the conceptual connections we presently lack.\textsuperscript{252}

7.e. Recognitional Concepts and Essential Conception

So far, we have sought to establish that there is no good reason to believe in MAR, because MAR provides no arguments for its central claim that phenomenal concepts essentially conceive their referents. Instead we find a series of strong and \textit{ad hoc} claims, exclusively motivated by a desire to combine the apparently conflicting aims of avoiding a problematic explanatory gap at all costs, and upholding the essentialist intuition that there is no appearance / reality distinction for consciousness. Nevertheless, we have still not shown that MAR \textit{could not} be true. Admittedly, MAR theorists have not found any satisfactory characterisation of phenomenal concepts that would make sense of their being both recognitional and essentially conceiving. Nevertheless, it could just be the case that we do possess just such a special and unique class of concepts able to play this combined role. Our next task, then, is to argue directly against MAR.

\textsuperscript{251} Block and Stalnaker (1999) “Conceptual analysis, dualism, and the explanatory gap”, does seek to qualify it, though.

\textsuperscript{252} If someone had proposed that \textit{water} was a direct recognitional concept, before its reduction to H\textsubscript{2}O became an explanatory reduction, they would have eventually been proved wrong. The same may happen to MAR, and we have no reason to think otherwise without an argument. McGinn’s cognitive closure thesis is similar in that for all he says, it may still turn out false. But at least he has an argument.
MAR, as we have seen, wants to position phenomenal concepts somewhere between standard recognitional concepts and theoretical concepts. Obviously, they want to deny that phenomenal concepts are simply concepts of brain states; this is a position that Smart mentioned only to reject in his original presentation of the identity theory, as a misunderstanding of the physicalist project based on a conflation of sense and reference.\(^{253}\) Obviously, in applying phenomenal concepts in introspection, we are not discriminating physical structures of our own brains in any straightforward sense. For this reason, MAR theorists want to say that phenomenal concepts are only recognitional concepts, which is the intuitively default position in any case.

However, MAR theorists also want to deny that phenomenal concepts refer by connoting functional descriptions satisfied by their referents, which is the normal way for a recognitional concept to refer. The reason they want to do this is that if phenomenal concepts did refer in this way, then there might be a problematic explanatory gap. For suppose that the referent of a phenomenal concept is a physical brain state, and the concept referred to that state by connoting the functional (or functionalisable) description of "looking red". In that case it is legitimate to request an explanation of why the brain state should satisfy the description of "looking red". Such an explanation might take the form of a D-N explanation, as on Levine's model, showing how the instantiation of a microphysical property necessitated the satisfaction of the functional description "looking red". This is just what we cannot do, according to proponents of an explanatory gap.

MAR contends that physicalism has no such explanatory obligations because phenomenal concepts do not refer by connoting functional descriptions of their

\(^{253}\) Smart (1959) "Sensations and Brain Processes", p. 144.
referents. Rather, MAR says that phenomenal concepts essentially conceive their referents. So to this extent, they want to align phenomenal concepts with theoretical concepts, which also refer in this way.

The problem with this is that recognitional concepts need to be less fine-grained than the properties that they are concepts of, if they are to perform their function as recognitional concepts. If we are thinking about a property under a merely recognitional concept, then there must be some slack between our way of thinking about the property, and the property itself. Without this slack, the concept’s function as recognitional, based only on recognition, is disabled. The core function is to introduce a certain immunity from empirical discovery: our ordinary ways of thinking about things should not be held ransom to the deliverances of science. The recognitional concept of a rainbow should retain the same use whatever we discover rainbows to actually be.

The standard, and perhaps only model we have for understanding what makes a concept recognitional, is that such concepts refer by connoting functional descriptions of their referents. Thus the concept of a rainbow refers to whatever satisfies our commonplace descriptions of rainbows, based on their appearances to us. However, MAR theorists reject this model for phenomenal concepts. But they still maintain that phenomenal concepts are recognitional concepts, because they want to say that we can recognise our own brain states in occupying them, whilst avoiding the manifestly implausible claim that we thereby ascertain their physical constitution.

The problem with this is that if phenomenal concepts essentially conceive their referents, then they cannot play the role of recognitional concepts. We can demonstrate this with a thought experiment.
According to MAR, the phenomenal concept of pain essentially conceives pain itself, rather than simply referring to whatever it is that meets an ordinary functional description associated with the concept of pain. For a physicalist, pain is either a physical property or a functional property. So if MAR is right, our phenomenal concepts of pain must essentially conceive certain physical or functional properties, though science has not yet progressed to the point where we can say exactly what physical or functional properties these are. In itself, this is not a problem for MAR, which insists that phenomenal concepts essentially conceive these properties only in a recognitional way. So there is no question of introspection usurping neuroscientific research: MAR is clear of that charge. Our phenomenal concept of pain might essentially conceive pain itself, but since it is conceptually and psychologically isolated from theoretical concepts, it provides no basis for determining what pain is in physical terms.

But now suppose that sometime in the future, brain scanning devices become so advanced that scientists become confident that they know exactly what pain is. What they find is that without exception, whenever there is a case of pain, there is a case of C-fibres firing. However, they also find that there is a slightly different functional organisation to the brains of half of the world’s population than there is to the other half’s. The difference is that in one half of the population (A), firing C-fibres interact with the rest of the subject’s brain in such a way as to realise functional property \( F \), and in the other half of the population (B), firing C-fibres interact with the rest of the subject’s brain in such a way as to realise functional property \( F' \).

On the standard, non-MAR way of looking at things, we can describe this situation (assuming a functionalist account of pain) as one in which the phenomenal
concept of pain turned out to refer to F for one half of the population, and F' for the other half of the population. The difference which has been discovered between A-people and B-people, is that A-people had been using their concept of pain to refer to F, and B-people to F'. Of course, this is a difference that nobody knew about, because before the scientists tested everybody, nobody had any idea whether they fell into category A or B (an esoteric neurofunctional characterisation that nobody would care about anyway). Everybody just used their ordinary, recognitional concept *pain* to refer to pain, though the scientists eventually discovered that for some people, pain was F, and for others it was F'.

If we describe the situation like this, then it would not differ in any salient respect from a situation in which the recognitional concept *water* turned out to have a different referent for different people. Suppose it were discovered that the Aral Sea is full of XYZ. The people who live around there never travel, and it turns out that even the rain in that region is XYZ. Consequently, these people have been using their concept of water (or what they call "water") to refer to XYZ. The secondary intension of their concept of water has turned out to be different to the rest of the world: the "that stuff" tied down by the functional descriptions associated with these people’s concept of water, is a different stuff than it is for the rest of the world. However, the functional descriptions themselves are the same as everybody else’s, so the primary intension of their water concept is no different; they have the same way of thinking about XYZ, based on their everyday dealings with it, as the rest of the world have about H₂O. If a swap occurred, nobody except the scientists would notice. This is because *water* is a recognitional concept.

---

²⁵⁴ The example, and the conclusions which are drawn from it, can be run for an identity theory by supposing that in A-people, pain is C-fibres firing, and in B-people, pain is D-fibres firing.
But now consider what MAR must say about the situation where pain is a different property for A-people and B-people. MAR cannot just describe this scenario as the discovery that pain is F for half the population and F' for the other half. Rather, they have also to say that the different halves of the population have entirely different concepts of pain. If phenomenal concepts essentially conceive their referents, then since one half of the population has a concept that essentially conceives F, and one half of the population has a concept that essentially conceives F', it turns out according to MAR that A-people and B-people have concepts of pain with different primary and secondary intensions. In other words, nothing about their first-person ascriptive ways of thinking about phenomenal pain turned out to be the same.

But if the situation described in the thought experiment ever obtained, it is hard to see why there would be any temptation whatsoever to say that something had been discovered about the primary intension of people’s concepts. There would be no more temptation than there would be in the situation where it is discovered that the Aral Sea is full of XYZ, to say that this discovery meant that the people of that region had a different way of thinking about water, based on the role it plays in their lives, to the rest of the world.

Empirical information is needed to tell us the secondary intensions of our concepts, but primary intensions are essentially an a priori matter. We know what we mean to some extent. But in trying to uphold the intuition that primary and secondary intensions coincide for phenomenal concepts (an intuition entirely at odds with the physicalist project), MAR holds our concepts ransom to empirical discoveries, when it seems that empirical discoveries can only bear on what our concepts refer to. The primary intensions of recognitional concepts need to remain immune to some extent
from the nature of the phenomena they are found to refer to.\textsuperscript{255} Finding out about the world should not be finding out about something that was already contained in our concepts, if only we had a way of reading them.

So MAR entails that empirical discoveries could bear on the primary intensions of phenomenal concepts. This seems to be a straightforward consequence of the theory, and it is hard to see how a MAR theorist could deny it. However, they might try arguing that this is a perfectly acceptable consequence along the follow lines:

Granted, if it were discovered that in some people, pain was F, and in some people, it was F', then on our theory, the primary (and secondary) intensions of their concepts would have to be different. But this is not a problem, because the phenomenal concept of pain is a very fine-grained recognitional concept, which just demonstrably labels a certain feeling. The feeling of pain \textit{just is} a functional property, so if the functional property is discovered to be different for different halves of the population, then the feeling would also be different. Consequently, different halves of the population would have different concepts to reflect that difference in feeling. The difference might reflect the slightly more “tingling” nature of pain in A-people, for instance.

But even if we accept that the feeling would be different for A-people and B-people (on the basis that the functional property is different, and the feeling is the functional property), we might still resist the conclusion that their concepts of those feelings should be different. After all, A- and B-people all live amongst each other, interacting within the same language communities. Consequently, even if their pains are different properties, they will not describe them as such, since A and B people have pains in exactly the same circumstances (when their C-fibres are made to fire), and have learnt to talk about them in the same way. So A-people would not \textit{describe} their

\textsuperscript{255} We can of course get clearer about our concepts through empirical discoveries; a new situation may instigate a degree of conceptual analysis from which we make implicit aspects of our concepts explicit.
pains as more “tingling” than those of B-people, or *vice versa*. And in these circumstances, it might seem odd to insist that they must nevertheless possess different concepts of pain. For this would be to allow for intersubjectively undetectable differences in *concepts*.

Even if we are prepared to accept this, though, the proposal has another, unpalatable consequence in that it rules out the possibility of developing our phenomenal concepts. For this, we need some distance between the feeling itself, and our concept of it, which MAR cannot allow for. Two different A-people, John and Mary, might possess phenomenal concepts of pain of differing sophistication, although *ex hypothesis*, pain is property F for both of them. Mary might be able to detect a lot more phenomenological detail in her pains than John can, and be able to make finer-grained distinctions between her pains. This could just be because she has thought about it more, and paid more attention.

Take wine tasting as an example. After attending a course in wine tasting, Mary might be able to detect an almond flavour in her favourite wine, which she used to be oblivious to. In this case, it seems natural to say that her phenomenal concept of the taste of the wine has changed. She can detect something in that flavour that she could not before, hence her way of thinking about that flavour has changed. We do not capture this by just saying that she has a new way of describing her experience, because if she had been told about the almond taste before starting her course, she might not have been able to detect that taste. Hence she would not have thought this description was applicable to the flavour of the wine at this point in time. First, her phenomenal concept needed to develop.
It might be countered that when she comes to taste the almond, the taste itself has changed (has developed into a new functional property). But this is still unsatisfactory, because we lose the sense of Mary's appreciation of the wine increasing if there is no constant between the start and the end of her wine tasting course. And besides, it would seem odd to be forced into adopting this sort of position because of the rigidity of one's theory of phenomenal concepts ruling out the possibility of phenomenal concepts and phenomenal properties altering with respect to each other.

In short, MAR's contention that phenomenal concepts essentially conceive their referents, does not allow them to behave as recognitional concepts. Recognitional concepts are used to think about things based on our superficial dealings with them. They are ordinary ways of thinking about things based on the roles they play in our lives. We never recognise things by essence; we discover essences through empirical testing and theorising. If we only recognise a thing, we do not know what it really is. To think that we can recognise essences is a strong rationalistic idea that physicalists should reject. And the qualification that we recognise essences in a way that does not reveal their essence does not help matters: it just threatens the position with incoherence, rather than plain falsity.

Recognitional concepts must leave room for any number of discoveries about the referent of the concept. Otherwise, empirical discoveries about the properties become discoveries about our ways of thinking about the properties. The only concepts to offer an exception to this requirement are theoretical concepts, which need to be as fine-grained as possible to reflect the world as closely as possible.\(^{256}\) However, in the particular instance of this class of concepts, the concept dictates the nature of

\(^{256}\) Mathematical and other such concepts can be included in this category of theoretical concepts.
the property in accordance with its place in a theory, and so if an empirical discovery is made that shows that the concept is inappropriate or inadequate, then it has been discovered that the concept does not refer to anything after all, and a new theoretical concept will be needed to take its place. This is the only time empirical discoveries should bear on essentially conceiving concepts: to tell us that nothing satisfies them, and thus that there is no such essence.

7.f. Inappropriate Affiliations

MAR is a surprisingly unified theory, given that it has a variety of different proponents. But though different MAR theorists offer very similar versions of the theory, there is no similar consensus as to whether MAR (1) closes the explanatory gap, (2) shows that there is no gap, or (3) shows that the gap cannot and need not be closed. Option (3) is the most popular, though the others have their advocates.²⁵⁷

Presumably, MAR theorists do not think much turns on this, and might put their differences down to terminology. But if we strictly employ Levine’s idea of an explanatory gap, and read (1), (2), and (3) accordingly, we find three quite different claims. To establish (1) would be to facilitate an explanation of phenomenology in terms of physical theory, to establish (2) would be to show that phenomenology is explicable in terms of physical theory, and to establish (3) would be to show why phenomenology cannot be explained in terms of physical theory. Laid out like this, it seems clear that (3) best captures the intentions of MAR; the whole point of it is to

²⁵⁷ (3) is the view of Tye (1995), Lycan (1996), Loar (1996), and Sturgeon (2000). (2) is the view of Papineau (1998) and Tye (1999). Block and Stalnaker (1999) seem to hold either (1), the view that MAR closes the gap, or else the similar view that MAR clears the philosophical ground for science to close the gap.
show why we ought not to expect one of Levine’s micro-macro explanations in the case of consciousness.\textsuperscript{258} And yet, according to Levine’s conception of an explanatory gap, there certainly remains an explanatory gap even after we have accepted MAR, though it is not one we should worry about, as MAR theorists explicitly committed to (3) always point out.\textsuperscript{259}

However, once we see MAR as a position which shows why phenomenology cannot be explained in terms of physical theory, whilst nevertheless upholding the truth of physicalism, then a surprising analogy emerges. The analogy is with Colin McGinn’s Transcendental Naturalism (TN) about consciousness.\textsuperscript{260} Both theories share the very particular combined aim of providing a principled reason why we should never expect to close the explanatory gap, whilst insisting that our inability to close the gap does not reflect badly on physicalism in any way. And not only do the two theories have the same aims, but they use the same means too. Both blame the problem on our different modes of access to the physical and the phenomenological. For MAR and TN alike, it is because of the disparity between the concepts we form on the basis of our different modes of access to physical properties, that we cannot connect the concepts in our understanding. This cognitive contingency presents no indictment of physicalism whatsoever, according to both theories.

Yet despite the kinship between these views, the moral of TN is supposed to be that consciousness presents us with an unsolvable mystery, whilst the moral of MAR is

\textsuperscript{258} MAR theorists who describe their theory as showing (1) or (2) could easily defend their terminology in the following fashion. They could say that showing there is nothing more to be explained is closing the explanatory gap, or they could say that showing there is nothing to be explained is showing that there is no explanatory gap. However, this takes us away from Levine’s particular idea of an explanatory gap, and I doubt there would be much resistance to the claim that as I have rephrased the options, (3) best represents the import of MAR.


supposed to be that physicalism solves the problem of consciousness without remainder. Ideologically, these views are at opposite ends of the physicalist spectrum.

We can explain these radically different associations by looking at perhaps the clearest difference between the theories. This is that MAR purports to show there is nothing about consciousness to explain, whereas TN only purports to show there is nothing about consciousness that subjects with our cognitive peculiarities can explain. MAR views the explanatory gap as a "cognitive illusion" whereas TN views it as a "cognitive predicament". Once we strip away the rhetoric of "demystification", or "recognising our cognitive limitations", this effectively comes down to a disagreement about whether or not a non-human with the right sort of cognitive constitution could close the explanatory gap. MAR says there is not a problem, but it seems like there is because of how we are built; TN says there is a problem, but we cannot solve it because of how we are built.

However, to deny that there is a problem, MAR denies that there are any functional descriptions associated a priori with phenomenal concepts. MAR insists that phenomenal concepts provide direct grasps of the essences of physical properties; they represent ways of thinking about those properties that mirror their intrinsic nature in some way. The concept of a sharp toothache and the concept of the taste of coriander are concepts of the ordinary properties of the brain that science can investigate. If we are inclined to dismiss this, as the first identity theorists did as a matter of course in introducing their theories, MAR theorists will tell us that we should not think of this direct manner of conceiving a property in the same way as we think of it for theoretical concepts. Phenomenal concepts essentially conceive in a special, recognitional way. And here the explanation must stop, as MAR rightfully insists. But
only because MAR has invented a special sort of concept which rules out any further questions about how physical and phenomenal concepts can share a referent.

