# Chapter 4 Consciousness and the Antipathetic Fallacy

## 4.1 Introduction

In chapter 1 I had occasion to mention dualism, which I characterized as the view that conscious mental states have features which cannot possibly be possessed by physical states. At th at stage I argued that the price of dualism was epiphenomenalism, on the grounds, roughly, that dualism requires conscious mental states to be distinct from the physical causes of behavioural and other physical effects. But I left it as an open ques tion whether the arguments for dualism, and in particular for its conception of consciousness, made epiphenomenalism a price worth paying.

In this chapter I want to argue that there is in fact no good motivation for the dualist view of co nsciousness, and that we should therefore uphold the simple physicalist position that all mental states, including conscious states, are identical with or realized by physical states. The advantage of this physicalist position is that, unlike dualis m, it allows us to view conscious mental states as genuine causes of behavioural effects.

It will be convenient to use the term "physical state" in a liberal sense throughout this chapter, to include not only strictly physical states in t he sense of chapter 1, but also any second-order or higher-order states which are realized by physical states.[1] The differences between these different kinds of "physical states" will not matter for most of this chapter, since most of the arguments between dualism and physicalism arise in exactly the same way whichever kind of "physical state" the physicalist identifies conscious mental states with.

At the end of the chapter, however, the difference between "first-order" and "higher-order" physical states will become relevant. So far in this book I have tended to assume that mental states are, if anything, identical with higher-order physical states, rather than with first-order ones: as I observed in chapter 1, it seems unreasonable to hold that extraterrestrials, or people with brain prostheses, cannot share thoughts with us, just on the gounds that their brains contain different kinds of molecules. With the specifically conscious features of mental st ates, however, the situation is somewhat different. For, as we shall see, there are persuasive arguments, based on various inverted spectrum thought experiments, for holding that such conscious features in particular depend on the physics of the bra in, rather than on its higher-order organization. This is fine-tuning, however. The prior issue is whether conscious features can be identical with any kind of physical property, first-order or higher-order. After that we can worry about which kind of physical property they might be identical with.[2]

I shall proceed as follows. In the next three sections I shall consider some recent arguments for dualism. I shall argue that they are quite ineffectiv e. Accordingly, in section 4.5, I shall ask why the intuitive pull of dualism is nevertheless so strong. My diagnosis will be that we are seduced by a fallacy, which I shall call the "antipathetic fallacy", into thinking of consciousness as so mething distinct from the physics of the brain. The final five sections of the chapter will then explore some of the consequences of this diagnosis.

## 4.2 What is it Like to be a Bat?

Much of the contemporary literature on consciousness begins with Thomas Nagel's article "What is it Like to be a Bat?" (1974). Nagel argues that conscious mental life involves certain essentially subjective facts, facts that can only be appreciated from the "first-person" point of vi ew, from the point of view of the subject of those conscious experiences. Such subjective facts contrast with objective facts, like physical facts, which are accessible from the "third-person" perspective, independent of any particular subjective po int of view. Nagel concludes on this basis that any physicalist account of mind must fail to account for the subjective aspect of mental life.

Nagel illustrates this thesis by inviting us to reflect on the echolocatory experience of bats. Nagel takes it that bats, like other mammals, have conscious experiences. In particular, he takes it that bats have conscious sensory experiences when they echolocate. But, he points out, we human beings are unable to adopt the ba ts' point of view, and so have no idea what those bat experiences are like. You might think that echolocation would be like flying about in the dark and hearing lots of high-pitched noises. But that would be what it is like for beings like us, with human perceptual apparatuses, to echolocate, and not, presumably, what it is like for bats to echolocate. And, indeed, the more we think about it, the more it becomes clear that we have no grip on the subjective nature of the bat's echolocator y experience.

Nagel focuses on bats, not because he has any doubts about bats being conscious, but rather because our complete inability to adopt the bat point of view highlights the existence of the subjective side of bat experience.&nbs p; In the case of other human beings, and perhaps even of chimpanzees and dogs and squirrels, we can put ourselves in their places, and imagine having experiences like theirs. And so it is easier, in these cases, not to notice that grasping the subj ective aspect of experience requires us to abandon the objective, third-person perspective. But since we can't put ourselves in the place of bats, this particular example forces us to recognize that a purely objective perspective does not in fact gi ve us any access to the subjective reality of experience.

Despite the plausibility of Nagel's line of argument, I think that physicalism can meet the challenge he poses. Let us proceed in stages. For a start, we should immedia tely concede that there is one sense in which we human beings are indeed cut off from the facts of bat experience. We do not have echolocatory experiences, whereas bats do. In this sense it is undoubtedly true that we "lack access to", "cannot appreciate", or whatever phrase you prefer, the "subjective reality" of bat experience. But this observation in itself clearly yields no argument against physicalism. For physicalists are just as well placed as anybody else to explain this di fference between bats and humans. Physicalists think that conscious experiences are identical with certain physical events in the brain. So physicalists can say that the difference between bats, who have echolocatory experiences, and humans, w ho do not, is simply that certain physical events, namely, those which constitute echolocatory experiences, occur in bats, but not in humans. In this sense the physicalist can happily agree that bats have access to experiences which humans cannot ap preciate.

This point is central to the physicalist view of conscious experience. Physicalism does not deny that there are conscious experiences, nor, if you wish, that "that it is like something to have them". The claim is onl y that this is nothing different from what it is to be a physical system of the relevant kind. Of course there is something it is like to experience pain, or to see red, or to taste cheese. And such things are highly important, especially for the subjects of those experiences. But, insists the physicalist, they are not non-physical things. What makes it like that for you is that you are you, that is, that you are a physical system of a certain sort. If you were physically dif ferent in the relevant respects, things would be different for you.

This is the initial physicalist response to Nagel's challenge. There is, however, a more persuasive line of argument suggested by Nagel's position. Suppose th at you did somehow come to have echolocatory experiences. Wouldn't you then differ from other human beings, not just in having had those experiences, but in then knowing something that other humans didn't know, namely, what echolocatory experiences were like? And wouldn't this be knowledge of an essentially subjective fact? For note that you could have known everything there is to know about bat experience from an objective point of view -- you could have been an expert on bat echolocati on, who knew all about the physics and physiology and computational workings of the bat's brain -- and yet, prior to having had the echolocatory experiences, you would not have known what they were like. So it seems to follow that after you have had an experience you acquire knowledge of certain facts -- the subjective, phenomenal features of the experience -- which are necessarily omitted by any objective, physicalist story.

## 4.3 What Mary Didn't Know

This "knowledge argument" is developed explicitly by Frank Jackson in "Epiphenomenal Qualia" (1982) and "What Mary Didn't Know" (1986). Jackson simplifies the issue by focusing on a case where it is mere happenstance, rather than the wrong cognitive appa ratus, that prevents somebody having certain experiences. He tells the story of Mary, who is an expert on the psychology and physiology of human colour vision. Mary knows everything there is to know about the goings-on in our brains when we wh en we see red, say. However, Mary has always lived in a restricted black-and-white environment. All the objects she has ever seen are black or white or grey. She has never herself seen anything red. Then one day she is presented wi th a red object. She then has the experience of seeing something red. And as a result she learns something she didn't know before. She now knows about the phenomenal nature of red colour experiences, when before she was ignorant of this. Remember, however, that Mary had always possessed complete objective information about colour experiences. So, once more, it seems to follow that there are items of information about experience that must be omitted by any physicalist account.

This "knowledge argument" adds an extra dimension to Nagel's original defence of subjective facts. But there is still plenty of room for physicalism to resist it. The natural physicalist response to this argument is to admit that there are indeed before-and-after differences in Mary, consequent on her having had her first experience of red, but to deny that these involve her becoming acquainted with some subjective feature of colour experience. There are other ways of c onstruing the changes in Mary,

which do not require the postulation of such subjective facts, and which do not therefore imply that a physicalist account of experience must be incomplete.

In this section and the next I shall outline a phs yicalist construal of these changes. In this section I shall consider changes in Mary's recreative powers of imagination and recall, and in her ability to reidentify her experiences. In the next section I shall consider changes in her concepts of experience. This will involve retreading some relatively familiar philosophical ground. But my aim is not just to block Jackson's argument -- other philosophers, referred to below, have already shown how to do that -- but rather to point t o a striking common feature of the experientially produced changes in Mary, namely, that they all yield ways of thinking about experiences that deploy versions of those same experiences. This point will be central to my subsequent diagnosis in secti on 4.5 of the "antipathetic fallacy" which I take to be responsible for the intutive pull of dualism.

The first before-and-after change to be considered concerns Mary's new powers of recreation. Once she has seen red, Mary can recre ate the experience of seeing red, in imagination and memory, whereas before she couldn't. Mary could of course always imagine, in the third-person, so to speak, that somebody else was seeing red, in the sense that she could imagine such-and-such phy siological or behavioural occurrences in that person. And, similarly, she was always able to remember, in the third-person again, that somebody had seen red. But now she has a new ability, the ability to imagine or recall having the expe rience itself, from the inside, as it were. She can now relive the experience, as opposed to just thinking about it.

An anti-physicalist like Jackson can account for this change in terms of Mary's new knowledge of a non-physical fac t. When Mary experiences red, on this anti-physicalist account, she discovers that the experience has a characteristic phenomenal feature P. And then, because she has this new knowledge, she can imagine the experience by entertaining the thoug ht that someone has an experience with property P. Similarly, she is now able to recall the experience by remembering that she herself had an experience with property P.

Physicalists will offer an alternative account. Supposet hat the kind of imagination and memory at issue depends on the brain literally recreating a version of the experience being imagined or remembered. That is, suppose that first-person imagination or memory requires that brain be in a state which is s imilar to the state constituting the original experience. It won't be exactly the same state, since imagining or recalling a pain is different from having a pain. But it could well be a similar state, a kind of faint replica, which would fit w ith the fact that an imagined or remebered pain shares to some slight extent the unpleasantness of a real pain.

This alternative suggestion yields as good an explanation of the fact that you can only imagine or recall experiences you have previously undergone as the theory which postulates new knowledge of phenomenal property P. For it seems highly plausible that the brain's ability to recreate an experience depends, as a matter of empirical fact, on its having at some time had an o riginal version of that experience, to give it, so to speak, the mould from which to make the replicas.[3]

What is more, this alternative account of Mary's new ability is quite consistent with an objective, physicalist account of conscious experiences. For on this alternative account the difference produced in Mary by her original experience of seeing red is

not that she acquires some new item of knowledge, but simply that she can now do something she could not do before, n amely, recreate that experience in imagination and memory.[4] The earlier account, which attributed new knowledge of phenomenal property P to Mary, implied that her previous third-person information about the experience left something out.& nbsp; However, since the new account does not credit Mary with any such new knowledge, there is now no implication that a physicalist account of conscious experience is incomplete.

Some readers may feel that this physicalist account of fi rst-person imagination and memory is an ad hoc theory whose only attraction is that it saves physicalism. But this would be unjust. For the account also has the positive virtue, noted in passing above, of offerring some explanation of why an i magined or remembered experience resembles the original experience itself -- namely, that such imaginings and rememberings literally involve a copy of the original experience.[5]

D.H. Mellor uses the term "secondary" to refer to this kind of copied experience, the kind of experience which ocurs when we recreate in imagination or memory those primary experiences we have previously undergone.[6] The existence of such secondary experiences which resemble their p rimary versions will be central to my eventual explanation of the antipathetic fallacy.

Of course this talk of "resemblance" between secondary and primary experiences needs further elaboratation, both to specify what kind of replication i s involved, and to explain how the resulting replicas mimic the original experiences in our cognitive workings. But I take it to be uncontentious that there is some phenomenon of resemblance here, and that the model of "secondary" replicas of primar y experiences offers a promising route to an explanation.

So far I have considered the new recreative powers of imagination and recall produced by Mary's first experience of seeing red. Another such before-and-after change is that M ary aquires a new introspective power to reidentify that experience when she has it again. Mary of course always had the ability to recognize "from the outside" when somebody was seeing red, from environmental or behavioural or physiological evidenc e. But now she has a new ability, to recognize, by direct introspection, that she herself is seeing red.

Again, one possible explanation of this new first-person ability would be that Mary discovers that the experience of seeing red has phenomenal property P, and that as a result she can now pick out experiences with property P as instances of seeing red. But, as before, this is not the only possible explanation of the new ability. For we can suppose instead that Mary si mply acquires a non-conceptual "template", in David Lewis's phrase[7], which can then be compared directly with further experiences, and cause Mary to believe that she is experiencing red again. She doesn't arrive at this belief by noting t hat the experience has property P, and concluding that it is an experience of seeing red. There is simply a mechanism in her brain which compares the experience with the template and yields this belief directly.

As with the "seconda ry experience" account of imagination and memory, the "template" account of introspective recognition both yields a plausible account of why we should need the original experience in order to acquire the recognitional ability (namely, because the brain ne eds the original to have the materials from which to form the template)[8], and remains consistent with physicalism (since it doesn't explain

Mary's new recognitional ability by attributing knowledge of phenomenal property P to her, but simply by postulating a new mechanism in her brain).

It is an interesting further hypothesis that the same cognitive operations may be involved both in recreative and in recognitional abilities. Perhaps the brain uses the processes constitut ing our "secondary experiences" themselves as the "templates" by which it classifies new experiences: that is, perhaps its mechanism for recognizing such new experiences is simply to compare them with the replicas which are activated in imaginationa nd recall. It does not seem inevitable that things should work like this: there is no contradiction in the idea of beings who could classify new experiences by some template process, and yet lacked the ability to recreate those experiences in imagination or memory; and perhaps it is even possible for there to be beings who could recreate experiences, but who lacked the second-order mental ability to classify them. But it seems clear that in human beings the two abilities always go together, and the natural explanation is that they do so because the same mechanism subserves both.

## 4.4 Concepts of Experience

The overall argument of the last section can be put as follows. In so far as Mary's first experience of red leads to her knowing something she didn't know before -- leads to her "knowing what the experience is like", if you want to put it that way -- this new knowledge can be construed as her knowing how to do something new, r ather than as her knowing that anything new. There are indeed genuine changes produced by Mary's new experience. But these changes are all a matter of her acquiring new abilities -- to recreate or recognize the experience -- not of her forming any new kinds of judgements about the world.

But can this be the full story? Surely, many readers will feel, new experiences doesn't just give us the abilities described in the last section. They also enable us to think new t houghts. Once you've seen red, then can't you think of that colour, and judge it to be vibrant, or threatening, or something everybody should experience at least once, in a way you couldn't before?

I agree. But I think that th is too can be accommodated by physicalism. The important question for physicalism is whether new experiences lead to our knowing about any new features of the world. Physicalists need to deny this. But they can consistently allow that ne w experiencs lead to our acquiring new concepts for thinking about those features. In Fregean terms, the change would be at the level of sense, not reference. Mary's thinking about the experience of seeing red would change, but what she was th inking about would be exactly the same thing as she used to think about when she was a scientist who had never herself seen red.[9]

In order to bring out this point, it will be helpful to switch examples slightly for a moment, an d consider, not Mary, but Jane, let us call her. Jane has always shared Mary's black-and-white environment. But Jane is no expert on colour vision. Indeed she has never heard of such things as colours and of people experiencing them.

Then one day Jane sees something red. Unlike Mary, she does not have available any public concept of the visual events that take place in people when they are presented with red objects. Indeed she may not even realize that the s ensation she is

currently experiencing is caused by some observable feature of her environment. Yet we would surely expect Jane to be able thereafter to to form beliefs about that sensation, such as that it was vibrant, threatening, something everyb ody should experience it at least once, and so on.[10]

Mary, on the other hand, does have a public concept of the visual events that take place in people when they are presented with red objects. But, despite this, there se ems no question but that Mary might acquire just the same kind of new thoughts as Jane does after experiencing red for the first time. For imagine that, the first time Mary experiences red, she does not know what it is -- she simply is n ot aware that the curious experience she is now having for the first time is the experience that is characteristically caused by red objects. In this case Mary will surely respond just like Jane, and start forming beliefs such as that this new exper ience is vibrant, threatening, and so on.

At first sight this might seem to substantiate Jackson's knowledge argument. Doesn't the fact that Mary follows Jane in forming new sorts of beliefs after her experience show that Mary's ori ginal set of physicalist beliefs must have left something out, namely, information about the sunbiective side of the experience? But this conclusion does not follow if, in line with my earlier suggestion, the novelty in Mary's beliefs lies at the le vel of sense rather than reference. And this of course is how the physicalist will diagnose the situation. Before Mary sees red, she has a "third-person" concept of this experience. Afterwards she also has a "first-person" concept. But they are concepts of the same thing. Mary is in the position of somebody who has thoughts about both Cicero and Tully, without realizing they are the same person.

A natural hypothesis about the structure of the new first- person concept acquired in common by Jane and Mary is that it involves a kind of exemplificatory reference by secondary experience. I earlier expressed the new belief formed by Jane and Mary as "that experience is vibrant". I now suggest that we take this construction at face value. Jane and Mary think: THAT experience is vibrant, accompanied by a secondary version of seeing red; they thereby secure reference to the experience of seeing red.[11]

; This account of first-person concepts of experience shows, as before, why you can't refer to experiences first-personally until you have had them (you need the primary mould to form the secondary replicas), and it does so consistently with physicalism ( since you don't become acquainted with any new facts, but just acquire new concepts).