McGinn, on the other hand, allocates the mysterious properties to our brains, rather than our concepts, and consequently leaves it open for different sorts of subjects to come along and solve our problem. He thinks that principles of naturalistic explanation require there to be some property of the brain from which sense could be made of the connection between conscious properties, and the “hardware” of the brain. But our cognitive constitution bars us from representing that property.

Either way, whether the mystery is built into our concepts so that we cannot sensibly ask about it, or left in the open but beyond our reach, the tactic employed by MAR and TN is clear. It is: let us find a way of shelving the problem of consciousness as a matter of principle. But so long as consciousness is conceived of as appearance, it is bound to be problematic for physicalism, as we have been trying to show.261

The common source for MAR and TN, which explains their similarity, is Nagel. Both develop Nagel’s idea of a cognitive defeater of conceivability, which says that given our disparate ways of thinking about the physical and phenomenal, their relations would seem to be contingent even if they were necessary.

However, the conclusions Nagel and MAR draw from this are quite different. For Nagel, empirical evidence points to a necessary connection between physical and phenomenal properties, which our psychological make-ups prevent us from grasping. So psychology presents an obstacle to our understanding of the nature of this necessary connection. But MAR holds that physical and phenomenal concepts

---

261 The new, intuitive problem that arises for MAR, might be expressed: “how is my concept of this supposed to be a concept of a brain state?” This is no less troubling than standard formulations of an explanatory gap, except that MAR insists that we should not expect an answer to this question. Thanks to S.R. Allen for stressing to me MAR’s transference of the problem from properties to concepts.
essentially conceive the very same property, and so the necessary connection in question is that of identity. Since identities need no explaining, as MAR theorists frequently point out, intuitions of an explanatory gap are illusory. So for Nagel, the psychological explanation is of a cognitive predicament, which obstructs our understanding of the connection between physical and phenomenal properties. And for MAR, the psychological explanation is of a cognitive illusion, which makes us think physical and phenomenal properties are distinct when they are not. On the basis of a difference of opinion over whether phenomenal and physical concepts essentially conceive distinct essential properties or the same essence (or the same essential properties), our psychological make-up is painted as either the source of the problem of consciousness, or its solution.

The difference of opinion stems from differing strengths of commitment to physicalism and to essentialist intuitions about consciousness. Nagel is inclined towards physicalism (though he can only imagine an opaque physicalism being true), but has no doubt that for consciousness, the appearance is the reality. Given his latter commitment, he already knows what consciousness essentially is, and since he can see that physical concepts do not capture phenomenology, he adopts the view that phenomenal and physical properties are different (but necessarily connected) aspects of single essences.

MAR, on the other hand, is mainly concerned to uphold physicalism as an unproblematic doctrine, but thinks it can combine this with essentialist intuitions about consciousness. So it says that the appearances which phenomenal concepts essentially conceive are ordinary physical or functional states. But an unproblematic (non-opaque) physicalism and the lack of an appearance / reality distinction for consciousness are
awkward commitments to combine, for this involves saying both that a conscious property (the taste of coriander) is what it appears to be, and is a physical-functional property. But if it is what it appears to be, then why does it not appear to be a physical-functional property? MAR theorists provide a unique account of phenomenal concepts, making them quite unlike any other sort of concept, in an attempt to get around this. But they run into problems, as we have seen, and it is hard not to suspect that this is because they have appropriated a model (and the intuitions that go with it) entirely inappropriate to their deflationary ambitions. Nagel’s scepticism about conceivability allows for some sort of physicalism to still be true once you have already granted an intuitive gulf between the physical and phenomenological.

The whole raison d’être for MAR is to uphold, within a physicalist framework, the intuition that consciousness is appearance, whilst undermining the significance of the conceptual independence of phenomenal and physical / functional concepts. But if the view does manage to paint the concept of appearance epistemologically, as a way of knowing or thinking about a property, it does so only by making our phenomenal way of thinking about ordinary physical / functional properties deeply problematic. If phenomenal concepts essentially conceive their referents, then MAR is right to say that there is no problem about consciousness; but only because, as McGinn would have it, it is a mystery, not a problem.
Chapter 8: The Problem of Conceptual Difference: A Cycle of Thought

8.a. Introduction

We have already seen that the root of the explanatory gap problem is a conception of consciousness as appearance. The difficulty in abandoning this conception ("it is just what we mean by "consciousness"), combined with the impossibility of functionalising appearance, has lead to a deadlock in the debate. Neither deflationism nor inflationism can win the argument, so long as they base their case on either reinforcing our conception of consciousness as appearance (based on V#2), or playing it down in favour of the world-view of physicalism (V#1). Progress requires resolving the apparent conflict between V#1 and V#2.

In this chapter, we shall amplify our conclusion, by making the case that the situation we have described in the explanatory gap debate actually forms part of a historical cycle of thought on the mind / body problem. The cycle dates back to Descartes (at least), and is promulgated by the problem of conceptual difference. This is the problem of explaining the difference between mental concepts (more specifically, concepts of consciousness) and physical concepts. Descartes gave a sceptical response to this problem, and it has still not been satisfactorily resolved. We shall trace four stages in the cycle of thought around this problem, culminating in MAR, the attempt to solve the problem by fiat.
8.b. The First Stage

It does not seem to me that the human mind is capable of forming a very distinct conception of both the distinction between the soul and the body, and their union; for to do this it is necessary to conceive of them as a single thing and at the same time to conceive of them as two things, and the two conceptions are mutually opposed.

- Descartes, letter to Princess Elizabeth, June 28, 1643

The fact that we have different concepts of the mind and the body provided Descartes with an argument for dualism, since the difference allowed him to conceive of the one without the other, and on that basis argue for their distinctness. However, Descartes also held the view that mind and body are intimately united in each individual person, rejecting the inference from the separateness of mind and body, to a conception of the mind as an isolated executive unit in control of the body. This idea of unity was confirmed by his reflections upon the phenomenology of embodiment, especially bodily sensations, and also motivated in some way by his desire to make the mind and body fully interactive.

But though Descartes did offer an account of how mind and brain could causally interact, he never explained what their unity consisted of. Instead, we find him appealing to our cognitive limitations as an explanation of why we cannot comprehend this unity, as seen in the quotation above. This tactic echoes Saint Augustine, who also considered the unity of mind and body “beyond the understanding of man”.  

For Descartes, then, our divergent concepts provide a route to demonstrating the

---

262 The Philosophical Writings of Descartes, volume 3: The Correspondence, p. 227.
distinctness of mind and body, but they block our understanding of the unity of mind
and body. This is because we apparently cannot conceive of the mind and body in
different ways, whilst also conceiving of them as unified.

Descartes held that we conceive of the mind and the body by essence; he
needed this assumption to move from the separability of the ideas of mind and body to
their actual distinctness. Physicalists agree with the import of Descartes' idea that we
conceive of the body by essence, to the extent that they think physical concepts
especially conceive.\textsuperscript{264} The difference is that whilst Descartes' conception of matter
came from his \textit{a priori} reflections, modern physicalists leave it to empirical science to
determine how matter is best conceived of. Physicalists also agree to a surprising
extent with Descartes about how we conceive of the mind, as we saw in Chapters 3
and 4. This is evidenced by the widespread acceptance of the doctrine that there is no
appearance / reality distinction for consciousness.

But if our concepts of mind and of body both conceive their referents by
essence, then an identification of those referents, which physicalists must accept in
some shape or form, looks \textit{prima facie} to be ruled out simply in virtue of the
conceptual difference and the logic of essentialism. For reflection alone seems to show
that it is inappropriate to apply mental concepts to physical phenomena, or physical
concepts to mental phenomena. So if these concepts essentially conceive their
referents, then since no one thing can have two different essences, it seems that they
cannot share a referent.

\textsuperscript{264} This sort of terminology presupposes an essentialist realism, which fits naturally with physicalism,
even though it is possible to be a physicalist within some other sort of framework. But the issue about
conceptual difference in the mind/body problem is not just a problem for essentialist realists; it will
arise in some form or another for just about any framework. Pragmatism, for instance, still holds that
some descriptions are to be privileged over others on pragmatic grounds. So the pragmatist has the
problem of saying why it is that in the mind/brain case we seem to have two privileged descriptions
for the same thing. See Section 2.e.
From Descartes' time until quite recently, the idea that mental and physical concepts might both conceive the same referent essentially would not have seemed an option, because it was naturally assumed that a single essence can only be conceived as it is in itself by a single concept, and this is evidently not a case of grasping the same concept twice since mental and physical concepts are conceptually unconnected, a fact which played a crucial role in Descartes' argument for dualism. So with this option ruled out, the physicalist who upholds the intuition that mental concepts conceive their referents essentially, would have to maintain that mental concepts are concepts of properties; it would be a property of a mental state that a mental concept like "pain" applied to. Then the disparity between mental and physical concepts would not essentially bear upon the claim that an experience is identical to a brain state, but would rather bear on any purported identity between mental and physical properties. The pressure the conceptual difference exerts towards substance dualism is thereby transposed into a pressure towards property dualism.

There have been physicalists ever since Descartes' time, including a significant number of the early psychologists from the nineteenth century. However, such views had as yet no response to the problem of conceptual difference. C.D. Broad formulated the problem as an argument against "Reductive Materialism", the view that mental and physical properties can be identified.265 Broad pointed out that even if the same thing could bear the properties of being an awareness of red and being a molecular movement (thus allowing for the possibility of a materialist monism), there would still have to be two distinct properties. Otherwise, "being an awareness of red" and "being a molecular movement" would be nothing more than different names for the same

265 The entire discussion of Broad within this chapter is based on chapter XIV of Broad (1925) Mind and its Place in Nature.
property, as are “rich” and “wealthy”. Yet this cannot be so, since “being an awareness of red” and “being a molecular movement” do not mean the same; it makes sense to ask if the awareness is clear, but not if it is swift, and it makes sense to ask if the movement is swift, but not if it is clear.

Broad’s argument, put in our terms, is that since mental and physical concepts are different, then even if they refer to the same thing (event, state, essence), they must still introduce different properties in order to account for the conceptual difference. Broad assumes that the concept of an awareness of red will introduce a mental property. Consequently, he finds Reductive Materialism a “preposterously silly” theory, and considers emergentism the more sensible option for a materialist to take.

Let us try to clarify the argument. In our previous chapter, we distinguished between concepts that essentially conceive their referents, and those that do not. However, even those concepts that do not essentially conceive their referents (such as recognitional concepts), still essentially conceive something: the property they introduce, or express. In the case of a recognitional concept, this is the functional property given by the descriptions connoted by the concept. The bearer of this property is the referent of the concept.

Now mental concepts like pain or red afterimage must be concepts of something physical, according to the physicalist. Any concept that refers must essentially conceive some property or another, as we have just said, whether that property be the referent itself (as with concepts that essentially conceive), or a property which reference-fixes on the referent (given by the descriptions connoted by

266 Unless eliminativism is to be adopted. But even if it is, mental concepts must still be imperfect concepts of something physical, for they are, at least sometimes, reports of something occurring, and so clearly not analogous to concepts of fiction or of mathematics, for instance. This means that a similar problem will arise for the eliminativist who wants to assuage worries about “mental phenomena”.


the concept, as with recognitional concepts). Whichever way the details go, there must be some property to fit any concept that refers, whether it be the referent itself, or else a property by which reference is secured in virtue of this property belonging to the referent. So for the physicalist, mental concepts must either be concepts which essentially conceive physical essences, or concepts which essentially conceive essential properties of physical essences, or concepts which essentially conceive contingent properties of physical essences.

All three options face serious problems. The problem with mental concepts essentially conceiving essences is that we already have physical concepts which do this, and no physical concepts are the same as mental concepts, or interderivable with them. The problem with mental concepts essentially conceiving intrinsic essential properties is the same: we already have physical concepts which do this. A possible reply here would be that mental concepts essentially conceive essential properties of physical states or events for which we have no physical concepts. But if we take this route, then there are essential properties of physical stuff that present physical theory tells us nothing about. This forces us to the conclude that our grasp of the essence of matter is seriously incomplete (opaque physicalism).

The problem with the third option, that mental concepts essentially conceive contingent properties, is that the physicalist thesis that the real nature of a mental state such as pain is physical, is only plausible if that physical nature is essentially conceived by the concept of pain. Otherwise, fitting the mental concept would just be something that an essentially physical state might or might not do, and this contingency conflicts with the intuition that conscious states have appearance as their essence. Once

---

267 They may however essentially conceive functional properties; this move is considered in the next section.
conscious states are conceived of in this way, it becomes absurd to suppose that a particular state of pain might have lacked its painfulness, because the state’s appearance is its painfulness.

Broad felt no need to set the argument up in anything like this detail, and is worried that he may be accused of “breaking a butterfly on a wheel” in his discussion. The case against physicalism was supposed to be really obvious. It was just that mental concepts are of something, and they are obviously not of anything physical (though they might refer to something physical in virtue of essentially conceiving a mental property).

The presumption existed, then, that the fact of conceptual difference meant that you could tell \textit{a priori} that physicalism is false (if physicalism is defined so as to constrain properties), just by reflecting on what our mental concepts \textit{fit}, in the sense of what they essentially conceive. The concept of a perceptual state as all this (introspectively demonstrated), or of a pain, or of an occurrent thought; none of these concepts look like remotely plausible candidates to essentially conceive something physical. Reflection alone seems to establish that mental concepts do not essentially conceive anything which is also essentially conceived by a physical concept. And so this conceptual difference, which according to Descartes prevents us from understanding the unity of mind and body, was thought to be enough to rule out the physicalist’s radical suggestion of identity as an explanation of mind / body unity.
8.c. The Second Stage

When Smart and others made physicalism a serious philosophical stance, it was in large part through making the case that a conceptual difference between how we think about the mind and body should not count against physicalism. The easy part was tapping into general optimism about scientific explanation by aligning mind / brain identity to other less controversial scientific reductions where it had been empirically discovered that two distinct concepts share a referent. But even if mental and physical concepts referred to the same thing, there remained the problem of the property essentially conceived by the mental concept, which must either be the referent of the concept, or the property by which the concept secures its reference. Either way, there remain a class of properties which are essentially conceived by mental concepts, and a priori reflection alone shows that they are not part of the physicalist’s ontology. These properties exerted a pressure toward property dualism.  

Smart’s response was to offer topic-neutral analyses of mental concepts. According to these analyses, the properties essentially conceived by mental concepts are not incompatible with those essentially conceived by physical concepts. Rather, they are contingent, quasi-functional properties of physical properties. The classic example is of the concept of a “yellowish-orange after-image”, which Smart topic-neutrally analysed as the concept of “something going on which is like what is going on when I have my eyes open, am awake, and there is an orange illuminated in good light in front of me, that is, when I really see an orange.”  

---

268 This problem about mental concepts introducing mental properties was put to Smart by Max Black, but the point had already been made by Broad. See footnote 208.  
269 Smart (1959) “Sensations and Brain Processes”, p. 149.
functional because "something going on" is more than a place-holder; it prejudges that whatever meets the functional characterisation will be a process or event.

Topic-neutral analyses were supposed to capture the vagueness of mental concepts, the idea being that we do not really have any idea of the nature of the thing we are reporting when we report the occurrence of an after-image, only that "something is going on" which resembles other "somethings" that have gone on before. This is supposed to capture the sense in which an ancient Greek, or a Cartesian, could talk about their mental images and mean much the same as we do; mental concepts are neutral between language-users who uphold a dualistic metaphysics, a monistic metaphysics, or no metaphysics at all.\textsuperscript{270}

Whatever air of plausibility such considerations lend topic-neutral analyses though, the main work done by emphasising the vagueness of mental concepts is that this rules out the possibility of such concepts essentially conceiving the essences or essential properties of their referents. This means that the properties introduced by mental concepts in virtue of being essentially conceived by them, present no problem for physicalism. For the properties essentially conceived by mental concepts are too vague to define the intrinsic nature of their referents; they can only be contingent properties, and this leaves it open for the physicalist to argue they are the contingent properties of physical processes, all the essential properties of which are essentially conceived by physical concepts. Moreover, the contingent properties themselves do not jeopardise physicalism, since they are relational rather than intrinsic properties. This leaves it open for the relations in question to be between physical items.

\textsuperscript{270} We have already cast doubt on the idea that this level of neutrality is attainable, in Section 5.c.
Though topic-neutral analysis does answer the problem for physicalism of the introduction of properties essentially conceived by mental concepts, two major problems remain. The first is that if mental concepts essentially conceive contingent properties by which reference to a physical state is secured, then it becomes a contingent state of affairs that any physical state counts as mental. This is the problem we encountered earlier.