It is interesting to consider what will happen if and when Mary figures out that her new experience is the kind of experience which is characteristicall y occasioned by red objects. The natural upshot, assuming that Mary herself is a physicalist,[12] it for her to conclude that she has two concepts with the same referent.[13] And then, as with anybody who realizes this, the t wo concepts will tend to "merge", each becoming merely an aspect of the unified concept with which she refers to the experience of seeing red.

The net result would be the same even if Mary were aware from the start that the new experience she is having is the one characteristically occasioned by red objects. For even in this case we could expect Mary to come to share the first-person

mode of thinking about this experience displayed by Jane, albeit that this new mode of thinking woul d now, supposing still that Mary is a physicalist, be "merged" ab initio with the concept Mary had when she was just a colour vision scientist.

Perhaps there is room to dispute whether such a "merged" concept is really a new concept, comp ared with the concept Mary had before seeing red. The merged concept will incorporate both the old third-person physical information Mary always had, plus the new first-person mode of thinking she shares with Jane. There are, however, familiar difficulties about whether such amplifications of existing concepts count as genuinely distinct concepts, rather than alterations of old concepts. But rather than getting bogged down in the knotty issue of concept identity, let us simply agree that Mary's concept of the experience of seeing red has been modified, in a way that would not have been possible if Mary had not seen red herself.

So, to sum up the argument of this section, once we have new experiences, we are led to form new sorts of beliefs about those experiences. But this does not show that we thereby come to refer to any distinctively subjectively phenomema. For the distinctive element in these beliefs need be nothing more that the deployment of first - person concepts, and, for all that has been said so far, there is no reason to suppose that such first-person concepts are not co-referential with third-person concepts of experience.

## 4.5 The Antipathetic Fallacy

I expect that, despite everything I have said so far, many readers will feel strongly that it is a mistake to conclude that "first-person" and "third-person" concepts of experience refer to the same things. For my arguments in the last three sections will have done nothing to shake the widespread intuition that conscious experiences and brain states are as different as anything can be.

Let me summarize the state of play. So far in this chapter I have considered the str ength of arguments against the physicalist identification of conscious experiences with brain states. And I take myself to have shown that these arguments are ineffective. There is no valid argument from "what it is like", or from "knowing wha t it is like", to discredit the physicalist view that having a given conscious experience is nothing more nor less than being a certain kind of physical system.

What is more, I take myself already to have shown, in chapter 1, that the cos t of viewing conscious mental states as something distinct from brain states is the denial of the efficacy of the mental: if you think that consciousness is non-physical, then you are forced to such undesirable conclusions as that your pain is never the cause of the motion of your arm.

I think that together these findings give us good reason to accept the physicalist view that conscious experiences are not distinct from brain states, and therefore to reject any intuitions to the con trary. However, it would be foolish to deny that such intuitions exist. Such non-physicalist intuitions exert a strong pull on all of us, even on us physicalist philosophers who are committed to rejecting them. So in this section I want to offer a diagnosis of these intuitions, with the intention of explaining why they arise even though they are mistaken.

In the previous two sections I have discussed a variety of ways in which we can focus mentally on conscious experienc es, a variety of mental acts which refer to types of experience. These acts can be divided into two main categories: those "third-person" acts which are possible prior to your actually having had the experience in question, and those those "fi rst-person" acts which are only possible after you have had the experience. In the former category are all the mental acts Mary could perform before she saw red: her "third-person" imaginings and memories of other people experiencing red; ; her non-introspective identification on behavioural or physiological grounds of certain events as experiences of seeing red; her "third-person" beliefs, conjectures, and other propositional attitudes about the experience of seeing red. In th e latter category are the "subjective" analogues of all these mental acts: the "first-person" imaginings and rememberings that involve internal recreation of an original experience; the introspective identifications of new experiences by direc t comparison with a "template"; the beliefs, conjectures, and other attitudes that can be formed by people like Jane whose concept of seeing red involves an element of ostension by internal exemplification.

The common feature of the se latter "first-person" acts, and what distinguishes them from the corresponding "third-person" acts, is that they all deploy a secondary version of the experience being referred to. This is the reason, I have suggested, why the first-person acts a re only possible after you have had the experience in question yourself. For it is only after you have had the experience that your brain will have the materials necessary to form secondary versions of that experience.

I think that this broad division between first-person ways of thinking about experience, which employ secondary versions which resemble those experiences, and third-person ways, which do not, is the source the strong intuition that conscious experiences involve someth ing more than the physics of the brain. For it is all too easy to conclude, when we reflect on the difference between these two categories of thought, that only the first-person thoughts really refer to experiences, while the third-person thoughts r efer to nothing except physical states.

The route to this conclusion begins with the perfectly accurate observation that first-person thoughts include an experiential element which is absent from the third-person cases. First-person thoug hts portray the relevant experience directly, so to speak, by giving the thinker a simulacrum, by recreating in the thinker a version of the experience being thought about. Third-person thoughts, on the other hand, do not do this, since they do not involve secondary experiences.[14]

So there is a sense in which third-person thoughts do indeed "leave something out": they do not give us (versions of) the experience being referred to. And this observation can then easily lead to the further conclusion that third-person thoughts are about something different from first-person thoughts: where first-person thoughts refer to the experience itself, in all its conscious immediacy, third-person thoughts merely refer to the external trappings of the conscious event, the physical goings-on which accompany it.

But of course this last step is a fallacy. The fact that we do not have certain experiences when we think third-person thoughts does not m ean that we are not referring to them. To make this move is to succumb to a species of the use-mention confusion: we slide from (a) third-person thoughts, unlike first-person thoughts, do

not use (secondary versions) of conscious experiences t o portray conscious experiences to (b) third-person thoughts, unlike first-person thoughts, do not mention conscious experiences. There is no reason, however, why third-person thought about experiences, like nearly all other thoughts about anything, should not succeed in referring to items they do not use.

I propose to call the above fallacy the "antipathetic fallacy". Ruskin coined the phrase "pathetic fallacy" for the poetic figure of speech which attributes human feelings t o nature ("the deep and gloomy wood", "the shady sadness of a vale"). I am currently discussing a converse fallacy, where we refuse to recognize that conscious feelings inhere in certain parts of nature, namely, the brains of conscious beings.

& nbsp; Let me be specific about the target of this charge of fallacy. My target is not the explicit argument against physicalist views of consciousness offered by Jackson. I take the points made in the last two sections already to have sh own what is wrong with Jackson's argument. Rather my target is a covert line of thought, whose fallaciousness is obvious once it is spelt out, but which I think has nevertheless seduced a great many thinkers into dualism: namely, the argument which moves from the true premise that third-person ways of thinking about conscious experiences do not use versions of those conscious experiences, to the false conclusion that those ways of thinking do not mention those conscious experience, but only ph ysical states.

Let me also be specific about what I take the identification of this fallacy to explain. It is supposed to explain why many people believe that some mental states are distinctively non-physical. It is not suppos ed to explain why some physical states are distinctively conscious. This latter kind of question will be addressed in the next section, and I shall there agree that our ability to think about certain states in first-person ways does nothing to accou nt for their possessing the distinctive inner light of consciousness -- though I shall also argue there that the desire to account for such inner lights rests on a confusion. My present concern, however, is not to explain why the states we can think about in first person ways are distinctively conscious, but rather to explain why these states are widely taken to be non-physical.

Both Thomas Nagel, in a well-known footnote in "What is it Like to be a Bat?" (1974, pp 446-7), and Willi am Lycan, in his book Consciousness (1987, pp 76-7), briefly allude to versions of the fallacy I am concerned with. My treatment here enlarges on their remarks in two respects. First, both Nagel and Lycan focus specifically on the contrast bet ween first-person imagination of conscious experiences and the third-person perceptual imagination of the associated brain states: the contrast, for example, between imagining having a pain and imagining the visual appearance of the relevant parts of the sufferer's brain. This is certainly one example of the kind of contrast I am interested in, but this exclusive focus underemphasizes the extent of this contrast. For, as I have observed, the contrast between first-person and third-person modes of thought is not restricted to imagination, but also includes memory, identification, and believing, desiring and other propositional attitudinizing. And even within the category of imagination, perceptual imagination is not the only kind of third -person imagination: if we can form non-perceptual beliefs and other propositional attitudes about brain states, as we surely can, then presumably we can imagine them non-perceptually too. (Nagel does mention "symbolic imagination", but only t o exclude it from his analysis.)

Second, neither Nagel nor Lycan emphasize the way that first-person modes of thinking about experiences deploy secondary versions of those experiences. Nagel does, it is true, say that first-person i maginings "resemble" the experiences being imagined. But when he goes on to explain how the fallacy arises, his explanation, like Lycan's, is simply that first-person and third-person imaginings are independent mental acts, each of which can happen witout the other, and that therefore we are inclined to conclude that they are about different things.[15] But this diagnosis fails to distinguish the antipathetic fallacy from all the other cases where different modes of thought abou t the same entity can create the impression that two different entities are being thought about. What is distinctive about the antipathetic fallacy, and what makes it so very seductive, is the fact that one set of ways of thinking about experiences -- the first-person ways -- involve versions of the experience itself, and so create the impression that the other ways of thinking about experiences -- the third-person ways -- leave something out. In general, when two different mode of thought cre ate the impression that two things are being thought about (for example, Cicero and Tully), the illusion is easily enough dispelled on receipt of evidence that there is in fact only one referent. But in the mind-body case the impression of differenc e continues even in the face of any amount of such evidence, precisely because of the extra feature -- the first-person use of secondary versions -- that makes it seem as if the third-person modes of thought omit mention of the experience altogether. [16]

## 4.6 Theories of Consciousness

So far I have argued that there are no effective arguments against the physicalist identification of conscious states with physical states, and that the admittedly strong in tuitions which run counter to this view can be explained away. It may still seem to some readers, however, that a further obligation faces defenders of a physicalist view of consciousness: namely, to answer the question raised briefly in the m iddle of the last section, and explain why some states are conscious and others not.

The obligation I am thinking of here is not just to provide physicalistically acceptable accounts of such specific conscious states as being in pain, see ing red, having an itch in your left finger, or so on. We can suppose for the moment that physicalists can somehow specify which physical occurrences constitute each of these specific mental states. The current challenge is rather to giv e an explanation of the generic difference between conscious and non-conscious states as such. Why is consciousness present when a person is in pain, or happy, or itching, but not when a stone is falling, or a tree is growing, or, for that matter, w hen an anaesthetized human is breathing?

Some philosophers of physicalist inclinations have proposed "theories of consciousness" in answer to this kind of question. I have in mind the kind of theory which aims to identify a physical istically acceptable characteristic common to all and only conscious states. Some such theories are based on assumptions drawn from everyday thought (for example, Armstrong, 1968, pp 92-99, holds that the states of any self-representing system are c onscious); others appeal to the resources of cognitive science (for example, Dennett, 1978, ch 9, suggests that cognitive systems with short-term buffer memories are conscious); and no such theory, I think, commands universal assent.

; However, we can leave the details of such theories to one side. For a natural reaction to all such theories is that they simply fail to address the philosophical question at issue. At best such a theory will specify some structural or other physically acceptable characteristic (A, say) which is coextensive with the class of states we are pretheoretically inclined to count as conscious. But then we still seem to face the question: why does consciousness emerge in just those cases? And to this question physicalist "theories of consciousness" seem to provide no answer.

I suspect that many philosophers regard the inability to answer this question as the fatal flaw in the physicalist approach to consciousn ess. Surely, they feel, any satisfactory philosophical view of consciousness ought to tell us why consciousness emerges in some physical systems but not others.

I think that physicalists should simply reject this question. For the question presupposes that there are two different features at issue, the physically acceptable characteristic A, and being conscious. The physicalist is then challenged to explain the relation between these properties, and in particular to expl ain why they are always found together. But the physicalist should simply deny that there are two properties here. Being conscious isn't something over and above having A, it just is having A. (In the section after next I shall ask some questions about the sharpness and determinacy to be expected from any A which might provide such a physicalist reduction of consciousness. But it will helpful to shelve such worries for the moment, and assume that some suitable property A is availab le.)

The idea that being conscious just is having some physical state might seem intuitively implausible: surely the difference between conscious and non-conscious systems is something more than the difference between having and lac king some physical feature. But the defender of a physicalist theory of consciousness, while not denying that these intuitions exist, can account for them as a further manifestation of the antipathetic fallacy. The earlier sections of this cha pter were concerned with the thesis that specific conscious states, like seeing red, are identical with specific physical states; and I argued there that our strong contrary intuitions can be explained away as due to the antipathetic fallacy. I would now like to suggest that a generalized version of this fallacy is responsible for the intuition that any physicalist theory of consciousness will necessarily be incomplete

We can think of the general property of being conscious as standing to experiences like seeing red as determinable to determinate. Seeing red, being jealous, feeling cold, and so on, are the determinate states which have in common the determinable state of being conscious. And so, just as the antipath etic fallacy makes us think that such determinate states as seeing red are distinct from any specific physical states, so it makes us think that the determinable state of being conscious is similarly distinct from any more general physical state. We are inclined to think of the determinable feature as a kind of generalized non-physical light, which stands to the non-physical features of particular experiences, as, say, the property of being illuminated as such stands to being illuminated with red li ght. But we shouldn't. Just as it is a mistake to think of experiencing red as something additional to the relevant physical property, so it is a mistake to think of being conscious as an extra inner light, over and above the physical feature A.

Once we fully free ourselves from the seductive "inner light" picture of consciousness, and take seriously the idea that being conscious may literally be identical with some physical A, then we should stop hankering for any further exp lanation of why physical state A yields consciousness. Consider this parable. Suppose that there are two groups of historians, one of which studies the famous American writer Mark Twain, while the other studies his less well-known contemporary , Samuel Clemens. The two groups have heard of each other, but their paths have tended not to cross. Then one year they both hold symposia at the American Historical Association, and late one night in the bar of the Chicago Sheraton the penny drops. They realize that they have both been studying the same person. At this stage there are plenty of questions they might ask. Why did this person go under two names? Why did it take so long to find out Mark Twain and Samuel Cl emems were the same person? But it doesn't make sense for them to ask: why were Mark Twain and Samuel Clemens the same person? If they were, they were, and there's an end on it.[17]

Similarly, the defenders of a physicalist theory of consciousness can say, with consciousness and the physical property A. The defenders of such a theory will take themselves to have discovered that consciousness and A are the same property. So they will allow that we can sensibly ask why there should be different concepts of this property, and why it took us so long to realize that they stand for the same thing; and indeed they can answer these questions, by explaining that there are ways of referring to conscious phenomena that use secondary versions of those phenomena, and ways that don't, and that this in itself makes it easy to succumb to the antipathetic fallacy of supposing that different things are being referred to. But, they will insist, there is no further question of why consciousness is always present when physical property A is. If they really are the same thing, then we can't explain why they are the same thing. Somebody who feels there is still a question here has simply failed full y to grasp the thesis that consciousness is identical with a physical property.

## 4.7 Life and Consciousness

It may seem to some readers that a physicalist theory of consciousness will come close to denying the existence of consciousness. But that would be a mistake. It doesn't deny consciousness, just a certain conception of consciousness.

It denies that consciousness is some kind of extra inner light, some further non-physical property whic h exists over and above any physicalistically specifiable property. But this is quite consistent with holding that consciousness is a real property which distinguishes some kinds of systems from others. This combination of views requires only that we accept that consciousness is identical with some property which is specifiable in a physicalistically acceptable way.

An analogy may be helpful here. In the nineteenth century there was a heated theoretical debate about the essence of life. The participants had a satisfactory enough working notion of life: they agreed about which kinds of behaviour and physical organization are characteristic of life, and in consequence were clear enough about where in practice t he line should be drawn. Everything from humans to microbes are alive, while planets and pebbles are dead. (Perhaps there were some borderline cases; but the penumbra of vagueness was not wide.)

Still, despite this wide degree of agreement on the nature of life, nineteenth-century thinkers took there to be a further question. Why are these systems alive? What mysterious power animates them? And why is this power present in certain cases, such as trees a nd oysters, and not in others, like volcanos and clouds?

These questions have disappeared from active debate. Biology textbooks sometimes begin with a few perfunctory paragraphs about the distinguishing characteristics of their subj ect matter. But the nature of life is no longer a topic of serious theoretical controversy. Everybody now agrees that the difference between living and non-living systems is simply having a certain kind of physical organization (roughly, we wo uld now say, the kind of physical organization which fosters survival and reproduction).

The explanation for this nineteenth-century debate, and of its subsequent disappearance, was that it was premised on the notion that living systems w ere animated by the presence of a special substance, a vital spirit, or elan vital, which was postulated to account for those features of living systems, such as generation and development, which were though to be beyond physical explanation. And of course, if you do believe in such a vital spirit, then you will want to know about its nature, and why it arises in certain circumstances and not others.