Now the intuitions of mental essentialism can be appeased without making conceptual difference problematic once more, and this along lines continuous with Smart's own, if topic-neutral analysis gives way to functionalism. Thus according to one sort of functionalism, at least, we say that mental concepts essentially conceive functional properties, as in effect the topic-neutral theorist does, but then instead of identifying mental states with physical states in virtue of these functional properties being contingent properties of physical states, the functionalist identifies mental states with functional states, which makes it the case that mental concepts introduce essential properties of mental states. This raises a number of complications, but shows the lines along which essentialist intuitions can be accommodated by making the switch from an identity theory to functionalism.

The second major problem comes from Broad's argument, which purports to show that even if it were true that an awareness of red is identical with a molecular motion, there must at least be two different properties introduced by the two concepts, since it makes sense to ask whether an awareness of red is clear, but not whether it is swift, and it makes sense to ask if a molecular movement is swift, but not if it is clear.

271 These were dealt with in Section 4.a.
Functionalism will not clearly help here, as it makes no more sense to ask whether a functional state is clear than whether a physical state is.

Smart addresses Broad's problem, though he never mentions Broad by name. His answer has two parts. For the first part, Smart points out that he is not claiming that "experience" and "brain process" mean the same, or even "have the same logic". The idea is that they mean different things, but have the same reference, as in the "morning star / evening star" model. Smart gives the example of "somebody" and "the doctor" referring to the same thing, although they have different uses. "Somebody" is the more circuitous and less specific of these two means of referring to the same person, and so Smart intends "somebody" as analogous to the mental concept.

Once more, the idea which does the work for Smart in avoiding this problem, and which gives substance to his claim that mental and physical terms have different senses, is the idea that mental concepts essentially conceive functional (or quasi-functional) properties. Since such properties present no challenge to physicalism, the problem of conceptual difference has been responded to without jeopardising physicalism. For if mental concepts essentially conceive functional properties, then we have an explanation of why mental concepts are different from concepts of physical intrinsic properties. Broad's point that "experience" and "brain process" cannot be different names for the same property can be concurred with, since Smart agrees that different properties are introduced.

However, though this explains why our concepts of mental states and brain states might be different, it does not touch Broad's point as applied to functionalism. If

---

272 Objection 5 of "Sensations and Brain Processes" is clearly concerned with the passage in Mind and its Place in Nature which we have been looking at, since Smart employs the same examples.
mental essentialist intuitions are to be appeased, so that it does not become a contingent state of affairs that a pain feels painful, then mental states will have to be identified with functional states. But then, the point that the conceptual difference between our concepts of mental states and physical states is explained by the fact that concepts of mental states are functional concepts, is no longer good enough to answer the problem of conceptual difference. This is because whatever mental concepts are, they are not detailed concepts of brain functioning, of the sort that might essentially conceive the functional states which are to be identified with mental states. We can see this by the fact that though it makes sense to say of an awareness of red that it is blurred but intense, it does not make sense to say of a functional property of the brain that it is blurred but intense.

To this, it might be responded that mental concepts only essentially conceive loose functional characterisations which tie down the more precise functional properties to be identified with mental states. Then there would still be two concepts to explain conceptual difference; the loose functional concept, and the fine-grained one. The problem with this response is that mental essentialist intuitions seem to be jeopardised once more, if meeting the loose functional characterisation is not essential to the mental state (which just is a fine-grained functional state). However, this is not necessarily a problem, since there is no reason that the functions essentially conceived by the mental concept might not form part of the specification of the scientifically defined functional property. Being "the sort of state that I enter when looking at a red object, etc." could perhaps be converted (via bridge principles to take us from ordinary language to a scientific vocabulary) to a partial specification of a fine-grained functional state.
However, the fact remains that even if mental concepts can be acknowledged to be different from scientific functional concepts, it remains nonsensical to say of any functional property (even one essentially conceived by an everyday mental concept), that it is blurred but intense. And this is because we have no such bridge principles between consciousness and brain functioning, of the sort that are obtained by reducing a property like being colourless to a certain spectral reflectance function. Consequently, it still makes no sense to say of a functional property of the brain that it is blurred but intense, even if we grant that mental concepts might be functional concepts.

The second part of Smart’s reply to Broad addresses this issue. Smart is of course talking about mind / brain identity, but the point is just as valid for the problem of conceptual difference facing functionalists. His suggestion is that such substitutions between mental and physical terms will come to make sense, perhaps as we become more scientifically sophisticated. The problem is not deep: we are just not used to knowing the referents of mental terms, the novelty being so great that substituting scientific for mental terminology still sounds nonsensical. But presumably, the same situation once prevailed when the molecular theory of water was at the vanguard of knowledge, long before it had filtered into common-sense. In those days, it would no doubt have sounded nonsensical to say “I drank a glass of H₂O”, yet most people would understand this sentence these days.

So perhaps the idea that mental concepts essentially conceive functional properties is enough to answer the problem of conceptual difference, and residual

---

273 Smart, op. cit., p. 152.
worries that mental concepts are different from any functional concepts are just a symptom of unfamiliarity with brain science.

8.d. The Third Stage

We have seen that Descartes' problem of the conceptual difference between our concepts of mind and body, which he supposed to bar any understanding of mind/body unity, cannot be overcome by simply pointing to the possibility that terms with the same reference might have difference senses. The reason for this is Broad's problem: difference of sense may not mean there will be a difference of reference, but it does mean that different properties will be essentially conceived by the different concepts.

Physicalists sought to overcome this problem in the second stage of the cycle, with the proposal that mental concepts essentially conceive functional properties. The introduction of functional properties by mental concepts explains conceptual difference, and since functional properties do not conflict with physicalist ontology, the monism of physicalism can remain a potential explanation of mind/body unity. This move allows you to grant that mental and physical concepts do not essentially conceive the same properties, whilst blocking the inference from conceptual difference to a class of mental properties which escape the physicalist net.

Both identity theorists and functionalists hold that mental concepts essentially conceive functional properties, and that these functional properties secure reference to the essences of the mental states to which mental concepts apply. The difference is that
the functionalist is free to hold that the property essentially conceived by the mental concept is essential to the state which is its referent.

But functionalism inherits a version of the problem of conceptual difference, because just as there are no conceptual links between mental and physical concepts, suggesting that mental concepts essentially conceive intrinsically mental properties, neither do there appear to be any conceptual links between mental and functional concepts, which makes it hard to see how mental concepts could actually be functional concepts. However, the fact that it does not seem as though our mental concepts essentially conceive functional properties, might just be a reflection of the paucity of our conceptual background knowledge. We may come to see that a mental concept of an appearance of red is really a functional concept, if our functional understanding of the brain provided insight into the appearance.

The third stage of the cycle of thought, however, reinvigorated the problem of conceptual difference by showing that our lack of understanding can be no minor affair, since there is a principled distinction between mental and functional concepts which needs to be bridged. This is Nagel’s claim to the effect that mental concepts essentially conceive and thereby introduce properties which are incompatible with the world-view of physicalism, at least as we now understand it. This is because conscious states are subjective. If mental concepts essentially conceive subjective properties, and no physical or functional properties are subjective, then physicalist proposals that mental concepts refer by essentially conceiving physical or functional properties must

\[274\] Nagel (1974) "What is it like to be a bat?"
be false, or at least far more opaque and expressive of faith in scientific progress than physicalists had intended.\textsuperscript{275}

Nagel’s idea is that our concept of a conscious state is the concept of something that it is \textit{like something} to be in. Since only the subject of the state is in the state, then only the subject has access to \textit{what it is like}, and consequently the concept of the state must be the concept of something subjective in order to reflect this. Given that all physical and functional concepts are of something objective, in that they essentially conceive physical and functional properties which do not need to be appreciated from any particular perspective, then \textit{prima facie} at least, it seems that it cannot be the case that mental concepts essentially conceive physical or functional properties. So we are apparently forced back to the view that mental concepts essentially conceive specifically mental properties, for the reason that no properties compatible with the physicalist’s ontology are subjective, or “like something” to instantiate. This is property dualism, a view which Nagel is favourably disposed towards, and which Broad considered a prerequisite aspect of any credible materialism. For Nagel, our need to acquiesce in this dualism reflects our lack of understanding, and being an opaque physicalist, he speculates that a future intellectual revolution will reveal mental and physical properties to be “essential aspects of a single entity or process”.\textsuperscript{276}

The difference between mental and physical concepts always meant that physicalism would conflict with our intuitions to some extent. Mental concepts obviously do not essentially conceive physical properties, as a physical concept such as

\textsuperscript{275} Smart, for instance, says that “whatever revolutionary changes occur in physics, there will be no important lesson for the mind-body problem”: Smart (1978) “The content of physicalism”, p. 340.

the concept of H\(_2\)O does. But when we make Smart’s switch to functional properties, the situation is no longer so clear-cut. And once we add in general reasons for believing the physical world to be complete, a physicalism about the mind that constrains properties becomes plausible, despite conceptual difference. Nagel turned the tide by pointing out that our mental concepts actually essentially conceive some properties which are alien to the physicalist’s world view. Our concepts of conscious states are not of anonymous states, but of subject-specific states. So the switch from physical to functional properties was apparently not as useful as had been supposed, since both are equally objective properties.

Nagel’s invocation of the subjective / objective distinction sparked off a new debate which focused on the subjectivity of mind, on the qualitative aspects of mental states, on phenomenal feel. The new focus was on the properties essentially conceived by mental concepts, which are apparently distinct in principle from any physical or functional properties. Nagel’s contribution was to point out a feature of consciousness conceptually isolated, and thus apparently inexplicable, from the resources of physicalism. And he did this, of course, by drawing attention back to our conception of consciousness as appearance.

The conceptual distance between mental and functional concepts is reinforced by a number of well known thought experiments,\(^{277}\) the distilled thought being that it is perfectly conceivable that any functional property of an objective physical system could be realised in the absence of the instantiation of the subjective conscious properties which functionalists want to identify with these functional properties. But as Levine has argued, and we have already agreed, conceivability arguments can only establish

that these scenarios are possible for all we know. In other words, they establish that we
do not know how to connect our ways of thinking about conscious properties with our
ways of thinking about functional properties. If a principle of understanding linking
functional performance with appearance were devised, the scenarios suggested by the
thought experiments might cease to be conceivable.

So recognising the subjectivity of consciousness does not necessarily rule out
physicalism, for it might be possible to provide a functional account of subjectivity.
The fact that functional concepts essentially conceive objective properties does not
threaten the functionalist’s claim that mental concepts are functional concepts so long
as subjectivity can be accounted for in objective, functional terms. However, if mental
concepts refer to fine-grained functional states by connoting coarser grained functional
descriptions, as has proved the least problematic route for physicalism to take, then the
fact that we have no principle to explain why the instantiation of a physical property
should be such as to satisfy these coarser grained descriptions is a substantial problem:
we have an explanatory gap.

We have an explanatory gap because these functional descriptions connoted by
mental concepts provide extrinsic characterisations of properties we are inclined to
describe as ‘‘subjective’’, and so if mental concepts really are functional concepts, we
will need some bridge principle to make sense of subjectivity in objective, scientific
terms. Otherwise, agreeing with Nagel’s point that consciousness is subjective would
amount to an abandonment of physicalism. This is a considerable explanatory burden,
and Nagel, for one, thinks that it would require a conceptual breakthrough that is
presently unimaginable. So Smart’s conjecture to the effect that a sentence like “this
functional property of my brain is intense but blurred” will come to make sense, is apparently hoping for no small thing.

8.e. The Fourth Stage

The latest move in the cycle of thought is another attempt to show that physicalism can unproblematically accommodate conceptual difference, which is quite different from the approach taken in the second stage of the cycle, that of functionalising the mind. This new approach is what we have called MAR.

According to MAR, it is possible for two different concepts to essentially conceive one property. If this is right, then there never really was a problem of conceptual difference, it just seemed as if there was one because of a false assumption that had been made. The reason for this is that the problem of conceptual difference, in order to be a problem, presupposes that different concepts cannot essentially conceive the same property (Loar’s “Semantic Premise”). If this is not true, then the difference between mental concepts and physical or functional ones does not necessarily present a challenge to physicalism.

MAR promises to clear up the problem of conceptual difference at a stroke. The problem was that given conceptual difference, then even if mental and physical or functional concepts referred to the same state, since only the physical or functional state could refer by essentially conceiving its referent, the mental concept would have to introduce another property to secure reference, and this property essentially conceived by the mental concept generated problems for physicalism. But if it is possible for the mental concept to essentially conceive the same state as the functional
or physical concept, as MAR claims, then no new property need be introduced by the mental concept. Consequently conceptual difference would leave no residue for physicalism to have to explain away.

Of course, physicalists always wanted mental concepts to just refer to something physical or functional, so the move MAR suggests cannot be entirely straightforward and unproblematic. The reason physicalists accepted that mental concepts would have to introduce properties distinct from their referents, was in order to explain conceptual difference, given that a priori reflection is enough to establish that mental concepts are not the same concepts as physical or fine-grained functional concepts. So MAR must take on the explanatory burden of accounting for conceptual difference given that both concepts essentially conceive the same referent.

The other related explanatory burden for MAR is to show that the option it points to is even possible. For prima facie at least, to essentially conceive a property is to capture its essence, or to definitively conceptualise its essence. But if there are two concepts, a mental and a physical or functional one, then it seems that either the single essence which is purportedly the referent of both concepts must be captured either by the one concept, or else by the other. There is no conceptual room for them both to do the job, since they are different concepts and there is only one essence. This seems obvious enough, and no doubt explains why nobody considered two concepts essentially conceiving the same essence to be an option until MAR came along in recent years.

MAR seeks to explain conceptual difference as a purely psychological difference. We have two sorts of epistemological access to some physical states: the mode of access we have to a state from perceiving it, and the mode of access we have
to a state from being in it. On the basis of these modes of access, we have formed different, psychologically unconnected concepts. But the difference between the concepts has nothing to do with what they conceive. It is put down entirely to the different conditions under which the concepts were formed, and thus the different purposes to which they are used.

To explain how it can even be possible for two concepts to essentially conceive the same essence, MAR theorists say a number of different things as we saw in the last chapter, but the basic idea common to them all is that mental concepts and physical or functional concepts essentially conceive properties in different ways. MAR insists that mental concepts are merely recognitional concepts, and consequently do not essentially conceive properties by elaborating them (or "structurally analysing" them\textsuperscript{278}), as might be the appropriate way to think of essential conception in the case of theoretical concepts. Rather, mental concepts essentially conceive their referents by simply latching onto them through acquaintance and recognition. Thus mental concepts do not capture the essence of their referents, as do theoretical concepts, but simply refer to their referents without the mediation of connoted functional descriptions. This supposedly alleviates the inherent tension in the idea of two different concepts essentially conceiving the same referent.

The idea of mental concepts as merely recognitional reveals MAR's similarity with the preceding attempt to solve the problem of conceptual difference within the cycle of thought. Smart's response to mental concepts not essentially conceiving physical properties, was to propose that they essentially conceive functional properties; they are the concept of being whatever it is that plays such and such a functional role.

\textsuperscript{278} Loar (1997) "Phenomenal States", p. 608.
This way, mental concepts do not essentially conceive an intrinsic nature, which accords with the intuition that it does not seem as though mental concepts are of anything physical, whilst leaving it open that they could in actual fact refer to something physical.

MAR works along similar lines. It says that mental concepts refer to something we are able to recognise, without thereby knowing its intrinsic nature. A common device employed by many but not all MAR theorists, is to align mental concepts with demonstrative concepts. Then the tactic is to say that mental concepts are no more committal to the nature of their referent than a concept of “this” or “that”. This way, no obstacle is placed in the way of the referents of these concepts turning out to be physical or functional, and an explanation is provided of why mental concepts do not seem to essentially conceive physical or functional properties, which is just that since they are only demonstrative concepts of properties, our possession of these concepts presupposes no knowledge of what these properties we are recognitionally acquainted with actually are.

MAR’s way of making mental concepts sufficiently noncommittal about the nature of their referent to avoid a conflict with physicalism, has the advantage over Smart’s in that it provides an answer to Nagel’s point that mental concepts are concepts of something subjective. On a view such as Smart’s, mental concepts introduce functional properties of their referents. But since functional properties are objective, functional analyses must either be false on the grounds that mental concepts essentially conceive subjective properties, or else there must be an explanatory gap pending an explanation of subjectivity in objective terms.
MAR has no such problem because it denies that mental concepts refer by connoting functional descriptions, which means that no functional properties are introduced. Consequently, there is no room for an explanatory gap, since this presupposes our inability to explain why these functional properties are realised. To explain the apparent subjectivity of consciousness, MAR invokes the idea that only the subject of a mental state can ever be in a position to apply a recognitional concept which has the function of classifying the state which the subject is occupying.