However, nobody nowadays believes in vital spirits any more, not least because it i s now generally accepted that the characteristic features of living systems can in principle all be accounted for in physical terms. In consequence, it no longer makes sense to puzzle about why living systems are alive. To be alive is just to be a physical system of a certain general kind. There isn't any extra property present in living systems, over and above their physical features, which distinguishes them from non-living systems. So we have stopped asking questions which presu ppose such an extra property.

I recommend that we do the same with consciousness. The apparently nagging question, "Why does consciousness arise in certain physical systems?", is premised, I claim, on the assumption that consciousne ss is some extra feature, over and above any physical characteristic. But if we accept, as I have argued, that there is no reason to view consciousness in this way, then we ought therewith to stop asking why consciousness is present in the relevant kind of physical system.

Of course the parallel is not complete. In the case of life, the motivation for postulating an elan vital is purely explanatory, a desire to find a cause for phenomena which do not appear to be physically ex plainable. In the case of consciousness, by contrast, there is also the extra pressure of the antipathetic fallacy. Still, this doesn't affect the point. There may be extra reasons for thinking of consciousness as non-physical, which don 't apply to life. But once we recognize that it is physical, we should do what we did with life, namely, stop asking why it arises in the right physical circumstances.

One last point about the analogy with life. Note that the rejection of an elan vital does not mean that there is no life. There may be nothing special about living systems except a certain kind of physical organization. But this does not mean that the difference between being alive and not being aliv e is not real. The postulation of an elan vital was simply one theory about the nature of life. We can reject this theory, and yet still uphold, as we do, the distinction between living and inanimate systems.

A similar point a pplies to consciousness. We should reject the theory that consciousness involves an extra inner light in addition to facts of physical organization. But we can reject this theory without rejecting consciousnness. Even if consciousness is just a kind of abstract physical organization, the difference between being conscious and not being conscious can still be perfectly real.

## 4.8 Consciousness is Vague

So far I have been assuming that there is som e well-defined and precise physical characteristic A which picks out just those states we are pre-theoretically inclined to count as conscious. However, I doubt that this assumption is justified. In this section I shall argue that any physical ist account of consciousness is likely to make consciousness a vague property. In the next section I shall argue that questions of consciousness may not only be vague, but quite indeterminate, in application to beings unlike ourselves. I do no t intend these points as criticisms of physicalism. Rather my aim is to show that if we take physicalism serously, some assumptions that we take for granted about consciousness may have to go.

The point about vagueness is suggested by the analogy with life. If life is simply a matter of a certain kind of physical complexity -- the kind of complexity that fosters survival and reproduction, as I put it above -- then it would seem to follow that there is no sharp line between lif e and non-life. For there is nothing in the idea of such physical complexity to give us a definite cut-off point beyond which you have enough complexity to qualify as alive. Rather as with baldness, or being a pile of sand, we should expect th ere to be some clear cases of life, and some clear cases of non-life, but a grey area in between where there is no fact of the matter. And of course this is just what we do find. While there is no doubt that trees are alive and stones are not, there are borderline cases in between, like viruses, or certain kinds of simpler self-replicating molecules, where our physicalist account of life simply leaves it indeterminate whether these are living beings or not.

But now, if consciousness is like life, we should expect a similar point to apply to consciousness. For any physicalist account of consciousness is likely to make consciousness depend similarly on the possession of some kind of structural complexity -- the kind of com plexity which qualifies you as having self-representing states,say, or short-term memories. Yet any kind of such complexity is likely to come in degrees, with no clear cut-off point beyond which you definitely qualify as conscious, and before which you don't. So we should expect there to be borderline cases -- such as the states of certain kinds of insects, say, or fishes, or cybernetic devices -- where our physicalist account simply leaves it indeterminate whether these are conscious states o r not.

Some philosophers regard this as a reductio ad absurdum of the physicalist view of consciousness. They take it to be intutitively obvious that there is a sharp line between conscious and non-conscious states.[18]&nbs p; So they conclude that there must be something more to consciousness than a certain kind of physical complexity.

I go the other way. I think that the phsyicalist approach to consciousness is correct. So I reject the in tuition that there is a sharp line between conscious and non-conscious states.[19]

I accept, of course, that such intuitions exist. But I regard them as a further consequence of the "inner light" picture of consciousness, t he picture into which it is so easy to be seduced by the antipathetic fallacy. For if you do think of consciousness as such an extra inner light, then you will no doubt think it is a sharp matter which states are conscious -- states which possess th e inner light are conscious, and those which don't are not.[20] On the other hand, if the idea of such an extra inner light is a confusion, as I take it to be, then we have no obligation to respect any further intuitions which stem fr om it.

If the line between conscious and non-conscious states is not sharp, shouldn't we expect to find borderline cases in our own experience? Yet when we look into ourselves we seem to find a clear line. Pains, tickles, visu al experiences and so on are conscious, while the processes which allow us to attach names to faces, or to resolve random dot stereograms, are not. True, there are "half-conscious" experiences, such as the first moments of waking, or driving a famil iar route without thinking about it. But, on reflection, even these special experiences seem to qualify unequivocally as conscious, in the sense that they are like something, rather than nothing.

However, I don't think that this dis credits my claim that the boundaries of consciousness in general are vague. For I think there is a special reason why we are able to draw a sharp line in our own case. Namely, that in our own case we can simply note which states are introspect ible, recreatible in imagination and memory, and otherwise accessible in first-person ways. States which are so accessible we count as conscious, and those which are not we consider non-conscious.

What exactly is the rationale and s tatus of this decision procedure? This is a tricky question, to which I shall return in the next section. But whatever view we take on this question, note that the decision procedure in question will not work for all beings. For once we move beyond the case of humans, to those many animals and other possible organisms who lack the ability to think about their own cognitive states, then the decision procedure in question ceases to apply. So it will be of no help in deciding whether the states of sharks, for example, or octopuses, are conscious.

So I think we should accept that sometimes it will be a vague matter which states of which beings are conscious. It would be a mistake to conclude from this, however, t hat consciousness is unimportant or unreal. Any number of genuine and important properties are vague. Consider the difference between being elastic or inelastic, or between being young or old, or, for that matter, between being alive and not b eing alive. All these distinctions will admit indeterminate borderline cases. But all of them involve perfectly serious properties, properties which enter into significant generalizations, are explanatorily important, and so on.

## 4.9 Consciousness is Anthropocentric

In this section I want to raise some more serious doubts about consciousness, doubts which suggest that consciousness is not only vague, but downright indeterminate.

The la st section was premised on the assumption that consciousness involves some kind of physical or structural complexity; the corollary was simply that consciousness, like other kinds of complexity, will therefore admit borderline cases. But what if

there isn't any specific kind of complexity common to conscious states, vague or otherwise?

It will be helpful to approach this possibility by returning to the suggestion, made in the last section, that in practice we decide which huma n states are conscious by considering whether they can be thought about in first-person ways. Now, there are two different ways of looking at this decision procedure. One would be to regard it as a test for the presence of some property that c an be independently specified, such as appearing in the short-term buffer memory, say. On this way of conceiving the matter, consciousness is a property that can be independently specified, and first-person accessibility is an empirical symptom of t he presence of this independently specifiable property. But there is a rather more plausible way of understanding the decision procedure, which analytically ties the test of first-person accessibility to our notion of consciousness. That is, s uppose that our notion of consciousness starts with the test of first-person accessibility, and that the reference of this notion is simply fixed as that feature which is common to all those states which can be thought about in first-person ways. Fr om this point of view, first-person access isn't an empirical symptom of some independently specifiable property, but the hook by which we pick out that property in the first place.

This alternative, however, leaves open the possibility t hat there isn't any such property in the first place, vague or otherwise, After all, the class of states which we humans can think about in first-person ways is extremely heterogeneous. As well as pains, itches, tickles, and the various modes of sense experience, there are emotions, cogitations, and moods. There seems no obvious reason, on the face of it, why there should be any structural or other physicalist property common to this whole genus. Each species within the genus may share some common physical or structural characteristic which renders it explanatorily significant. But why suppose that there is some further such characteristic, common to members of all these species, which binds them all together?