Now we have already presented a detailed case against MAR in the last chapter. The problem with MAR which we shall bring up here is much more general, and applies equally to the attempt to give a functional account of mental concepts. It is that all attempts to defuse the problem of conceptual difference by claiming that concepts of conscious states are essentially vague, simply being concepts of “whatever plays a certain functional role” or concepts of “this” or “that”, where we have no idea about the nature of the “whatever”, “this” or “that”, are guilty of underdescribing the physicalist’s predicament. Mental, specifically phenomenal concepts, are not metaphysically innocent. Rather, they are concepts of appearance.

To see this, consider the fact that when philosophers talk about the concept of phenomenal red, they are plainly not talking about the concept of a property of an object, such as a spectral reflectance function. Rather, they are talking about our concept of what it is like to experience this property of red objects. This is even true for a strong intentionalist like Tye, who is an externalist and an objectivist about colour.\textsuperscript{279} For Tye, to possess the concept of phenomenal red is to possess a concept applicable to representational states you are occupying, whose content is objective red.

\textsuperscript{279} Tye (1995) Ten Problems of Consciousness, pp. 144-150.
However, the concept is still a concept of a representational state, and is distinct from a concept of objective red, as can be seen from MAR’s insistence (with which Tye agrees) that third-person physical and functional concepts are typically formed on the basis of perceiving properties (i.e. objective red), and first-person phenomenal concepts are formed on the basis of occupying states (i.e. a mental state whose content is objective red).

But as soon as phenomenal red is marked off from objective red, appearance has been marked off from reality. To insist that phenomenal red is just a representation of objective red, or some other physical or functional state, is just to express a commitment to a physicalist account of appearance. But regardless of physicalist commitments (from the belief that V#1 must be true), the distinction between appearance and reality has been made (V#2 has been interpreted in a certain way, and accepted).

As soon as the distinction between appearance and reality has been made, the problem of conceptual difference will arise as a major obstacle to physicalism. And it cannot be solved by saying that concepts of appearance are essentially vague, noncommittal concepts, which might, for all we know, refer to physical or functional states. This sort of solution will never be satisfactory, for we have strong intuitions that concepts of appearance are quite determinate and substantive, and in previous chapters we have drawn attention to the deep metaphysical commitments attendant upon the concept of appearance.

Our concepts of phenomenal red, or what it is like to smell coriander, or to taste a vintage Colares red wine, are as determinate as any of our concepts. This is because we have a particular thing in mind, a definite idea of what we are talking
about. We use our phenomenal concepts to make fine-grained distinctions (consider wine tasting), to vividly imagine experiences we are not presently undergoing, and to think about properties to which we have demonstrative access ("that aching sensation").

Levine makes this point in saying that phenomenal concepts have a "'thick', substantive mode of presentation". He explains what he means by "substantive" in saying that when thinking of phenomenal red, "the reddishness itself is somehow included in the thought". He adds that the conception is "determinate" because the "reddishness presents itself as a specific quality, identifiable in its own right". The reason phenomenal concepts are substantive and determinate is the epistemic closeness of appearances. Since appearances are subjectively presented to us, phenomenal concepts provide a complete grasp of appearances (there is nothing more to an appearance than how it appears, and consequently, how we think of it as appearing), and since nothing else is presented in this way, we have concepts for these properties which are distinct from any other concepts.

In other words, \textit{a priori} reflection alone is enough to tell us what phenomenal concepts essentially conceive, namely appearances. MAR can agree with this, but will disagree with Levine that we have any determinate conception of what appearances are, beyond our ability to recognise them. However, as soon as it is agreed that appearances are the target, physicalism cannot be so unproblematic. For what makes our conception of appearance so determinate is the whole conceptual apparatus in which this concept is irrevocably embroiled.

---

281 Ibid., p. 8.
282 This is an interpretation of Levine's comments on the "connection between the distinctive nature of phenomenal concepts and subjectivity", at ibid., p. 9.
The concept of appearance is tied up with its opposite, reality. It forms part of an explanatory system devised to explain the epistemological fact that whilst we can be sure about how the world seems to us, we cannot be sure that how the world seems to us is how the world actually is. Appearances have an epistemological role: they inform us about external conditions, and explain why our take on external conditions is fallible. To achieve this, they need the epistemic peculiarity of being known more intimately than any other properties, so that we can be sure about the presence of these properties, without being sure whether the objective properties which the appearances purport to reveal to us are actually there or not.

These presuppositions about appearances, mean that they cannot just be something we neutrally identify by means of our phenomenal concepts, prior to uncovering their underlying reality. Smart’s proposal that the concept of a yellowish-orange after-image is the concept of “something going on which is like what is going on when there is an orange illuminated in good light in front of me”, seriously underdescribes the physicalist’s predicament. For as soon as we mark off our attention away from an actual orange thing, we have already classified the object of our attention as an appearance, with all the attendant conceptual baggage that entails. The analysis would have to read “an appearance like appearances which resulted from looking at an orange.” Likewise, MAR theorists cannot gloss phenomenal concepts as simply demonstrative concepts; they could at most have a demonstrative element, being concepts of “that sort of appearance”.

Physicalism deals only in realities, and so if phenomenal concepts essentially conceive properties of appearance, then physicalists must show that properties of appearance are actually properties of reality, i.e. physical or functional properties. But
since the concept of appearance carries an epistemological burden, then a physicalistic reduction requires the physical or functional properties identified with appearances to take on this burden. So the physicalist must explain how a functional state, for instance, can be known in the particularly intimate way that appearances are known, and why no distinction can be made between how properties of such a state seem to be to the occupant of the state, and how they actually are.

This means that so long as a physicalist targets their account of consciousness on appearance, accepting and working with the traditional appearance / reality distinction, then they inherit all the explanatory obligations that the traditional distinction was supposed to discharge. The special epistemic properties which appearances have by definition, now need to be explained in physical terms. Since nobody can do this, for we have no principle with which to connect our concept of an appearance with physical or functional states, there is an explanatory gap.

So long as consciousness is conceived as appearance, there will be a problem of conceptual difference; attempts to get around this problem whilst working within the framework of appearance are ultimately futile. Once the appearance / reality distinction has been made in the traditional way (V#2 has been given a certain interpretation), then we will think we have a determinate grasp of two different sorts of properties. Phenomenal concepts will essentially conceive one sort of property, and physical or functional concepts will essentially conceive the other sort. We can tell this by reflection alone, because the traditional appearance / reality distinction is a product of reflection (by reflecting on appearance, we distinguish it from reality).

MAR’s claim that phenomenal concepts essentially conceive physical or functional states, must be problematic if those states are supposed to be appearances.
To say that our concepts of appearances are nothing other than concepts of physical or functional states formed on the basis of occupying those states, fails to address the epistemic distinction between properties which we are aware of simply in virtue of instantiating them (subjective properties), and properties we are aware of through the instantiation of other properties (objective properties). Even if these properties are the same, we still need some account of why our concepts of them are so different. To say that phenomenal concepts are unique in their applicability to properties simply on the basis of our instantiation of them, simply obviates further questions by fiat.

The problem of conceptual difference will remain a problem for those who accept a conception of consciousness as appearance, so long as no explanation of appearance is available. Until then, it will continue to seem that appearances are a distinct set of properties that phenomenal concepts essentially conceive. And sentences such as “this molecular motion is intense but blurred” will not make sense, because we will retain the intuition that the molecular motion is one thing, and an appearance (which might be legitimately described as “intense but blurred”) is another.

8.f. Summary and Conclusion

(1) The problem of conceptual difference is affirmed. Descartes considered the difference between our concepts of mind and of body to be an insuperable obstacle to achieving an understanding of the unity of mind and body; mental concepts will never seem applicable to the body, and vice versa. This problem of conceptual difference was inherited by physicalism, since it seems evident that mental concepts are not concepts of anything physical. Broad sharpened this objection up: just because mental concepts
do not essentially conceive physical essences, it does not follow that they do not refer to physical essences. However, if they do refer to physical essences, they must do so by essentially conceiving mental properties, since whatever it is that they essentially conceive is not physical.

(2) The problem of conceptual difference is denied. The great breakthrough made by early physicalists like Smart was their attempt to show that conceptual difference was no objection to physicalism. We could have two different sorts of concept used for referring to the same thing. Just because a priori reflection will not reveal mental concepts as concepts of something physical, it might be discovered that they do in fact refer to something physical. There is no need to introduce distinctively mental properties; conceptual difference is explained by the fact that mental concepts essentially conceive functional properties.

(3) The problem of conceptual difference is affirmed. Nagel argued in effect that the properties essentially conceived by some mental concepts possess the special characteristic of being subjective. Consequently they could not be physical or functional properties, and so the difference between mental and physical concepts does compromise physicalism after all.

(4) The problem of conceptual difference is denied. Smart said there is no need to assume that mental and physical concepts refer to different properties just because they are different concepts. The Modes of Access Response says there is no need to assume that mental and physical concepts essentially conceive different properties just because
they are different concepts. So since they both essentially conceive the same referent, and no other properties are introduced in virtue of there being two sorts of concept for that referent, there are no subjective properties to account for.

The cycle feeds on our conception of consciousness as appearance, for there is no unproblematic resting place for appearances within physicalist ontology. Attempts to solve this problem have tried to convince us that the concept of mind is rather less specific, rather less committal than all that. But we have a clear conception of what appearance is supposed to be, and a clear grasp of the properties of appearance themselves. It seems that if we try to move away from this conception, we risk losing our grasp of what consciousness is supposed to be, but if we stick with it, physicalism inherits an explanatory requirement that it cannot meet.

The next stage in the cycle of thought, stage (5), would presumably be (and perhaps will be), a novel reaffirmation of the problem of conceptual difference. It would doubtless draw on some aspect of the conception of consciousness as appearance in order to show in a new and evocative way that mental concepts cannot just essentially latch onto physical of functional properties without residue.

Any attempt to break the deadlock between inflationism and deflationism which does not consist in a serious reappraisal of the idea of consciousness as appearance, is liable to (if not bound to) form part of the cycle of thought. The current deadlock in the explanatory gap debate is no isolated affair; it is an inheritance of the recent history of the philosophy of mind. And the paths taken by that recent history which we have traced in this chapter, have been determined to no small extent by the conception of appearance found (if only implicitly) in Descartes and Locke.
Significant progress has been made in the course of the cycle, however, as the conceptual heart of the problem of appearance has been gradually exposed, culminating in its present incarnation as the explanatory gap problem. MAR marks an important development, since the invention of the position reveals an awareness that the problem resides with our phenomenal concepts, and what they essentially conceive, rather than just what they refer to, which if physicalism is true, must be something physical or functional in any case.

Still, the fact remains that nobody really knows what to do with the conception of consciousness as appearance once they have taken it on board. For as soon as the traditional appearance/reality distinction is accepted, it is ensured that phenomenal concepts and physical or functional concepts will essentially conceive different properties. If the physicalist wants to say that some pairs of phenomenal and physical or functional concepts essentially conceive the same properties, then there must be some explanation of the peculiarities traditionally attributed to appearances, in physical or functional terms. This would require bridge principles, and the only way to meet this requirement would be to functionalise appearance. But this will not work precisely because of the epistemological peculiarities of appearance, as we saw in Chapter 5.

The problem gets a hold, as we have already seen, because of the apparent conflict between the background commitments of V#1 and V#2. Denying either of them would be implausible. What we need is a non-tenuous interpretation of them according to which they are compatible. The problem is not really with V#1; that is just an expression of scientific optimism, and commitment to scientific methods of explanation rather than rationalistic or superstitious belief systems. The problem is with the interpretation of V#2, which lands us with a conception of consciousness as
appearance. The only way to break out of the cycle of thought (and the deadlock between inflationism and deflationism), is to abandon any conception of consciousness as appearance, difficult (or even impossible) as this may seem. In the next chapter, we shall see what we can do with this, by now, obvious suggestion.
Chapter 9: How could this be a state of my brain?

9.a. Disentangling Consciousness and Appearance

In this chapter, we shall try to make sense of the idea that consciousness is not appearance. For if consciousness is not appearance, then the problems we have encountered for a physicalist account of appearance are not ipso facto problems for physicalism about consciousness. The intuition that there is no appearance/reality distinction to be made for appearance itself would be stripped of its power to generate essentialist intuitions about consciousness. The impossibility of functionalising appearance would not necessarily preclude a functional account of consciousness. Inflationist and deflationist strategies could lose their salience to physicalism about consciousness. And the problem of the conceptual difference between concepts of appearance and physical or functional concepts, might become tangential to issues about consciousness, so undermining the cycle of thought.

However, abandoning a conception of consciousness as appearance is no easy matter. It invites the response that to do so is to abandon consciousness itself, since appearance is just what we mean in talking about consciousness.\(^{283}\) This is the main problem facing deflationism. Deflationists can adduce considerations to cast doubt on traditional ideas of appearance, but as soon as a conception of consciousness as appearance is disavowed, we lose our grasp on what consciousness is supposed to be. The deflationist may then proceed to draw up a physical or functional account of

consciousness, but this invites the question of what makes their physical or functional account an account of consciousness. Moreover, as we have seen,284 deflationists are rarely if ever consistent in their rejections of appearance. They present critiques of inflationist conceptions of consciousness, but then in discussing their own physicalist accounts of consciousness, it often becomes clear that they take themselves to have provided an account of appearance.

The near universal assumption is that by “consciousness” we mean appearance, and that the real problem for physicalism is to show how appearances fit into the physical world. The possibility that consciousness might be something other than appearance is entirely neglected. But once the assumption that consciousness is appearance is in place, the results are bound to be unsatisfactory, because we have an independent grasp of appearance, provided by our conceptual background in the philosophical idea of an appearance / reality distinction, and by the apparent fact that we have direct awareness of appearances in introspection. So given a physicalist theory of what consciousness “really is”, we are at a loss to see how this could be a theory of appearance, something which we already have a definite conception of, a conception which the physical or functional theory will inevitably fail to make any conceptual connections with.

Assuming that it is obvious what we mean by “consciousness”, and focusing on the question of what we are referring to, is exactly the wrong approach for physicalists to take. If physicalism is true, then we are referring to something broadly physical. This is not the main problem for physicalism. The main problem is to connect what seem to be two totally different ways of thinking about something broadly physical, or what

284 See Chapters 4 (especially Section 4.h.) and 6.
comes to the same thing, to explain what first-personal thoughts about our own minds could have to do with the physical or functional properties which these thoughts apparently refer to. But if this is the problem of conceptually connecting co-referring concepts of appearance and physical or functional concepts, then as we have seen, there is little prospect of success.

Deflationists are prepared to assume that we know what we mean in talking of consciousness, because they make the mistake of thinking there must be some neutral sense of "appearance" they can avail themselves of, without endorsing the metaphysical conception of appearance which inflationists defend. But to mark appearance off as something itself, assuming that appearances are properties or events which we can isolate in introspection, is already to fall in with a certain interpretation of the appearance / reality distinction (and V#2), and thereby the conceptual apparatus of an epistemological and metaphysical explanatory system. The concept of appearance is philosophically loaded; it essentially conceives properties according to the epistemological roles they must play, rather than simply demarcating the territory towards which reductions must be targeted.

The assumption that in talking about consciousness, we mean to talk about appearances, has rarely been seriously challenged by physicalists. According to identity theories and functionalism, in talking about consciousness, we mean to talk about phenomena which play certain functional roles. This does not challenge the assumption that in talking about consciousness we mean to talk about appearances, given that the phenomena in question are supposed to be appearances, or what we call

---

285 One reason why physicalists keep making the mistake of concentrating on the reference rather than the meaning of mental discourse, is that they persist in setting up their arguments as arguments against dualism. In the process of defeating the dualist, they shirk the explanatory obligations of their own theories.
"appearances". The same goes for MAR, according to which phenomenal concepts are for picking out phenomena on the basis of recognition; again, the phenomena to which these concepts are supposed to be applicable are what have been called "appearances".

An exception to this rule is provided by analytical behaviourism, which actually challenges the assumption that in talking about consciousness we are talking about something called "appearance", rather than trying to accommodate that assumption. The analytical behaviourist enacts a conceptual analysis on mental discourse, translating it into discourse about actual and possible behaviours. This is supposed to present an alternative to the Cartesian theory that mental discourse is discourse about something called "appearance". So on the behaviourist view, appearances are pushed right out of the picture.

This is no doubt the reason why analytical behaviourism is so widely considered to be implausible. One standard complaint is that behaviourism misses out the inner states responsible for behaviour. This is particularly clear in the case of a sensation which is currently being undergone, but which is not eliciting any behaviour: to analyse discourse about it in terms of possible behaviours that could result, seems to miss out its current reality. In other words, the appearance itself has been missed out in restricting the analysis to what "appearances" are supposed to do. And given the strong intuition that consciousness is appearance, this seems to constitute a denial of consciousness. So one major reason for the implausibility of behaviourism is our attachment to consciousness as appearance.

Challenging the assumption that we just mean appearance in talking about consciousness has its risks, then. For without a link to appearance, it is hard for a physicalist theory to justify its claim to still be about consciousness. Nevertheless, it
simply will not work to take it for granted that talk of consciousness is talk of appearance, and then purport to provide a physicalist account of appearance. Either way, whether appearance is disavowed or not, the intuition will remain that consciousness has been missed out, whether because the theory is not supposed to be about appearance, or because though it purports to be about appearance, the theory fails to make the prerequisite conceptual connections for explanatory reduction.

The reason for this is that pending some good reasons why consciousness is not, and could not be appearance in any sense, thinking of consciousness as appearance will continue to seem all right to us. Technical critiques of some specific conception of appearance may convince us that nothing could possess the properties that appearance has traditionally been supposed to possess. But they do not challenge the apparently innocent, and apparently unquestionable intuition that nevertheless, regardless of whatever false philosophical presuppositions may have been made about something called “appearance”, still there is something which appears to us, and whatever the appearance is, be it a property, event, or state, it is a property, event, or state of consciousness.

This thought seems to be innocent; we do not have to think that the properties we are aware of in introspection are irreducibly subjective, intrinsically mental, or infallibly known, in order to think that they are nevertheless the properties to which a theory of consciousness must address itself. But in taking this line, we are not dealing with an innocent notion of appearance. If what appears is something distinguishable from what appearances are of (or what they represent), then V#2 has already been interpreted in such a way as to introduce properties which can only be known from a
certain perspective, and appearances have been attributed epistemological roles which will generate principled difficulties for physicalistic reduction.

So any attempt to claim that consciousness is not appearance must provide good reasons why consciousness is not, and could not be appearance in any sense. And to make this claim plausible, it will need to be backed up with an explanation of why we naturally thought consciousness was appearance, and what our mistake was. Otherwise, the intuition will endure. It is bound to endure if the rejection of consciousness as appearance leads us to lose our grip on the idea of consciousness. To avoid this, consciousness as appearance needs to be exposed as a mistaken conception of consciousness, and replaced by something better.

9.b. Introspective Access to Consciousness

Without the assumption that we have introspective access to properties of consciousness, there can be no explanatory gap problem. In introspection, we are aware of bright colours, tastes, and thoughts: “technicolour phenomenology”, as McGinn has called it.\(^{286}\) Physicalism requires that these properties are nothing over and above the ordinary physical world, and so must be realised by the instantiation of microphysical properties. But our conceptions of phenomenology and the physical world seem to have nothing to do with each other, so the intuition arises that our understanding of consciousness in physical terms must be seriously inadequate. A typical way to express this intuition would be to ask rhetorically: how could this be a state of my brain? The question could not arise if we did not have an introspection-

based conception of consciousness, with which we felt our physical or functional concepts do not connect.

If an introspective grasp of consciousness were not being presupposed, then the explanatory gap debate would not be so filled with talk of “mystery”, on the one hand, and “illusion” on the other. There could be no mystery surrounding the supposed identity of a conscious state with a physical or functional state, if we did not already have an intuitive grasp of conscious states provided by our introspective awareness of them, making it seem that conscious states possess properties which physical or functional states do not or could not have. Neither could there be anything like an illusion, if we did not have an intuitive grasp of conscious states alluring us to the mistaken supposition that consciousness presents some principled obstacle to a physicalist treatment.

Introspective access to consciousness is described in a variety of different ways, but despite different terminologies and emphases, there is an underlying idea which remains the same. This is that in introspection, we direct our attention upon the properties of appearance itself, rather than on what our appearances are of. Since there is a universal assumption that consciousness is appearance, which was the burden of our earlier chapters to establish, then this amounts to the idea that in introspection, we direct our attention upon consciousness itself, rather than what occupying a conscious state makes us conscious of. Most of the disagreement over introspection comes down to disagreement about how to understand “directing our attention”.

McGinn sees introspection as akin to perception, interpreting introspection as a type of faculty of awareness of our own conscious states and the properties of those states, in virtue of occupying them:
“Introspection” is the name of the faculty through which we catch consciousness in all its vivid nakedness. By virtue of possessing this cognitive faculty we ascribe concepts of consciousness to ourselves; we thus have immediate access to the properties of consciousness.  

For McGinn, the faculties of introspection and perception are sensitive to different parts of the world. He says:

these two faculties - sense perception and introspection - have different fields, meaning by this that they take different kinds of object of apprehension. They are used to detect different regions of reality.

The different “regions of reality” are what our experiences are of (which could be just about anything), and the experiences themselves.

Armstrong also likens introspection to perception, and construes it as an awareness of the mental state we are occupying:

By sense-perception we become aware of current physical happenings in our environment and our body. By inner sense [introspection] we become aware of current happenings in our own mind.

Perception, for Armstrong is “the acquiring of information or misinformation about our environment”; introspection is no different, except that the part of the environment it is concerned with is our own mental state. Introspection is the directing of one mental state upon another, a sort of “self-scanning process in the brain.”

---

288 Ibid., p. 61.
290 Ibid., p. 326.
291 Ibid., p. 324.
The basic form of Armstrong’s view holds widespread assent. Lycan endorses a very similar self-scanning model of introspection.\textsuperscript{292} Hill is basically in agreement, though he thinks a self-scanning theory of introspection also needs to accommodate the phenomenon of “volume adjustment”, according to which scanning one’s own mental states can heighten the “intensity and degree of internal articulation” of a sensation.\textsuperscript{293} And Dennett seems to have once endorsed a theory of introspection very similar to Hill’s.\textsuperscript{294}

Rosenthal holds a view which is similar to the self-scanning view, except that he disowns the analogy between introspection and perception on the grounds that, Perceiving something involves the occurrence of some sensory quality, which in standard circumstances signals the presence of that thing. If our being conscious of a mental state is like perceiving something, our being conscious of it will involve the occurrence of some mental quality; otherwise the analogy with perception will be idle.\textsuperscript{295}

But in introspection, the only “mental quality” of which we are aware is that of the mental state we introspect; there is no additional sensation associated with introspecting the state.\textsuperscript{296} So Rosenthal adopts the view that introspecting is just having a higher-order thought about your mental state. In cases of introspection, this higher-order thought is a third-order thought. This is because conscious thoughts (in the sense of any appearance, be that an occurrent thought or a sensation) must have a second-order thought directed upon them: this is what makes them conscious.


\textsuperscript{293} Hill (1991) \textit{Sensations: A Defense of Type Materialism}, p. 120.

\textsuperscript{294} Dennett (1978) \textit{Brainstorms}, chapter 11.


\textsuperscript{296} Of course, inner-sense theorists have responded to this criticism; see Lycan (1996) \textit{Consciousness and Experience}, pp. 27-30.
Introspection occurs when a third-order unconscious thought is directed upon the second-order thought, thereby making the second-order thought conscious as well. Normally, the second-order conscious state would be unconscious.  

Searle too rejects the analogy with perception, on the grounds that conscious states are not like objects that we can take a distance on and inspect. He also makes the same point as Rosenthal: that there is no experience associated with introspecting as there is with perceiving. Introspection, for Searle, is just thinking about our own mental states. In introspecting our conscious states, we direct one conscious state on another. This is very similar to Rosenthal's account, except that no unconscious third-order thought is needed to make the second-order introspecting thought conscious, since Searle has different views about what makes a mental state conscious.  

However, Searle also holds a certain view about the "self-consciousness" of conscious states, which seems very close to the perceptual model of introspection:

we can always shift our attention from the object of the conscious experience to the experience itself. We can always, for example, make the move made by the impressionist painters. Impressionist painters produced a revolution in painting by shifting their attention from the object to the actual visual experience they had when they looked at the object. This is a case of self-consciousness about the character of experiences.  

This idea that we have the capacity to direct our attention upon appearances themselves, rather than what appearances are of, does not seem to be any different to McGinn's conception of introspection. Searle criticises McGinn's account for requiring a "distinction between the act of inspecting and the object inspected", but this need  

299 Ibid., p. 143. For an alternative philosophical interpretation of what the impressionists did, see footnote 331.  
300 Ibid., p. 105.
be no more than the act of paying attention, and the experience attended to, as endorsed in the quotation above. Searle seems to have read something far more weighty into the perceptual model of introspection than is in fact there.\footnote{Searle seems to think the idea is of a capacity to discover what mental states we have just by “looking” (whether I really like caviar, whether I am afraid of going on the rollercoaster, etc.). But all McGinn seems to have in mind is that for any conscious state we are occupying, we can turn our attention inwards to the state itself; the experience itself is a different sort of “object of apprehension” to the object of the state, in McGinn’s terminology. These are intuitions which Searle seems to share exactly.}

Chalmers has another similar view. He says that,

> although we have the *ability* to notice our experiences, most of the time we notice only the contents of the experience, not the experience itself. Only occasionally do we sit back and take notice of our experience of the red book; usually we just think about the book.\footnote{Chalmers (1996) *The Conscious Mind*, p. 221.}

For Chalmers, introspecting is a matter of making “second-order” judgements about our experiences, rather than making “first-order” judgements about the contents of our experiences.\footnote{Ibid., p. 176 & ff.} So instead of simply judging that an object is red on the basis of a perceptual experience (a first-order judgement), we also have the ability to notice that we are in an experiential state that is representing the object to be red, and so judge that we are instantiating an appearance of red (a second-order judgement). This position seems to combine a perceptual view, with the view that introspecting is thinking about our thoughts. For we have the ability to notice our representations themselves, rather than the contents of our representations. But introspecting involves turning our attention on the representations themselves and making judgements about them, thereby self-ascribing concepts of consciousness which are applicable to appearances, rather than what appearances are of.
Modes of Access theorists are best interpreted as taking a perceptual view of introspection, since they think we have a distinct awareness of conscious states in virtue of occupying them. They hold that the illusion of an explanatory gap arises from our possession of two modes of access to the same things, such as brain states, for which we have formed two sorts of concepts, physical concepts and phenomenal concepts. Here again, we have the idea of a distinction between awareness of the contents of representational states, and awareness of representational states themselves. Physical concepts are formed on the basis of our awareness of the contents of representational states which we occupy. In this way, we can acquire physical concepts of brain states on the basis of perceiving them; this is one mode of access to our brain states. The other way is by occupying those states, whereby we are introspectively aware of being in certain representational states, and thus form phenomenal concepts of them.

This idea of introspection, then, is the idea of an awareness of appearance attained by occupying a state that either is an appearance or has properties of appearance, as opposed to an awareness of the world engendered by appearance (perception of the world, thought about the world, etc.). Phenomenal concepts are formed on the basis of introspection, since they represent our way of thinking about states we are occupying (or have occupied, or can imagine occupying). Tye expresses this idea, using pain as his example, in the following way:

---

304 So long as a sensory element to this sort of “perception” is disavowed; consider Hill and McLaughlin (1999) “There are fewer things in reality than are dreamt of in Chalmers’s philosophy”, p. 448: “Sensory states are self-presenting states: we experience them, but we do not have sensory experiences of them. We experience them simply in virtue of being in them.” Sturgeon disavows any perceptual analogy for introspection (Sturgeon (2000) Matters of Mind, p. 47). This is on the grounds that the epistemological peculiarities governing the application of concepts of appearance (“Anything so conceived [through concepts of qualia] will be a domain in which appearance is reality.” Ibid., p. 49) provide an adequate disanalogy with the application of concepts on the basis of perception.
Fully understanding pain requires grasping how it feels, its distinctive phenomenal character, what it is like to undergo pain. That, in turn, requires applying to pain the concept that is typically applied when people introspect pain and pay attention to what it is like subjectively. This concept is a phenomenal concept.\textsuperscript{305}

Phenomenal concepts, then, are introspection-based concepts, applicable to appearances on the basis of being subject to such a state of appearance.

We have now briefly worked through a number of different models of introspection. They are representative of current thinking on introspection. None of them seriously challenges the idea that by introspecting, we fix our attention on properties of appearance itself, rather than what appearances are of. Whether this amounts to a special awareness we have of conscious states in virtue of occupying them (which could be some sort of self-scanning process in the brain, or the exercising of a special faculty), or whether it is the making of a judgement about the conscious states we are occupying, or the entertaining of a conscious thought about the conscious state that we are occupying; whatever it amounts to, the basic thought remains the same. Perhaps it would be said that paying attention to conscious states themselves, rather than their contents, is just what "introspection" means.

This natural propensity to separate appearance and reality, by distinguishing between our awareness of representational states and our awareness of their contents, has been a part of physicalism since its outset. Consider Smart’s rendering of having a yellowish-orange after-image as having something go on which is like what goes on when seeing an orange in normal conditions. The "something going on" is the conscious state or appearance which we know in introspection, and an orange is an example of the sort of reality which similar states inform us of. The same appearance /

reality distinction is presupposed when we say, in a more contemporary idiom, that conscious states and properties are just what our phenomenal concepts are concepts of, and these concepts are formed on the basis of introspecting. It is a natural way of thinking. But to deal with the problems we have encountered in previous chapters, we need to challenge it.

This is exactly what we shall do in the next section, Section 9.c., by presenting some considerations which have a direct bearing on the idea of consciousness as appearance. The considerations are familiar, but the inference we shall draw is not: an ability to introspect appearances is not thereby an ability to introspect conscious states. If we cannot introspect conscious states, except in a sense far removed from the intentions of any of the philosophers we have discussed in this present section, then this suggests that the philosophical conception of appearance which has generated the explanatory gap debate and turned the cycle of thought we traced in the last chapter, is a misconception. Of course, even if we cannot introspect conscious states, that does not in itself show that we cannot introspect properties of conscious states. However, we shall argue this case as well.

9.c. Five Considerations

Our first consideration is that consciousness always has a selective focus. Experience is never a unified array that can be picked out as “that”, for it is irreducibly gestalt. There may be any number of different features in my environment at any one time that I could be directing my consciousness upon: a dusty smell in the air, the warmth of the sun, a tinny stereo playing dance music, a coldsore on my tongue, some
people in the street, the trees and houses etc. But I am not aware of them all at any one moment. Rather, my attention flicks between different objects of attention. Focusing can be deliberate or non-deliberate. I might deliberately focus my attention on the music, but if something obtrusive and untoward occurred, my attention would normally be drawn away.

When philosophers directly confront intuitions of an explanatory gap, perhaps asking themselves “how could this be a state of my brain?”, they are always deliberately focusing on something or another, such as the feeling of the coldsore or the warmth of the sun. This focusing cannot be the same focusing as is normally involved in consciousness, because in introspection, we are supposed to be aware of our experiences themselves, rather than just the contents of our experiences. So perhaps the idea is that in introspection, we focus on a conscious state, which is itself focused upon some specific part of the world.

Now we have already said that consciousness has a selective focus. In that case, it does not seem contrived to say that every state of consciousness involves a focusing upon one thing or another. But the focusing itself does not appear. When we introspect our experiences, we do not find ourselves focusing on something; rather, we find what we are focusing on. If this is right, then since the focusing does not appear, but is a constituent part of a conscious state, that means that conscious states are not states of appearance. States of appearance are what they appear to be, and there is more to a conscious state than what appears. However, this point does not

---

306 My own view, as shall be made clear later on, is that it is indeed exactly the same focusing.
307 Could we perhaps say that the focusing is not a part of the conscious state? This seems unmotivated, since we can deliberately focus on something, and be quite aware that we are doing so.
challenge the view that appearances are properties of conscious states, such as the
property of how the thing focused on appears.

The second consideration is that focusing on an appearance in introspection
presupposes thinking about it as an appearance, rather than as what it is an appearance
of. This accords with a general feature of consciousness, that we always subsume the
objects of consciousness under some aspect or another. This is because consciousness
is representational, and always represents its objects in some way or another. If we
were not thinking of the conscious states that we introspect as conscious states, then
we would simply not be introspecting. Conscious states provide an awareness of the
world, so if we were not specifically thinking of a conscious state as a conscious state,
we would be aware only of the world; we would be perceiving or thinking, rather than
introspecting.

Now from the fact that we always think about things under some aspect or
another, it does not follow that the aspect is one which really is applicable. I can think
of an apple as a pear, or of glass as a solid, and the Empress can think of her horse as
the re-incarnation of her late husband. So when we introspect, and focus on a
conscious state which presents some part of the world to us under some aspect, for
example an experience of a pear as a pear, and then we concurrently think of what we
are focusing on as a conscious state, it is possible that we are wrong. What we are
focusing on might not be a conscious state at all.

The third consideration is that when we introspect a conscious state, we do not
thereby become aware of the subject of the state. In introspection, we become aware
of appearances. But any appearance is necessarily an appearance to a subject. When
subjects pick out an appearance as "this", their attention is not turned towards
themselves. All they find is the appearance.

Certainly, self-consciousness is implicated in the act of introspecting. The act
involves focusing on a particular appearance, and thinking about it as an appearance,
and this presupposes a subject who focuses, and who thinks about what they focus on
as something mental, as an appearance rather than as what the appearance is an
appearance of. But typically, the subject does not appear. Hume found the subject
systematically elusive, but it is not our purpose to rule out introspective awareness of
the subject. Our point is only that this is not typically the case, and there is certainly no
awareness of the subject when conducting the sort of thought experiments that tend to
preface books on consciousness. When introspecting an appearance such as the taste
of coriander, the subject remains anonymous as an object of attention, although the
appearance could not exist anonymously. But if conscious states are intentional, they
are appearances to a subject. So they are not states of appearance.