; What about the property of being first-person accessible itself? This is a kind of structural property, and therefore physicalistically acceptable; and it is unquestionably common to all those states which can be thought about in first-perso n ways. But this property is ill-suited to provide an analysis of consciousness. For, even if first-person accessibility provides a reference-fixing description, our notion of consciousness seems clearly to be a notion of some other property w hich is responsible for first-person accessibility, not just the concept of first-person accessibility per se.[21]

This is why most people think it obvious that higher mammals, like cats, and bats, and human infants, have conscio us states, even though these animals are not capable of thinking of their own states in first-person (or any other) ways. These animals may not have first-person access to their own cognitive states. But their sensory and other states seem so closely similar to our own in every other respect that it seems natural to conclude that they must share the property that underlies the first-person accessibility of our own conscious states, whatever that property might be.[22]

However, to repeat the question, what if there is no such property? What if there isn't anything physically or structurally in common to all our first-person accessible states? We may still feel it is uncontroversial that other higher mammals are conscious, because of the close overall similarity between their states and our

own. But once we start considering beings that are less closely allied to us, like fish or toads, not to mention Proxima Centaurians and other extra-terrestrials, t hen we are left with nothing to go on, and it becomes quite indeterminate how the notion of consciousness should apply to their states.[23] The problem here isn't just be the kind of vagueness discussed in the last section. At t hat stage I was assuming we knew what kind of organizational complexity was at issue. The only problem was how much of it fish and Proxima Centaurians needed to qualify as conscious. But now we are facing the possibility that there is simply n o fact of the matter about what kind of physical or structural features you need to qualify as conscious, let alone how much.

Even this needn't make us reject talk of consciousness altogether. Maybe consciousness isn't an explanatorily im portant property, the kind of property that enters into laws and serious explanations. But the concept can still be useful in characterizing humans and closely related beings. We might draw an analogy with concepts like good-looking, or witty. These are perfectly useful concepts, and indeed ones which play an important role in human affairs. But nobody would think that they cut nature at the seams, or that it made any significant sense to apply them to beings like fish or toads or Proxima Centaurians.

This view of consciousness may seem to have awkward moral consequences. For questions about consciousness often have moral significance. Whether fish are conscious, for example, seems crucial to the issue of how we should treat them. But if there is no fact of the matter as to whether they are conscious, then doesn't it follow that that there is no right and wrong about how to treat them?

I agree that the position I have reached does have unexpected moral consequences. But I don't think that this shows there is anything wrong with the position. Rather, the position helps us to think better about certain moral questions. I take it that the consciousness of fish and similar beings can only be morally important if there is a definite fact of the matter. If there isn't a definite fact of the matter, we will do better to base our decisions about fish directly on information about the organization of their brains and nerv ous systems, and not on the supposed further issue of whether this physical organization makes them conscious. Indeed, the idea that this is a further issue of moral importance here seems to me not only theoretically misguided but morally dangerous.

Perhaps we might be persuaded by the physical facts that it is wrong to injure certain beings, even though we felt unsure, prior to addressing this moral question, whether they should be deemed conscious. In such a case, should we count them as conscious because we regard them as objects of moral concern? I am sure that we would do so in practice. It may seem odd to hold that certain beings might be conscious because they are morally significant. But the thought i sn't that how it is for them depends on the moral conclusion -- merely that the moral conclusion would give us a motive for refining the indeterminate notion of consciousness in such a way as to include them in the category of conscious beings .

## 4.10 Pains, Shapes and Colours

In this final section of this chapter I want to return to such specific mental states as pains, tickles, visual experiences and emotions, and consider whether these states are

det erminate, even if consciousness is not. For nothing in the last section rules out our identifying these specific mental states with specific physical or structural properties, thus making it definite which beings have them, even if there is no way o f doing this for the overall genus of consciousness. In the terms used earlier, perhaps there are physical equivalents for the determinates like pains, sensory experiences, emotions, and so on, even if there is none for the determinable property of consciousness itself.[24]

Apart from its intrinsic interest, this possibility would make a difference to the moral issues touched on at the end of the last section. It wouldn't matter too much if there is no prin cipled basis for deciding whether fish are conscious, if there is a fact of the matter on whether they feel pain.

However, when we investigate this issue, we shall see that there are problems about projecting even such specific conscious states as pain or colour experience onto beings other than humans or higher mammals. For once we abandon the seductive picture which identifies these states with different kinds of inner light, as I have argued we must, then we must face up to the p ossibility that there is nothing else to decide whether some alien being has the same experience as you have when you see something red.

In a sense such specific states as pains and colour experiences raise a converse problem to that rais ed by the generic propery of consciousness. In the case of the generic property, we started with those states which the test of first-person accessibility identifies as conscious, and asked what phyicalistically acceptable property might tie them to gether. The problem was that there may not be any such property, since the different species of human consciousness are so various. On the other hand, if we start with the states we identify as pains, or experiences of red, and so on, the diff iculty isn't so much that they may share no physical features, but that they seem to share too many.

Let me explain. It seems likely that human beings who share pains, or colour experiences, or other sensory states, will do so becau se they have determinate physical properties in common. So far, so good for physicalism. But the trouble is that it also seems likely that such physical commonalities will appear at a number of different levels of abstraction. For exampl e, it may be that two human beings who are both in pain will both have certain kinds of nerve cells firing. But, if so, then they will also share further properties, such as the functional property of having-some-property-which-plays-a-certain-causa l-role. The propblem for physicalism is to decide which of these competing properties pain is identical with.

Lycan (1987) has emphasized that there are likely to be a large number of different levels at issue here, starting with ve ry strictly physical levels, which are describable only in the language of fundamental physical science, through physiological levels, and on to various functional levels, which will themselves be distinguished by the fine-grainedness of the causal role t hey involve. I think Lycan is quite right about this. But for my present purposes nothing will be lost if we revert to the familiar philosophical oversimplification, and pretent that there are only two competing levels at issue, which we can t ake to be the physiological level ("C-fibres firing", to adopt the conventional philosophical shorthand for the physiology of pain) and the folk-psychological functional level ("a state which mediates between bodily damage and the desire to avoid the caus e thereof").

As I observed in chapter 1, there is an obvious rationale for identifying mental states with functional states rather than physiological ones. Namely, that the choice of physiological states would have the "chauvinist" implication that beings with different physiologies, like toads, perhaps, or silicon-based Proxima Centaurians, certainly, could not share our mental states. Yet it seems unreasonable to conclude that Proxima Centaurians cannot believe that the univ erse is expanding, or, for that matter, that they cannot feel pains, just because they are made of silicon and not carbon.

Yet in the case of conscious mental states, states that it is like something to have, there are also strong contrar y intuitions in favour of the equation with physiological states. These intuitions are best elicited by spectrum-inverting thought-experiments. Imagine that you have your retina altered at birth so that you respond physiologically to green obj ects in the way other people respond to red objects. After the operation you are then raised normally, so that you learn to call red objects "red", post letters in post boxes, eat red and not green tomatoes, and so on. In consequence, the stat e produced in you by red objects plays the same causal role as normal people's experiences of red. But the physiology of this state will be like the physiology of normal people's experiences of green. What will it be like when you see a red ob ject? A widespread intuition is that it will be like most people's experience of green. According to this intuition, the subjective nature of your colour experience is fixed by what physiological processes are taking place in your brain, and n ot by what causal role those processes play.[25]

So there seem to be two conflicting intuitions: the anti-chauvinist intuition that wants the Proxima Centaurian to share our mental states, and so equates those states with f unctional states; and the spectrum-inverting intuition that wants people with abnormal retinas to see red where we see green, and so favours the equation with physiological states.

David Lewis (1980) has developed a theory which aim s to accommodate both these conficting intuitions. In Lewis's view, experiences go with physiology for similar beings, but with functional role for different kinds of beings. Lewis considers pain rather than colour experience. He imagine s a human (a "madman") who is spectrum-inverted with respect to pain. The madman is arranged so that the physiological state which realizes pain in normal humans is produced in him, not by bodily damage, but by moderate exercise on an empty stomach; and it doesn't cause him to writhe or try to alter the state, but rather to snap his fingers and think of mathematics. Lewis takes it that the madman will share the experience of pain with normal humans. So pain goes with physiology for humans.

But Lewis does not therefore think that an extraterrestrial being (a "Martian") cannot feel pain. He takes it that a Martian will feel pain just in case it is in the physiological state that realizes the functional role of pain in normal Martians. So a normal human and a normal Martian who both feel pain will share the functional state of being-in-some-state-with-the-relevant-causal-role. Pain goes with functional role for normal beings from different species.&n bsp; (Within the Martian species it goes with physiology again: there could be a mad Martian who feels a pain, not because any of its states play the functional role of pain, but because it is in the physiological state which plays that role in norm al Martians.)

The attractions of Lewis's theory are obvious. It accommodates the intuition that the experiences of spectrum-inverted people depend on their physiology, but avoids the chauvinist consequence that beings of other speci es cannot share our experiences.

It does, however, have an odd consequence. Imagine that Martians and humans are similar enough to interbreed, in virtue of the fortunate fact that their genes are effectively identical; the onl y substantial exceptions are the genes that direct the development of the pain mechanism, where, as it happens, the Martian genes are dominant.

So a Martian-human hybrid would have its pain mechanism realized by Martian rather than human physiology. Now imagine that such a hybrid exists, and that its pain mechanism is activated. Is the hybrid in pain? If we count it as a Martian, then it will be: for it will be in the physiological state that realizes the role of p ain in normal Martians. But if we count it as a human, it won't be: for, although it is in a physiological state that plays the functional role of pain, this isn't the state that plays that role in normal humans.