The human mind unifies disparate experiences into a coherent consciousness of
the world. Our thoughts and perceptions are experienced as connected, a
connectedness which allows us to take up perspectives upon the world, and to plan
and engage with projects within it by forming conceptions of ourselves. Appearances
do not, as a rule, take us by surprise; we experience them as related to each other, and
this is because they all refer back to the same subject, a subject that has concepts of

---

308 F.H. Bradley made a similar point, though for entirely dissimilar reasons. The relevant passages
are in Bradley (1893) Appearance and Reality, pp. 103-126 & pp. 223-240. He says such things as:
"The actual subject is never, in any state of mind, brought before itself as an object ..... [it] never feels
that it is all out there in its object, that there is nothing more left within." (p. 111); "We may confuse
the feeling which we study with the feeling which we are. Attempting, so far as we can, to make an
object of some (past) psychical whole, we may unawares seek there every feature which we now are
and feel." (p. 232); "by its essence the "this" is self-transcendent" (p. 233). Thanks to J.J. Valberg for
the reference to Bradley.
what the appearances are of. It is because of the mind’s function of imposing unity upon experience, by tying various appearances together and bestowing significance upon them, and that we speak of “subjects”.

What exactly a subject is, is another question. Philosophers once thought of them as akin to things, substantial egos, but common contemporary views are that subjects are just human beings, or that the idea of a subject is of a logical constant to which experiences are referred. To the point in hand, it makes no difference; however we understand the notion of a subject, it is clearly not something that we always find when we introspect our experiences. We find what the conscious state refers back to the subject, not an object being referred back to a subject.

In short: intentional states present objects to subjects. Conscious states are intentional states. In introspection, we are aware only of the object of an intentional state. So we are not aware of the intentional state itself, and consequently, we are not aware of a conscious state. What we are aware of in introspection is appearance (what appears to us). So consciousness is not appearance.

This time the point goes through for properties as well; appearances cannot be conscious properties. If in introspection we are aware of conscious properties, then our attention must be fixed on properties of our experiences, rather than properties represented by our experiences. So the property focused upon must be an experiencing of a property in the world. But to count as an experiencing of a property in the world, a conscious property must be a presentation of a property in the world to a subject. We do not find these properties in introspection.

If we could introspect conscious properties, then the only difference between experiencing properties in the world, and introspecting the conscious properties which
make us aware of these properties in the world, would be a difference in intention. If the redness of my experience of a red book is a property of my conscious state which I can introspect, then it cannot also be a property of the red book. It must be rather that most of the time, I think of and experience the redness as a property of an external object, and intend to refer to the objective red which my experience of red represents, but sometimes (in introspection) I intend the property of the conscious state itself.\footnote{The alternative, that I can “flick” my attention at will between two different reds, phenomenal red and objective red, is absurd.}

This way of looking at conscious experience is reminiscent of sense-data theories, since experientially familiar properties like redness become properties of conscious states. The counterintuitive nature of this way of looking at experience is hardly dissipated by insisting that seeing objective red simply consists in instantiating the conscious property of redness (perhaps adding that the instantiation be the upshot of the right sort of causal chain). It still remains that perceiving and introspecting red are both a matter of focusing on a property of a conscious state, the difference being whether your intention is to focus on the objective property of red (which the property represents), or the conscious property itself.

This picture is not obligatory. We avoid it altogether by insisting on something we would naturally suppose in any case: that consciousness presents us with properties in the world (rather than with properties of itself which represent properties in the world). Conscious properties could only be presentations of external properties to the conscious subject (or properties of these presentations). They cannot be properties which are presented (or self-presented). Consequently, they do not appear to us, are not properties of appearance, and so are not introspectable. All conscious properties necessarily incorporate a subject which is aware of the experienced property; without
the subject, the property of the experience cannot count as an experiencing of a property in the world. If we leave out the subject, then we have only the content.

We can now draw together these three considerations, and come to some conclusions.

We started out by noting that in consciousness, we are always focusing on something or another. Introspection is supposed to be a form of consciousness, consciousness of our own consciousness, so it provides no exception. But the focusing involved in introspecting is not supposed to be the same focusing as is normally involved in consciousness, because in introspection, we are supposed to be aware of our experiences themselves, rather than just the contents of our experiences. All the same, there does not seem to be anything else which we can focus upon, except for the contents of our experiences.

There is a simple reason for this, which is that there is nothing else there for us to focus on, nothing except for the world which consciousness presents us with. What happens when we think that we are introspecting a conscious state is that we focus on the same thing that we were focusing on before (there is only one act of focusing), but think that we are focusing on something different: the conscious state itself. What changes is that we now think about what we are focusing on as a conscious state or property, i.e. as an appearance.

Now the second consideration we brought up was that consciousness always subsumes its objects under an aspect, but that the aspect is not necessarily appropriate to the object. This is just the sort of mistake which occurs when philosophers think they are introspecting consciousness. Whatever it is that they are focusing on, such as the redness of a book, or the taste of coriander, is thought of "in introspection" as an
appearance or property of consciousness. But it is not: it is the same property of a
book, or property of some coriander, that they were focusing on to start with.

The third consideration draws the previous two together by showing why the
idea of an introspective grasp of consciousness must be a mistake. This is because
consciousness is intentional, and intentional states present objects to subjects. This
means that all that can appear to the subject are the objects of consciousness.
Conscious states are presentations of objects to the subject; these presentations
themselves do not appear to the subject. We cannot direct higher-order conscious
states upon other conscious states in the way that has been supposed, for conscious
states have structural aspects which do not appear: they involve a focusing on a
particular aspect of the environment, and present that aspect to a subject as part of a
unified and coherent experience. And there are doubtless other structural aspects that a
more detailed analysis would bring out.

If we direct a “higher-order” conscious state (it is really just a conscious state,
or occurrent thought) onto an appearance, then we have failed to direct it upon a
conscious state; we are just thinking about what consciousness presents to us as
something that it is not. We can of course entertain conscious thoughts about
conscious thoughts (in order to engage in the philosophy of consciousness), but if we
end up thinking about appearances, we have gone wrong (the present critique
excepted, of course).\(^\text{310}\) What we should really be thinking about must include the sort
of structural properties aforementioned; we should be thinking about how
consciousness must be given how the world is presented to us.

\(^\text{310}\) Although Rosenthal’s theory of introspection looks \textit{prima facie} to just be conscious thought about
conscious thought, he must be thinking about directing conscious thoughts upon appearances. If he
were not, then the third-order thought would not be needed to make us conscious of the first order
state itself, rather than what that state is of.
We have adopted a very sceptical attitude to introspection. But the word “introspection” has an ordinary use which is perfectly legitimate, and which has only a tangential connection with the notion of introspection employed in philosophy. For a start, introspection has not got much to do with consciousness; when we introspect, we are usually concerned with our dispositional mental states. Typical cases of introspecting involve trying to find out what you really think about somebody, often yourself, or about something. It is more akin to self-analysis, to working things out, rather than “self-scanning”, or using a “second-order” state to find out what the “first-order” one is.

So, for example, you might wonder whether you really do hate Henrietta, or whether you really believe in ghosts, or whether you really do like conceptual art. Introspecting to find out might involve remembering how you reacted when Henrietta last made conversation, and how you would react if similar circumstances arose again. Or you may try to work out what exactly is wrong with Henrietta, and decide whether this is worthy of hate. Or introspecting might involve performing a thought experiment: if the stakes were high, would I bet on ghosts existing? Or it might involve considering the supposed merits of conceptual art, and deciding whether or not you believe that the works you have seen displayed these merits.

Sometimes you do introspect conscious states. For instance, you might go to the doctors to report a feeling in your chest. The doctor asks whether it hurts. You might then introspect, by focusing on the feeling, and asking yourself whether this feeling really constitutes pain, perhaps by comparing it to how you remember some past paradigm of pain. But this does not involve making the state of pain the object of your attention (in virtue of scanning it, or directing a thought upon it). The sensation in
your chest which the state presents to you is still the object of your attention, and your thoughts are concerned with how best to describe it.

In short, introspection is the process of working out what your mental states are, or must be. We do not employ an awareness of mental states, and we do not direct our thoughts upon states we are currently occupying to find out what they are. The nearest we ever come to this is when we focus on what a state presents to us, and try to ascertain what state we must be in to receive such a presentation. From now on, we shall have nothing more to say about real introspection, and shall reserve our use of the word for the notion employed in the explanatory gap debate.

Consciousness is always somebody's consciousness. Consciousness presents us with the world to demonstratively pick out, but does not present us with ourselves representing the world, and focusing on a part of the world. Any state of consciousness must include a subject, just as any conscious event or conscious property must include a subject. This is because conscious states, events and properties, are intentional states, events, and properties. For this reason, there can be no true exclusively introspective conception of consciousness.

However, despite what we have said, the intuition will remain that “consciousness” just means appearance, or things appearing in a certain way. There may be more to consciousness than the appearances on the surface, in might be said, but nevertheless, it is the surface that we have in mind when talking about consciousness. The problem, so goes this objection, should not be defined away.

But this is not our intention. The tradition is mistaken in conceiving of consciousness as appearance. It is simply not true that this is the only way we have of conceiving consciousness, and we shall now present two further considerations which
illustrate this fact rather more directly than the previous ones. A conception of
consciousness as appearance is not simply a datum to be accommodated, since
common-sense thinking has room for conceptual distinctions between consciousness
and appearance.

Our next consideration concerns the old thought experiment everybody has
tried at some point, in which you try to imagine body swapping with somebody. You
imagine taking over somebody else’s body, a situation in which you suddenly look out
of the other person’s eyes. The thought experiment is very much a part of the popular
imagination, and has been a plot device in various books and films.\(^3\) It demonstrates a
natural distinction that can be made between the concept of consciousness and the
concept of appearance.

There are two ways of running the thought experiment, depending on whether
you take an active or a passive role after body-swapping. Here is an example of the
first scenario. I press the button on my desk to body-swap with John, and suddenly I
see the inside of the pub where John is having a drink. I can see out of his eyes: what
he was seeing at the moment I body-swapped with him. I have a pint in my hand, and a
tattoo on my forearm: it is John’s arm. John’s friend asks me a question, and I answer,
trying hard to say the sorts of things that John might say. As I talk, I hear my words
cloaked in John’s voice.

The alternative way of conceiving body-swapping is more passive; I enter into
John’s body and see and hear what he sees and hears, but I have no control over his
body. I listen to what he says (hearing the way John’s voice sounds to John), and

\(^3\) A recent example is the film “Being John Malkovich”, in which a business is set up selling the
opportunity of taking over Malkovich’s body for a short period.
observe the passing scene as he walks around and does things. I might even be aware of his occurrent thoughts as well.\(^{312}\)

Whichever way we run this thought experiment, it involves only acquiring access to another person’s *appearances*, or how the world appears to another person. On pressing the button, the world appears to me from the perspective that it usually appears to John. The difference between the two versions is just whether or not you are able to actively change what appears to you (by walking around, saying things), as you can when you “occupy” your own body.

The thought experiment does not involve taking over another person’s *consciousness*. Such an idea is not even coherent. It is hard to see what it could even mean to say that on pressing the button I take over John’s consciousness. Presumably, pressing the button would bring about my death, since my consciousness would be no more. But what sort of fact could make it the case that *I* now had John’s consciousness? To “have” John’s consciousness is to be John. The idea of swapping consciousness does not make sense. But taking over somebody’s appearances does make sense. So there is conceptual distance between consciousness and appearance; consciousness does not just mean appearance.

Our final consideration concerns the location of sensations. We ordinarily talk about sensations being located within parts of our bodies. For instance, I might have a pain in the finger, and if asked where the pain is, I could hold up my finger, or point to it. However, on reflection, I would obviously not want to say that the conscious property, the sensation itself, is located in the finger, like some ethereal vapour. But equally obviously, the appearance of the pain *is* in the finger, otherwise I would not

\(^{312}\) This ambiguity in the thought-experiment is exploited in the above mentioned film, in which the central character wrestles (with Malkovich) to assume active control over Malkovich’s body.
call it a “pain in the finger”. So since we want to say different things about the location of the appearance and the location of the conscious state, there must be some separation between our concepts of appearance and consciousness. The appearance is in your finger. The conscious state or property is not. So consciousness is not appearance.

All that we are doing here, is taking the idea that there is no appearance / reality distinction for consciousness seriously, treating the idea that conscious states are what they seem to be consistently. Philosophers particularly attached to this idea are rarely so consistent. Consider for instance Searle’s remark that,

In a sense all of our bodily sensations are phantom body experiences, because the match between where the sensation seems to be and the actual physical body is entirely created in the brain.\(^{313}\)

However elsewhere he says quite categorically that,

Where appearance is concerned we cannot make the appearance / reality distinction because the appearance is the reality.\(^{314}\) [Searle’s italics]

The first quotation says that sensations seem to be in parts of our bodies, but they are not really, since the brain creates a kind of illusion. The second says that sensations are what they seem to be. But either they are what they seem to be (sensations located in parts of our bodies), or they are not. And the dodge that are seemingly located pains will not work, because they do not appear to be seemingly located pains: they appear in a certain location, they do not appear to appear in a


certain location. Difficulties and inconsistencies such as these demonstrate the foibles of trying to maintain the rationalistic doctrine that some things are what they appear to be, whilst also trying to uphold a scientific world-view.

9.d. Human Representation

The fundamental error which has generated the explanatory gap debate, and with it the opposing camps of inflationism and deflationism, is the assumption that we experience consciousness. But we do not experience consciousness. We experience the world in virtue of being conscious. Human beings are incapable of switching their attention from what their experience is of, to the experience itself. Our representations are not made available to us when we occupy the conscious states that are those representations. The representations themselves are only something we can think about, and devise theories of.

The error, which generates a conception of consciousness as appearance, is to look at the world and expect to find consciousness there. But consciousness is just that which allows us to look at the world. Consciousness does not make itself transparent to the reflecting subject, or reveal any of its properties. All it reveals is the world as represented by the subject. And since it is just vacuously true that we can only experience the world as we experience it, there should be no temptation to try to reify the “world as we experience it” into a special object of philosophical inquiry.

We can of course think about consciousness to contemplate its identity with something physical or functional. But this is not introspection. Consciousness is a theoretical notion, inseparable from the idea that an experiencing of the world should
be distinguished from the world which is experienced. To think of consciousness is to think of an experiencing of the world in which a subject is presented with the world appearing in some way or another. If we start thinking about a visual experience of a painting, or the taste of coriander, then we have gone wrong: such appearances of parts of the world are only the acid test for the presence of consciousness. The state of consciousness should not seem like anything to the subject thinking about it, though if the subject were in that state, or is able to simulate that state, then the world would seem a certain way.

The theoretical nature of consciousness does not jeopardise the indubitability of its existence: we know, of course, that there is consciousness, for otherwise we would never experience paintings and coriander. My experiencing of a painting is a mental state, and I can of course intend to refer to that experiencing, just as I can intend to refer to a brain state. But the experiencing itself could be just about anything for all I can tell from reflection alone. I can intend to refer to it, and succeed. But I cannot pick it out in introspection. In thinking about a conscious state, I have no associated images, as I do in thinking about a brain state; consciousness itself does not seem like anything as an object of thought. So, as we said, it could be just about anything. Therefore there are no grand conceptual obstacles to physicalism about the mind; introspective conceptions of consciousness are fallacious.

Right at the outset of contemporary physicalism, Smart insisted that he did not want to identify an after-image (i.e. an appearance) with a brain process, but rather wanted to identify an experiencing of an after-image (i.e. a conscious state) with a brain process. He was right that physicalism should eschew anything like an act-object account for mental phenomena like after-images. Lycan is an unusual participant to the
explanatory gap debate, in that he still makes this point, often stressing the tendency inflationists have of "slipping tacitly from talk of mental states and events to what amounts to talk of phenomenal individuals".\(^315\)

However, a conception of appearance survived this fundamental physicalist point, when the emphasis switched from experiencings themselves, to the properties of experiencings, such as the particular redness of an after-image.\(^316\) Even if an object experienced in perception is a part of the objective world, and it is only the experiencing of that object which is to be identified with a brain state, there remained the problem of accounting for "what it is like" to be experiencing that object. In introspection, it was thought, we can form a conception of "what it is like" by focusing on the conscious, subjective properties of our experiencings. Thus the problem of accounting for appearance was renewed. Even Lycan is aiming to reduce what he finds when he "monitors" his consciousness in introspection. But on our view, we are never aware of conscious states or any of their properties, not even their "surface properties".

The theory of consciousness we are defending is a form of intentionalism, strong intentionalism in fact, since consciousness is exhausted by its intentionality. But it is distinguished from any of the current physicalist strong intentionalist theories by rejecting introspective access to consciousness as such, and thus any conception of consciousness as appearance. Appearances are not representational states. They are not any states, properties, or things at all. The world appears to conscious subjects.

To see the contrast, consider Tye's strong intentionalism. Tye quite rightly says that to have a visual experience of red, is to be in an intentional state that represents


\(^{316\) This switch corresponds to Stage 3 of the cycle of thought we traced in Chapter 8.
some distal object as being red. But to him, this means that "what it is like" to experience red is a property of representing the object as red, which makes it amenable to physicalism:

the "what it is like" aspects of phenomenal states are second-order, broadly physical properties that are realized by objective, first-order, physical properties of those states.\(^{317}\)

We know "what it is like" to see red, because we can introspect and apply phenomenal concepts to our representational states.

Thus Tye supposes that we can direct our attention on properties of our representations in introspection. This means that he is still attached to a conception of consciousness as appearance. His intentionalism exists within that framework. He thinks that appearances can be reduced: they are properties of representational states. Consequently, he is concerned by the explanatory gap problem, the problem of connecting our introspective concepts of representational states with our theoretical concepts of them, and adopts a version of MAR.

Our approach is quite different. To have a visual experience of red, is to be in an intentional state that represents some distal object as being red. But the redness is not a property of the representation; it is a property of the distal object. We are not aware of objective red in virtue of being aware of our representation of objective red, and we cannot in introspection turn our attention on the representation itself. Rather, we are aware of objective red in virtue of being in a state that represents objective red, a state which we cannot in principle make an object of awareness. To talk of "what it is

like” to experience red is simply a misleading reference to objective red as experienced. To know “what it is like” to see red is to possess a recognitional concept of red.

Phenomenal concepts are not concepts of properties of consciousness. They are not even concepts of representational states that we apply in virtue of being in them. Rather, they are ordinary recognitional concepts of properties in the world. The concept *redness* and the concept *pain*, belong to the same category as the concept *water*. They are all ordinary recognitional concepts that refer by connoting functional descriptions of their referents. They are formed on the basis of how the world is presented to us in consciousness, and contrast with theoretical (physical) concepts which are formed on the basis of our theories about how the underlying structure of the world must be.

The concept of redness is a recognitional concept of a property of certain objects in the world; we can describe the property by mentioning typical things which we think have this property, loosely carving out the causal role it plays. However, there might not really be any objective property in common to all these things; in the case of redness, there apparently is not. But the concept is only a recognitional one.

We could make the same mistake with the concept *water*, by lumping H₂O and XYZ together. The mistake to avoid is that of jumping to the conclusion that the apparent unity is explained by all the different properties causing us to experience the single property phenomenal red. This would be like saying that H₂O and XYZ both cause us to experience phenomenal water.

---

318 Georges Rey (1995) “Towards a Projectivist Account of Conscious Experience”, p. 137: “Many of us are deeply disappointed to learn that the redness (that appears) in rainbows, grease spots, tomatoes and roses is no single “natural” property, but an enormously complex disjunction of such properties, unified only by the fact that they have a certain effect upon (some of) us.”
All we need say is that there are a variety of properties in the world that we cannot tell the difference between, our recognitional concept red being insensitive to the distinctions. The alternative, that our concept of red is of phenomenal red, lands us with the strange position that we have no recognitional concept of these various properties of objects in the world, only theoretical concepts. For we have no other recognitional way of thinking about these properties.319

Levine has argued that phenomenal concepts have a “thick”, substantive mode of presentation. We have, he says, “a fairly determinate conception of what it is for an experience to be reddish”, in contrast to the MAR view that we are simply labelling some “we know not what”.320

The reason for this, on our view, is that such concepts are recognitional concepts of properties in the world. They are as determinate as any such concepts: a way of thinking about something we recognise. The reason that it is so often supposed that they are not “thick” concepts is that once the assumption has been made that the property in question is phenomenal, an appearance, then we have precluded anything that might have been said of it. We are no longer talking about a property of an object, but a property caused by experiencing the property of the object, a property which is what it appears to be.

The idea of consciousness as appearance, as something we can introspect, which pervades even intentionalist accounts, displays a misunderstanding of human representation. This comes out time and again when it is assumed that we can isolate

319 Phenomenal concepts, like all recognitional concepts, can be developed. Just as you can learn to recognise the difference between elm and beech trees, so can you learn to recognise an almond taste in your favourite wine (recall the example of Mary in Section 7.e.) The properties (of the trees and the wine) do not change, but you can develop more sophisticated concepts of those properties.

properties of our representations in introspection, properties that we are aware of (such as "what it is like" to be experiencing something).

Television sets represent. They represent in such a way that we can isolate some patch on the screen, and say what it represents. For instance, we could run our fingers around a section of the screen, and say "that represents a house". We could also isolate a section of the screen, and say "that represents red". But this is nothing like human representation. We make a mistake if we focus on a patch of red "in introspection" and think to ourselves "that represents red (i.e. some objective property in the world)". This is because the representational states of consciousness, unlike those of a TV set, are presentations of things in the world to a thinking subject, who is deliberately focusing on something and entertaining a concurrent thought about it. Any property you can pick out is a property presented to a subject. Consequently, it cannot be a conscious property, for what makes a property conscious is that it is a presentation of an object or property to a subject. The things and properties we have access to in consciousness do not represent anything, unless they happen to actually be representations, such as images on a TV set.

IMAX cinemas are much better at representing the world than televisions, but they still represent in the same sort of way. Even if the technology advanced to the point where we could not tell whether we were looking at an IMAX film, or an actual scene in the world, it would still be the case that we were looking at something (a screen through goggles) that represented something else. In consciousness, we have access only to what is represented; the representations themselves stay out of reach in principle.
This understanding of intentionality recalls the conception Sartre and Merleau-Ponty had of it, as the property consciousness has of putting us in contact with the world, but not itself. Consciousness leaves itself behind and presents something else, like a hand reaching out and touching something; but whereas a hand can always become the thing touched, consciousness itself can never be phenomenologically presented as an object.\textsuperscript{321} A similar point is made in the development of J.J. Valberg's "horizontal conception of experience", or consciousness. Consciousness, for Valberg, is "that within which the world is present". It is the horizon within which phenomena of the world turn up, and consequently could never itself turn up within the world: "experience has no character of its own".\textsuperscript{322}

Unlike the philosophers mentioned above, we shall take the point as an exclusively phenomenological one, without ontological import.\textsuperscript{323} We are not trying to draw up an ontological conception of consciousness; nothing we have said should lead us to deny that consciousness supervenes on the brain, for instance. Neither will we rule out physicalistic reductions of intentionality such as covariation or teleosemantics; so long as they can accommodate the fact that human representations have the structure of a subject being presented with a world.

\textsuperscript{321} For Sartre, since consciousness can never be made an intentional object, to understand its unity with the body, we must imagine the body in its purely explorative role as a subject as well. The difficulties we feel in trying to conceive the unity of mind and body, "all stem from the fact that I try to unite my consciousness not with my body but with the body of others": Sartre (1958) \textit{Being and Nothingness}, p. 303. The same idea occurs in Marcel: "the moment I treat my body as an object of scientific knowledge, I banish myself to infinity." Marcel (1949) \textit{Being and Having}, p. 12. Merleau-Ponty discusses "double-sensations", the phenomenology of our hands alternating the roles of subject and object, in his (1962) \textit{Phenomenology of Perception}, pp. 93-4.

\textsuperscript{322} Valberg (1992) \textit{The Puzzle of Experience}, chapter 6, esp. pp. 124-5.

\textsuperscript{323} For instance, Merleau-Ponty thinks that since our eyes are a "means of knowledge" of the world, the "objective thought" of science and common-sense makes a mistake in construing them as "bits of matter" (Merleau-Ponty, op. cit., p. 70). And Valberg thinks that since our conception of consciousness is of "that within which the world appears", this rules out the possibility of its being something within the world. He says, for instance: "Try to juxtapose these things in their stark disparity: the horizon within which the objects around you are present, and something you conceive of as occurring inside your head" (Valberg, op. cit., p. 140).
However, the phenomenological point we have endorsed bears directly on the idea that there are *explanatory* problems for physicalism about consciousness. The explanatory gap problem arises because of our supposed direct awareness of consciousness. But since we have none, there can be no requirement of making conceptual connections between introspective concepts and theoretical concepts: there is nothing to explain on that front.

9.e. The Obsoletion of Appearance

We should take the idea of “appearance as something real” for the oxymoron that it is. A great part of the obstinacy of appearance is the apparently irrepressible intuition that appearance must *be* something. Thus it is blithely contrasted with reality, when the natural contrast to something real is something not real. Appearances are not things or properties, rather things and properties appear to us. Nothing is what it appears to be, nothing *is* an appearance.

Appearance is a necessary condition for consciousness, for the reason that consciousness is intentional. In order to be in a conscious state, something must be appearing to the subject. Consequently, there will always be “something it is like” to be in a conscious state, because something must be presented to the subject, and the subject must always be able to tell that something is being presented. Sometimes, we represent things without being able to tell that we are doing so: these states are unconscious. It is a scientific question to determine the difference (functional, perhaps) between states in which we can tell we are representing something, and those in which we cannot.
To speak of things appearing to us in consciousness should register nothing more than the fact that we can tell that we are representing something, and that the thing in question may not be there at all; our conscious state might be misrepresenting the world. This is the kernel of truth behind the confused web of epistemological and metaphysical ideas associated with a concept of appearance, and belonging to an obsolete explanatory system that physicalism must leave behind.

There is an appearance/reality distinction to be made for everything, or at least for everything that can appear to us (perhaps not neutrinos). Nothing reveals its essence to unaided perception or introspection. So no recognitional concept could essentially conceive an essence. Physicalism accords the authority to describe essences to only our best theories. Observation is the springboard from which underlying structures are investigated. In the case of consciousness, there is literally nothing to observe. But we can reason about it from what we can observe. The concept of consciousness is, and always has been, theoretical; it is not something we find in the world, but something that we have realised must be there to make sense of our finding things in the world.

Of course, sometimes there is no object, but it seems as if there is, as in hallucination or sciatic pain. But physicalism has the resources to deal with these cases without resorting to the idea that we are aware of appearances. We can at least allude to some intentionalist approaches to these cases. For instance, in hallucination no object is present, but the subject thinks that one is. In this case, the subject misrepresents the world; there is nothing there. The subject's inability to tell the difference is perhaps because they have been put into a conscious state resembling those in which the object which appears to be present is actually present. The notion of
one conscious state resembling another might be cashed out in terms of a physical or functional similarity between the states, accounting for our inability to tell them apart, or at least for our judgements of similarity.

In pain, we are conscious of disturbances within our bodies. There is great temptation to think that the concept of pain is of the feeling itself (something introspectable), rather than the bodily damage: perhaps it is the concept of “what it is like” to be in such an intentional state. But the feeling itself is just the bodily damage being presented to the subject; it is an experiencing of bodily damage. Consequently, the concept of pain (and of other sensations) is best construed as a recognitional concept of properties in our own bodies. With sciatic pain, either some abnormal bodily happening is the object of the state, or the subject has been put into a state which resembles one in which there is a bodily disturbance.

In general, it seems to be a good hypothesis to treat our capacity for bodily sensations as a perceptual modality for making us aware of our own bodies. Sensations do, after all, provide recognitional access to the internal states of our bodies, just as perception provides recognitional access to other parts of the environment. There should be no more temptation to think that something else is needed (a “what it is like” property) to account for a perception of bodily damage, than there is to account for perception of an object like a tree. We could perhaps visually perceive bodily damage (through some sort of scanner), but we can also both look at and touch a tree. Phenomenal concepts are restricted to one sense modality, but so are some other recognitional concepts, like our recognitional concept of a rainbow.

Moods may also involve a perception of a state of our body (something chemical, perhaps). But there must also, generally, be a dispositional component; a
large part of what we mean in talking about being in a certain mood is that we are disposed to think certain thoughts about ourselves, other people, or the world in general, of either a negative or positive character, depending upon our mood.

Mental images and occurrent thoughts may involve active simulation of another conscious state. So picturing a red square may be something like putting yourself into a state resembling past states in which red squares were present. Afterimages may be similar, except that we are passively put into these states, and the perceptual state which is resembled was occupied only moments ago. Occurrent thoughts, our silent soliloquies, may involve passing through states that resemble those you would be in whilst hearing yourself talk.

This is all very sketchy of course, for it is not our project (which is just to analyse and resolve the explanatory gap debate). But a credible intentionalist analysis running the whole gamut of mental phenomena might well be possible. And it will not run into any conceptual difficulties so long as it is remembered that in an intentional state, all you are aware of is the object of that state, i.e. what you experience, so you cannot introspect the state since you have no experience of it: it is your experience of something else.

Let us return to the crucial step which leads to intuitions of an explanatory gap, and thereby the various tactics employed by inflationism and deflationism. It is, as we have seen, to think that in our experiences of colours, tastes, feelings etc., we can in some way, in special moods, be aware of consciousness.

---

324 The proposals about what is being represented in any particular conscious state would have to be led by scientific results. The content attributions made by the analyses would have to be plausible given, for example, what is known to be going on in the subject's body at the time they report the mental phenomenon under analysis.
In its most plausible, strong intentionalist incarnation, the idea is that though objects and properties in the world are the objects of our conscious states (there are no other, mental objects, like sense-data), we can in introspection focus on the properties of the vehicle of representation, rather than the object. Or even if this view is disavowed, it might still be said that we can focus on the “what it is like” aspects of being in a certain representational states, or its “intrinsic character”; these are supposed to be conscious properties.

But no matter how far such views are pushed to secure unmediated contact with the world, human representation has been misunderstood so long as it is thought that we can be aware of any of the properties of consciousness. All conscious properties are presentations of objects to a subject that is focusing on that object. The subject of consciousness is never in a position to pick out a property of consciousness, only a property of the world.

This mistake leads to (or perhaps results from) a conception of consciousness as appearance. But I think there is another mistake endemic to the thought experiments which philosophers of consciousness conduct, and which present them with the vivid intuition that the idea of consciousness being physical is something beyond our comprehension (or at least that it seems that way because of an “illusion”).

When contemplating the explanatory gap, it is natural to imagine the physical world from an “objective” point of view, i.e. you imagine fields, buildings, people etc. When you imagine the people, you imagine them as objects composed of organic matter. You imagine inside their skulls, and there is the grey matter of the brain. You imagine the inside of your own skull as well. But you do not find any consciousness.
So it seems that you can imagine the whole of the world without any consciousness in it. Physicalism, it seems, is bound to miss out consciousness.

But now consider exactly what you have imagined. You have imagined green fields, grey buildings and the grey matter of the brain. You have imagined the ordinary perceived world, the world as it is presented in consciousness. But you do not need to bother imagining this; you experience it everyday. You experience other people and do not experience their consciousness. It is just a truism that you cannot perceive other people’s consciousness, or your own when you look at a brain scan.

If the thought experiment is to establish anything, you would have to imagine the world without your own consciousness. Not a technicolour world including your own grey brain presented in your consciousness, but an exclusively objective world. But this is something we cannot imagine: it is the world as it is in itself, considered from no perspective whatsoever. Kant showed that this is something we cannot imagine, but can at most entertain negative thoughts about what it is not. And later philosophers questioned whether it is meaningful to talk about it at all. But in any case, since we cannot imagine it at all, we cannot imagine it with or without consciousness. The whole thought experiment, which provides us with a palpable sense of mystery, is flawed from start to finish.

With this mind, let us return to some of the inflationist positions that take the problem for granted and draw up the consequences (deflationist positions, as a rule, try but fail to engage with the problem).

Nagel argues that third person knowledge does not equip you to know the consciousness of other subjects, for consciousness can only be experienced from a first person perspective. Consciousness, for Nagel, is “what it is like” to be a particular
subject, and so something only the subject in question has access to. But in trying to draw attention to the subjectivity of consciousness, Nagel ended up excluding the subject from his analysis. Nagel’s “what it is like” is clearly not meant to be generic; it is always what it is like for a particular subject. But in that case, the idea of “what it is like” cannot define autonomous properties of consciousness; to be a conscious property, something must be presented to a subject.

A common deflationist complaint to Nagel’s claim that the consciousness of a bat is inaccessible to us, is that knowing all about a bat does not thereby make you a bat, and hence there is no reason you should be able to tell what it is like to be a bat just in virtue of understanding the principles of bat physiology, echolocation, and so on. And it is of course right that only bats experience the world as a bat does. But it is the bat’s perceptual world that is inaccessible to us, not its consciousness. The bat’s consciousness itself is the bat as a self-aware subject, focusing on some part of the world. Bats experience the world in a certain way because they are conscious. We, on the other hand, can only experience the perceptual world of a human being, and not that of a bat, or of course, the world as it is in itself. The inaccessibility Nagel points to is just the old Kantian point that we cannot reach the world except through our own consciousness of it.

Nagel’s conception of consciousness as “what it is like”, rather than a subject focusing on something, leads him to hope that:

---

325 This is a case where the problem purports to be about consciousness, but is really about our inability to imagine the world in any other way than from the point of view of human experience; we cannot imagine it as it is in itself, and we cannot imagine it as it is for a bat.
some day, long after we are all dead, people will be able to observe the operation of the brain and say, with true understanding, ‘That’s what the experience of tasting chocolate looks like from the outside’.