Lewis is not unaware that his theory has this kind of consequence. Although he does not consider such an extreme case, he does observe that attributions of experiences will depend, given his theory, on which populations we assign individuals to; and he adm its that such questions of classification will not always admit of hard-and-fast answers.

Still, even if Lewis is aware of it, this consequence is still pretty odd. Surely, one feels, whether a given being is in pain is a determinat e matter, quite independent of what population we might choose to classify it under. (The hybrid's state isn't going to stop hurting, just because the Earth Government changes its immigration regulations to allow that a single human parent qualifies you as human.)

Odd as this consequence is, I don't think that it should lead us immediately to dismiss Lewis's theory. It is possible that our conviction there is a fact of the matter about alien pains stems from the antipathetic f allacy and the associated picture of extra inner lights. For, if pain were an extra inner light, separate from the physics of the brain, then it would in principle be determinate which brains were illuminated by it. But if there is nothing the re, apart from the physics of the brain, then it may be indeed be arbitrary how to classify beings whose brains are like ours in some respects, but not in others.

I have illustrated this possibility with respect to pain, as this is the ca se that Lewis focuses on. But in fact pain is a somewhat unconvincing example of the possibility. While I do think that there are some sensations whose possession is an indeterminate matter, I don't think that pain is one of them.

T his is because I do not think that the intuitions in favour of identifying pains with physiological states carry much conviction to start with. Let us go back to Lewis's madman. According to Lewis, the madman's pain is caused, not by injury, b ut by moderate exercise on an empty stomach. And it doesn't make him writhe or want to alter his state, but simply to snap his fingers and think of mathematics. Given all this, it doesn't seem to me to make much sense to say the madman is in p ain. The madman may share the physiology of normal humans in pain. But if this

physiological state causes the madman no discomfort, if he lacks all inclination to make it go away, then I'm inclined to say that it doesn't hurt, that it's just not a pain.

The concept of a conscious pain, it seems to me, is the concept of being-in-a-state-which-disposes-you-to-certain-sorts-of-behaviour. Something just isn't a pain unless your initial reaction is to get rid of it. If this is right, then pains must be equated with functional states, rather than physiological ones. So "madmen" and "mad Martians" are not in pain, even though they share the physiology of their normal conspecifics. And this now removes the earl ier indeterminacy: the human-Martian hybrid is unequivocally in pain, however we classify it, for its state plays the functional role of pain.

This disambiguation may of course still leave us with a penumbra of vagueness. ; Even if pain is firmly tied to functional role rather than physiology, there may remain an indeterminacy about how complex this functional role has to be before it qualifies as pain. But vagueness is a different issue, as we saw earlier. Our current concern is what kind of physical or stuctural complexity pain should be identified with. We can have a definite answer to this question even if we are vague about how much of that complexity is needed..

Which other sensatio ns are like pain in being conceptually tied to behaviour? Sensations like these will be unequivocally identifiable with functional rather than physiological states, and in consequence their ascription to beings other than ourselves will be determina te, up to the boundaries of vagueness.

There is good reason to regard visual experience of shapes as like pain in this respect. A number of recent works have focussed on such experiences, and their arguments strongly support the vie w that visual experience of shape goes with functional dispositions to behaviour rather than with the phsyiology of the normal viewer. A test case would be a person who is in the physiological state that normally goes with seeing something square, b ut tries to draw the shape in question by making circular arm movements. Intuition strongly favours the view that this person must have the conscious experience as of seeing something circular, and thus supports the identification of the experience with functional role rather than physiology.[26]

Indeed, in the case of spatial perception, there seems to be direct empirical evidence in favour of a functional over a physiological identification. I am thinking here of th e well-known psychological experiments in which subjects wear glasses with "inverting lenses".[27] When they first wear the lenses, subjects faced with an upright drinking cup, say, will have both the physiology, and the dispositions to beh aviour, that normally go with an upside-down cup. Accordingly, we can all agree that at this stage the subjects see the cup as upside down. But after a while such subjects learn to adjust their behaviour, so that they come to behave in the way appropriate to upright cups, even though they still have the physiology that normally results from upside-down cups. And at that stage they then say that the cup "looks the right-way up" again. This obvious fits with the thesis that conscious spatial perception is tied to behaviour rather than physiology.

In fact this experiment is less straightforward than it seems. For the inverting-lens experiment doesn't so much test the thesis that spatial perception is tied to beh aviour (after all, I am treating this as a conceptual truth), as the conjunction of this thesis with the further assumption that subjects can tell what kind of experience they are

having, even after they have been turned into "spatial madmen". To co nfirm this, note that somebody who holds that spatial perception goes with normal physiology, rather than with normal behavioural function, can accommodate the inverting-lens experiment simply by arguing that retrained subjects can no longer be relied on to report accurately which how things look to them.[28]

Now consider colour experiences. In this case it seems unlikely that there is any conceptual tie between seeing something as red, say, and behaving in any particular w ay. When I introduced the colour-spectrum-inverting thought experiment earlier in this section, I said that after the operation you would "call red objects 'red', post letters in postboxes, eat red and not green tomatoes, and so on". Most of t he behaviour involved in this functional characterization (saying "red", using red postboxes) depends on nothing more than social convention, and so can scarcely be part of what it is to see red. (We don't want to say that you can't see red unless y ou know the English word "red".) And the non-conventional behaviour associated with seeing red (eating tomatoes and similar fruit) still seems too thin and topic-specific to tie down the experience.[29]

So colour experience s are different from pains and spatial perception. I do not want to deny that such experiences as seeing red have a characteristic functional role. After all, common sense criteria, which define the functional role for red, are clearly suffici ent in practice to decide which human beings are experiencing red. (In this connection we should not forget the central fact that red objects normally cause red sensations.) But, by contrast with pains and spatial perceptions, colour experienc es do not have a stock of non-conventional desires or actions to call their own. And, because of this, it seems unconvincing to argue, as we did for pains and spatial perceptions, that colour experiences are determinately tied to functional roles, r ather than to physiology. Where there is a direct link between an conscious experience and something we non-conventionally do, then it seems natural to hold that this functional link fixes the nature of the experience. But with experiences whi ch lack any such intrinsic tie to action, there seems to be no corresponding rationale for holding that functional role, rather then physiology, determines the experience.

So I conclude that with colour experiences (and similarly for tast es and smells[30]) there is a real indeterminacy about how to project our categories beyond the case of normal humans. As long as the physiology and the functional role continue to go together, then there is no problem. But when we ha ve one without the other, as with the subject of the spectrum-inverting operation (the "colour madman"), -or a Martian who comes to earth and learns to make our colour discriminations, then I don't think there is any fact of the matter about whether they have the same experiences as us.

There is still David Lewis's strategy, which decides such cases by seeing whether the difficult individuals share the physiology of the functionally normal members of their group. But then, as we saw , it may be indeterminate which group we should consider the difficult individuals to be part of. Lewis's strategy does place some extra constraints on our ascriptions of colour experiences to difficult cases. But, by making such ascriptions d epend on assignments to groups, Lewis in the end only hides the underlying arbitrariness of experiential classifications under the cloak of a different arbitrariness.

I realise that some readers will think it ridiculous for me to suggest that it is an arbitrary matter whether or not colour madmen are counted as have the same experiences as the rest of us. (Surely either they do or they don't). But let me recall a point I made at the end of the last section. I am not sugg esting that how it is for the colour madman will depend on how we classify his experience. Of course it won't. My claim is only that it is indeterminate whether the madman's experience is the same kind of experience as our experience of red.&n bsp; That is, I don't think that there's anything lacking in the colour madman. It's just that the notion of sameness of colour experience breaks down when we come to such cases.

No doubt some readers will find even this absurd.&nbs p; Even if I am not saying that we can alter feelings by linguistic fiat, isn't it bad enough for me to be saying that experiential comparisons are indeterminate? Take one of the madman's colour experiences. Now imagine what it's like to see a bright red tomato. Surely the madman's experience is either like that, or it's not. What could be simpler?

But I don't think think it is that simple. The reason it seems simple is that we naturally suppose that, when we have (or imagine) a visual experience, we switch on an inner light. And so we all we need to do is compare that shade of inner light with the shade illuminating the madman's mind. But there isn't any such inner light. There are just the physical and structural features of the relevant brains, some of which we share with the madman, and some of which we don't. So our conviction that either the madman must feel the same or feel different is based on a false picture. Wittgenste in had a good analogy: "You surely know what 'It's 5 o'clock here' means; so you also know what 'It's 5 o'clock on the sun' means. It means simply that it is just the same time there as it is here when it is 5 o'clock." (1953, §350.)