This they may well do, but their understanding will not link what they see to the taste of chocolate, which is what Nagel is supposing. Consequently, they will not understand something presently incomprehensible. Perhaps we will one day be able to see what somebody’s experience of the taste of chocolate looks like. But note that we do not need to say: looks like “from the outside”. The qualification drops out of consideration here, because someone’s experience of the taste of chocolate is not something that can be seen “from the inside” anyway, unlike the taste of chocolate. “From the inside”, the taste of chocolate is available, because that is just chocolate for a particular subject representing the world.

McGinn thinks that the mind-body problem cannot be solved because the only access we have to the mind is through introspection, and the only access we have to the brain is through perception. We have no faculty capable of straddling the nexus between the two, with which we could form a concept of the naturalistic property that explains their connection.

If consciousness were appearance, something known in introspection rather than perception, then McGinn’s position would be an attractive one. For there is no conceivable connection between “technicolour phenomenology” and the brain; appearances will not reduce, as we have seen, and so if physicalism is true, the hypothesis that our concept forming capacities are lacking looks like a good one.\(^{327}\) Even if Nagel’s “intellectual revolution” came about, it is hard to see how it could

---


\(^{327}\) Though even taken on its own terms, there are some problems McGinn's position would have to overcome. See Section 6.c.
come up with a concept that did not relate exclusively to either the perceptual or the introspective domain.

However, consciousness is not appearance, and if we do have essentially different concepts of appearance and reality, it is not because they were formed on the basis of distinct faculties of perception and introspection, but because of an inherited philosophical conception of appearance. We have no special access to consciousness; consciousness is something we can only think about, like microphysical and functional states of the brain. Nothing in fundamental physicalist ontology can be straightforwardly observed, and anything that can be picked out as "this" within consciousness, cannot be a property of consciousness. So there is no intuitive obstacle to identifying consciousness with something physical.

In a paper called "The Hidden Structure of Consciousness", McGinn argues that there must be more to consciousness than the "surface" which we introspect, just as he thinks language has a deeper logical structure than its surface grammar. This sounds like a basis for denying that appearance is consciousness. But that is not McGinn's conclusion; he concludes that the deep structure of consciousness itself is beyond us as a matter of principle, something we are cognitively closed to. But it is not; consciousness involves, at least, a subject focusing in a certain way. This is part of the structure which solves McGinn's mind-body problem.

For Searle, consciousness is a surface property of certain physical systems, just as solidity and transparency are surface properties of certain physical systems. The brain is just a biological machine, the microstructure of which causes these surface

\[^{328}\text{McGinn (1991) The Problem of Consciousness, pp. 89-125.}\]
properties. Though conscious properties are observer-independent (intrinsic), they are also subjective.

Searle’s view exhibits some characteristic traits of inflationism. There is the conviction that conscious properties are naturalistic properties like any other. There is the realisation that appearance cannot be reduced, plus a particular response to this, which in Searle’s case is to say that appearance is caused by the brain. And there is an awkward concession to the concept of appearance, which in this instance is the idea that conscious properties, though ordinary physical macroproperties akin to liquidity or solidity, and thus a part of the scientific world-view like any other, are nevertheless “ontologically subjective”. Searle explains this as follows:

the brain causes states of itself that are ontologically subjective in the sense that they only exist as experienced. A conscious state differs from, for example, a mountain or a molecule, in that it only exists as experienced by some subject.\(^\text{329}\)

However, since these properties are unique in being both observer-independent (they are not observer-dependent properties like economic or political properties), and existing only as experienced, Searle’s deflationary comparison with macroproperties like liquidity entirely breaks down.\(^\text{330}\) If it is just a property of the brain, why should it only exist as experienced?

Searle’s position results from assuming that the concept of appearance provides an unproblematic macro-description of consciousness.\(^\text{331}\) But in fact, we have

\(^{329}\) Searle (1999) “Mind, Language and Society: An Interview with John R. Searle”, p. 78. In this interview, Searle talks about conscious states, and not just properties, as possible objects of awareness.

\(^{330}\) Searle’s properties are just as unique as MAR’s concepts.

\(^{331}\) This is made particularly clear when he says that the impressionists switched their attention from the objects of their experiences to the experiences themselves (see Section 9.b.) What the impressionists really did was attempt to paint exactly what they saw, rather than imposing preconceptions upon the landscapes they painted (by filling in aspects they could not really see, etc.)
no such macro-description. In ordinary observation, we subsume the world under macro-descriptions, and these are the foundation and ultimate testing ground for our more precise descriptions of the world. The mind is the source of all macro-descriptions of the world. But the mind does not provide us with a macro-description of itself. It is not a function of the source of all macro-description to subsume itself, that source, under a macro-description. Consciousness gives us the world, not itself.

Levine thinks that there is an explanatory gap in our understanding of consciousness, because we cannot deduce the realisation of conscious properties from the microphysical structure of the human brain. If we are given a microphysical story about $H_2O$ molecules, kinetic energy, surface tension etc., along with some bridge laws, then we can deduce that water will boil at a certain temperature. But given a microphysical / functional story about pyramidal cell activity, neural distribution, and even objects in the environment that the subject is interacting with, we still cannot deduce that chocolate will appear a certain way to the tasting subject.

Levine's conception of an explanatory gap is independently important for the light it sheds on our models of reduction, explanation and conceivability. It also has the merit of putting the problem of conceptual difference concerning consciousness into sharp focus. However, the particular explanatory gap concerning consciousness which worries Levine and the inflationists and deflationists who have followed him in the explanatory gap debate, is premised on a conception of consciousness as appearance. This is not the real explanatory gap which exists in our understanding of consciousness.

We will never see an intuitive connection between brain activity and the taste of chocolate. We can perceive chocolate through the sense modality of taste, but we can
never escape this perspective to learn *something else* about chocolate as perceived by
the sense modality of taste. Further explanation can only concern tastings of chocolate,
i.e. the taste of chocolate for x or for y, which is not something we can perceive,
introspect, or otherwise be aware of in any straightforward way. If we grant Levine
that closing the gap requires a D-N explanation, this will still only be a deduction from
neural activity to a *tasting of chocolate*, something which incorporates the subject.

There should be no doubt that we do have more to learn about tastings of
chocolate and the like, for consciousness has not yet been adequately explained. There
is a scientific explanatory gap in our understanding of consciousness, just as there once
was in our understanding of water. Philosophy can contribute by proposing and
debating analyses of mental concepts for the various mental phenomena, to uncover
the structures that need explaining, and if states are said to be "representational" or
"dispositional", then another task will be to reductively account for these notions. As
science progresses, such analyses may be confirmed or disconfirmed. For instance, as
we discover more about what actually goes on in the brain and body when people
report the occurrence of mental phenomena, we will have more of an idea of what
these states could plausibly be interpreted as representing.

However, a *tasting of chocolate* is just as theoretical a notion as any in science;
it is something we know is going on when we taste chocolate. If and when the gap is
closed, it may be inconceivable to us that certain brain activity should occur without
the subject of that brain activity tasting chocolate.332 But conceivability arguments in
which we imagine the taste of chocolate are irrelevant to issues about explaining
consciousness, as we have tried to make clear.

---

332 Though perhaps only within a certain environmental context.
9.f. Conclusion

For McGinn, we are constitutionally incapable of understanding the connection between consciousness and the brain. Our possession of distinct faculties of introspection and perception bars us in principle from making any connection between the colours, sounds and tastes we know by introspecting, and the brain states that we can only perceive.

For Papineau, to wonder how a conscious state could be a brain state, is like wondering how Tony Curtis could be Bernie Schwartz. Given that Tony Curtis is Bernie Schwartz, there is no sensible question to ask of how he “could be”. Likewise, consciousness is the brain, “and there’s an end on it”.

For Searle, consciousness cannot be identified with the brain, since conscious states are subjective appearances, and brain states are not. But this presents no metaphysical difficulties; the solution to the mind / body problem “has been available to any educated person since serious work began on the brain nearly a century ago”. It is that the brain causes consciousness, i.e. how things appear to us.

For Tye, when we consider a brain state in isolation, it is bound not to seem like we are considering consciousness. This is because conscious states are representational states; they are brain states representing external properties. So in introspection, we are aware of states that represent the external world, and our phenomenal concepts (of colours, tastes, etc.) are concepts of what it is like to be in those states.

---

My view is that we have no introspective grasp of consciousness, and so we have no reason whatsoever to suppose that consciousness could not be the brain, or brain functioning. Consciousness gives us the world, and the world only. It does not give us itself, not even in our unusually reflective moments. There is no special attitude we can adopt to contemplate our conscious states themselves, or the properties of those states. Being conscious, we have access only to what we are conscious of.

McGinn “introspects” colours and tastes, finds no conceivable connection between them and neural happenings, and proposes a principled reason for this. Papineau “introspects” colours and tastes, considers it a reasonable scientific hypothesis that they are brain properties, and realises that if they are, then there is no sensible question of why they are. Searle “introspects” (or is aware of through “self-consciousness”) colours and tastes, accepts a principled distinction between them and neural happenings, and sees the main problem of consciousness to be finding out how these appearances are caused by neural happenings. Tye “introspects” colours and tastes, and puts the apparent difference between them and neural happenings down to the fact that the neural happenings represent external states of affairs.

But when you “introspect”, you find nothing new, only the same old world of colours, tastes, etc. that was there before you decided to “introspect”. Making the decision marks only a change of intention, the intention now being to focus on experience itself. But intending to do this does not mean that we actually can. What really happens in “introspection” is that the philosophical apparatus of the appearance / reality distinction is brought to bear upon what we find in experience, so that the world, which is all that we ever find in experience, is now misinterpreted as experience itself, i.e. appearance. Once we draw on this philosophical apparatus, and form a
conception of experience as something that we can experience, then we introduce a conceptual difference between our way of thinking about consciousness (appearance), and our way of thinking about the physical world. And then there seems to be a very peculiar explanatory gap.

The philosophical apparatus of an appearance/reality distinction is part and parcel of a false philosophical system designed to explain the metaphysical fact that the way we represent the world to be is not necessarily how the world actually is, and the epistemological fact that we can be sure about how the world seems to be without being sure about how the world is. The problematic “explanation” was that “the way we represent the world to be” is something we can experience, i.e. appearance, which may or may not match how the world actually is.

One development of this way of thinking was the theory that we are only ever aware of “sense-data” in perception. This lost popularity in favour of the view that we do actually perceive the objective, physical world. But the old way of thinking lived on in the idea that though we do most of the time experience the external world, we have, in introspection, the ability to direct our attention upon aspects of our experiences themselves, rather than the world which our experiences present to us.

Physicalism has ample resources to provide alternative explanations for the facts that the concept of appearance was introduced to explain. The way we represent the world to be is not necessarily how the world actually is, because broadly physical representational states are fallible; sometimes we misrepresent the world. And the reason we can be sure about how the world seems to be, without being sure about how the world actually is, is that “how the world seems to be” is just how we would judge the world to be on the basis of how our representational states present the world as
being. Since we know that we sometimes misrepresent the world, we cannot be sure that such judgements will be accurate.

So long as we conceive of consciousness as appearance, thereby interpreting what we find in consciousness according to an outmoded canon of philosophical explanation, we will have concepts of consciousness that we will not be able to connect with our ordinary physical and functional concepts. Introspective awareness of consciousness will provide us with these problematic “phenomenal” concepts, as well as a direct acquaintance with the mystery of consciousness, as we “look” at experience and “look” at the brain, and struggle over the apparent need for there to be a connection between the two.

The explanatory gap problem is a perfect distillation of the mind / body problem, for it brings out its intuitive basis: the inability of physicalism to provide an explanation of consciousness. And the reason physicalism has no such explanation is that nobody has solved the problem of conceptual difference, by finding a way to connect concepts of appearance with physical concepts. But the problem of conceptual difference is a sham; the concept of appearance is a philosophical inheritance, devised for the express purpose of defining something epistemically closer to us than reality, and necessarily distinct from it. And the properties we apply the concept of appearance to in “introspection” are not properties which the concept fits: they are just properties in the world.

Objects and properties appear in consciousness: there is an ordinary sense of “appearing” that we can salvage, as well as an ordinary sense of “introspecting”. But appearance itself is something which philosophy made up. Consciousness is not. And yet it is appearance that causes all the problems for physicalism. Ayer once said that
"philosophical problems are not settled simply by our taking care that they should not arise".\textsuperscript{337} This is not true if the philosophical problem arises because of a philosophical mistake.

\textsuperscript{337} Ayer (1956) \textit{The Problem of Knowledge}, p. 98.
Chapter 10: New View

10.a. Introduction

We began with two views, V#1 and V#2, which are influential in determining physicalist positions on the nature of consciousness. They seem to be incompatible if a conception of consciousness as appearance is brought to bear on V#2. But we can now see a way in which the main insights of both can be preserved, without any incompatibility arising.

10.b. V#3

We live in a physical universe. All objects, large and small, are made up from and connected by the tiny particles, forces, waves, and suchlike, which physical science describes. We ourselves are no exceptions. Humans are physical things in the world, like anything else. What does mark us out as exceptional though, is consciousness. This special physical constitution means that we can know the rest of the physical world around us. Vases, by contrast, are unconscious, and as a consequence they are oblivious to their surroundings.

Consciousness gives us a perspective on the world. Functional states of human brains are representational states which present parts of the world to a subject. These are states of consciousness. The subject of a conscious state is presented with the intentional objects of these representational states, which are parts of the world from a
particular perspective. Conscious human beings, in virtue of occupying these representational states, are confronted with the world from the perspective they occupy, and the world only.

The subject is presented with the world, not its own representational states, since any such state includes the subject itself. The subject is an aspect of consciousness which, though not a part of the world being presented by consciousness, must nevertheless be present for us to make sense of what is presented to us. We find the world, but we can choose what to look at. We can adopt an attitude to what we find, or find ourselves taking such an attitude. Ever present are our recent memories, following us like shadows, and more distant memories which situate current experience as if on a familiar landscape.

Which parts of the world are available from our "first-person" perspectives, depends on where in space and time we are, and hence where our conscious states are. This is no more mysterious than the fact that which snooker balls a moving cue ball can interact with at any particular time depends upon which snooker table it is on. The presentation within consciousness of the world from a particular perspective is "private" because consciousness presents the world to the subject of the state only, that subject which is a constituent of the conscious state, a state which is a physical-functional particular in a certain place at a certain time, like any other concrete particular. The brain does not broadcast its representations for popular consumption, like a TV set.

From the perspective on the world with which we are provided by consciousness, we describe the world. We have developed objective descriptions, descriptions of the things we represent which aim to abstract from the multiplicity of
ways in which we can represent them. These descriptions of the physical world are
objective, valid for everyone. However, they are only descriptions of the world
presented to conscious human beings. Though we hold the conviction that the world
exists in abstraction from any particular perspective, we can only describe it as we find
it. With physical science, we take a further step away from the perceived world
towards absolute objectivity, with theories of unobservable properties that explain the
observable features of the world. These theories may also explain the interactions we
observe between other types of subject and the environment. However, we can never
uncover the world of other subjects (because we do not have their consciousness), or
the world in itself (because we are conscious).

We can think about both what an experience is, and about having an
experience. In thinking about what an experience is, we can think abstractly about its
structure as a representational state, as a presentation of some part of the world to a
subject. Or we can visualise the physical state that realises it. Thinking about having an
experience is a matter of imagining instantiating that physical state ourselves, i.e.
imagining the world as it would be presented to the subject of that state. Either way,
we are just thinking about parts of the world, though different parts.

Since different human beings have different physical constitutions, different
histories, and occupy different spaces at different times, their mental states have
different subjects, and they focus on the world from different standpoints. We abstract
from these standpoints to describe the common world we all represent. Each person’s
own consciousness provides them with representations of the world, and from patterns
of similarity between those representations, we learn observational descriptions of the
world. Science provides us with more accurate descriptions. Since consciousness is for
representing the world, not itself, we have no observational descriptions of
consciousness; our ordinary conception of consciousness comes from reasoning based
upon observation. Science promises more accurate descriptions here as well.

Consciousness presents an incredibly complex array of properties in the world,
properties of the world around us and properties of our own bodies. We are presented
with both sorts of properties all the time and at the same time, and we can form
recognitional concepts of any of them. Recognitional concepts refer to properties in
the world by connoting functional descriptions of those properties. Thus they are
relational conceptions of the world, for the reason that having something presented in
consciousness does not put us in a position to know its intrinsic nature. Science may
discover that there is no single represented property to correspond to a recognitional
concept, since recognitional concepts are something of a blunt instrument: the noticed
similarity may be due something else, a functional similarity between states with
different contents, for instance. Scientific descriptions are more precise; microphysical
concepts essentially conceive their referents, since we accord to physics the authority
to provide our most accurate descriptions of the world.
BIBLIOGRAPHY


Valberg, J. J. (Unpublished) “Why there could not be an explanatory gap”.