1. This follows standard practice in this area: see Horgan (1984, pp 147-8) and Tye (1986, p 1).

2. Moreover, most of the arguments between dualism and physicalism arise in exactly the same way between dualism and any more general non-physi calist "objectivism" about conscious mental states. I shall formulate the issues as a matter of physicalism versus dualism, however, since I think there are good arguments -- namely, those presented in chapter 1 -- for preferring physicalism to othe r kinds of objectivism. Even so, much of what follows should also be of interest to objectivists who are not physicalists.

3. An immediate qualification is needed here. For we can obviously imagine complex experiences, like seeing a unicor n, as long as we've previously experienced the elements separately. And we can perhaps imaginatively extrapolate to intermediate experiences, like imagining a colour which is spectrally between others we have previously experienced. But these possibilities are clearly consistent with the general thesis that the brain needs to acquire the materials for the replicas from previous experiences, and so in accord with the fact that we can't imagine experiences of a radically unfamiliar kind, like se eing colours at all, or echolocating, until we have actually had those experiences.

4. The view that Mary acquires new abilities rather than new knowledge is urged by Lewis (1988) and Lemirow (1990).

5. Note that the alternative non-physicalist ac count, in terms of phenomenal property P, does nothing at all to explain why the exercise of our recreative abilities should in some sense make us re-experience the original mental state. Thinking of or remembering something as an event with some pr operty P can in general have any experiential nature, or none at all. Of course, it could be argued that, in the particular case of some phenomenal property P, thinking of or remembering an event with that property involves recreating in your brain a copy of the experience characterized by the property. But, once this last move is made, then it becomes unnecessary to bring in the phenomenal property P to explain Mary's new imaginings and memories in the first place -- for now we can simply exp lain these imaginings and memories directly, by appealing her recreative abilities.

6. For this terminology, see Mellor (1992, p 11).

7. Lewis (1983, pp 131-2)

8. Again a qualification is needed to accommodate the fact the we can recognize nov el complex experiences, as long as their components have previously been experienced: in such cases the brain doesn't need an original complex experience to form a complex template, but only the originals of the component experiences to form templat es of the components.

9. Cf Peacocke (1989, pp 67-9).

10. Isn't Jane ruled out by Wittgenstein's private language argument? Well, she'd better not be, if Wittgenstein's argument is any good, since Jane is clearly possible. I don't thin k there is in fact any tension here. I take the moral of Wittgenstein's argument to be that there must be room for error in people's judgements about their experiences, not that those judgements must necessarily be expressed in a language used by a community. And I see no reason to suppose that Jane cannot make mistakes about her own experiences.

11. This suggestion is central to the response made to Jackson's argument in Horgan (1984).

12. Of course, if Mary isn't a physicalist, then she will be disinclined to make this identification, and will no doubt maintain that the first-person concept she shares with Jane refers to a phenomenal attribute, whereas her scientific concept refers to a physical phenomenon.

13. Won't this r ealization involve some new information, of a kind Mary couldn't have had before her experience? After all, someone who discovers that a = b, where [a] and [b] express two modes of presentation of the smae object, will generally acquire the informat ion that the property invoked by [a] is co-extensional with the property invoked by [b]. So won't Mary acquire the new information that the property of having such-and-such neurones firing is co-extensional with the phenomenal with the phenomenal property of red? However, this argument assumes that Mary's first-person mode of presentation of the experience of red invokes some phenomenal property. In contrast, I have just suggested that this is an indexical construction. If this is right, Mary no more acqu ires new non-physical information than someone who suddenly realizes that it is noon now.

14. Or, if they do involve secondary experiences, as when we think about somebody being in pain, say, by thinking about the visual aspect of their behaviour or b rain state, then they will be different secondary experiences, secondary version of visual experiences, not secondary versions of pain experiences.

15. In line with this, Lycan calls the fallacy the "stereoscopic fallacy".

16. In a generous gestur e of help to his physicalist opponents, Nagel points out that the fallacy in question provides an answer to Kripke's modal argument against mind-brain identity. Kripke (1972) appeals to the principle that identity statements involving rigid designat ors are necessarily true, and then challenges physicalists, who identify mind and brain, to account for the apparent contingency of mind-brain identity statements. Nagel's suggestion is that, instead of looking for some non-rigid way of reading the terms in these statements, which is how we account for other apparently contingent identity statements, like "water = H20", physicalists should simply explain the appearance of mind-brain contingency by reference to the fallacy that makes us so convinced that mind and brain are different to start with. On this suggestion, we won't explain away the appearance of contingency by finding some non-rigid reading which is violated in other possible worlds, as we do with the other cases. Rather, we si mply account for the appearance of contingency by explaining why we are so disinclined to accept mind-brain identities in the first place. I agree with Nagel that this is the right way for physicalists to respond to Kripke's argument.

17. Ned Bl ock offered this story to me; I don't know where it originated.

18. Cf McGinn (1982, pp 13-14).

19. An alternative physicalist response to the intuition that consciousness is not vague would be to seek some physicalist characteristic A which does provide a sharp dividing line. But this strategy strikes me as unlikely to succeed.

20. Thus McGinn, ibid: "The emergence of consciousness must rather [unlike the emergence of life] be compared to the sudden switching on of a li ght . . ."

21. No doubt the idea of an inner light as such a property is partly responsible for our having this notion of consciousness. But it would be a pity, I think, to build the inner light itself into our notion of consciousness.

;

22. This is of course the standard objection to self-monitoring theories of consciousness like Armstrong's.

23. Chris Hughes suggested to me that the relevant question is whether the states of toads and similar beings would be first-person acces sible, if they occurred in beings who could introspect, imagine, and so on. But I doubt this really removes the indeterminacy. Exactly which counterfactual possibilities are we supposed to consider? Is the question whether we humans coul d introspect toad states, if they occurred in us? Or are we supposed to consider super-toads, who stand to toads as we do to monkeys? But then what is supposed to stop us considering super-trees, say, or super-stones?

24. If so, couldn't w e just disjoin the determinates to get a physical equivalent for the determinable? But we still lack a principle to generate all instances of the genus. We might be able to cover all human determinates by brute enumeration. But, in the a

bsence of any property equivalent to consciousness as such, there will be nothing to decide which states of the Proxima Centaurians should be included in the disjunction.

25. Note how this thought experiment differs from the traditional vesrion, in wh ich physical identicals have inverted spectra. I take this traditional version to be discredited by the general arguments for physicalism. The modern version, by contrast, involves people with different physiologies. So it doesn't presen t a problem for physicalism as such, but only for functionalism.

26. See in particular McGinn (1989, pp 58-94), Davies (1992). It should be said that this literature is more concerned with whether spatial experiences have broad or narrow content s than with their phenomenal identity as conscious states. But the two issues are connected. See Davies, op cit, sect III. Davies's concern with the issue of content leads him to distinguish carefully between internal inclination to beha viour, and actual external behaviour (with phenomenology going with the former, and content with the latter). But from our perspective these are alike matters of functional role.

27. See Gregory (1977, ch 12).

28. Some readers mig ht feel it would make more sense for sensory states to be incorrigibly tied to introspective reports, instead of to further links to behaviour. But it seems wrong to rule out introspective mistakes in this way. Apart from anything else, there are brain abnormalities that seem to affect introspective abilities rather than anything else. Morphine is a good example: it makes people say that the pain is still there (even though they don't mind it); but, if Daniel Dennett (1978, p p 208-9) is right, these people are not in pain on anybody's account, since not only do they fail to display pain behaviour, but they also lack the normal physiology of pain.

29. Janet Levin (19xx) takes the contrary view that even experiences of colo ur, taste and smell might be definable by their links with non-conventional behaviour. In line with this, she suggests that spectrum-inverting operations with these modalities might turn out like the "inverting lens experiment": at first the m admen's responses will involve both the "wrong" physiology and the "wrong" behaviour ; but after a while they will adjust their behaviour to make it "right"; and then things will seem conscioulsy "right" to them once more. This argument raises a number of issues. Central is whether the inversions would lead to systematically inappropriate behaviour of a kind that could be remedied by a systematic (rather than picemeal) rewiring of our behavioural responses. I agree with Levin that if this were so, then it would be appropriate to describe the rewired people as having the "right" phenomenology again. What I doubt is whether there are such systematic links between the relevant experiences and behaviour to start with.

3 0. Sounds raise yet further issues, which I shall not pursue. Note that the categorization of experiences in terms of their links with behaviour does not coincide with the division between primary and secondary qualities. I have argued that a constitutive link to behaviour is present both in experience of shape, which is a primary quality, and painfulness, which can be thought of as a (hyper-)secondary quality.