# *The Subjective Qualities of Experience*

MICHAEL TYE

It has frequently been urged that there are facts about our sensory experiences, both perceptual and bodily, which no amount of physical information, including that of a functional sort, can capture.[1] The facts I am referring to are familiar to all of us in our everyday conscious lives: they pertain to the subjective phenomenal qualities or 'qualia', as they are sometimes called, which characterize our pains, our itches, our sensations of colour, our feelings of love, hate, and despair. I believe that the arguments adduced for the view that these facts lie outside the physicalist's net are unsound. I also believe, however, that at least one of the arguments, which I call 'the Argument from Knowledge', deserves a great deal more careful attention than it has received heretofore. This argument in one form or another has exerted a powerful influence on many philosophers, and it has been a thorn in the side of physicalism for a number of years. In what follows, my primary concern is to present a conclusive refutation of this argument.

The structure of the paper is as follows. In Section I, I distinguish one main strand of argument for the claim that there are subjective experiential facts which physical information cannot capture. I call this 'the Argument from Possibility'. I briefly take up a number of variants on this argument and I suggest that they have little force. In Section II, I turn to a detailed examination of my primary topic, the Argument from Knowledge, versions of which constitute the second main strand of argument from qualia against physicalism. I here give clear analyses of knowing what it is like to undergo such-and-such a type of subjective experience, and knowing what a given token experience is like. I maintain that once these concepts are properly understood the physicalist has nothing to fear from the Argument from Knowledge. I also show that the physicalist can accept the thesis, dear to the hearts of empiricists, that there are (or can be) terms for the subjective qualities of experience whose meanings cannot be fully understood

[1] Here, and throughout the paper, I use the term 'physical' in a broader way than is usual. I adopt this usage partly for ease of exposition and partly because it seems to me that since the non-causal descriptive vocabulary in standard functionalist analyses is physical in a narrower non-functional sense, the functionalist analyses themselves may reasonably be classified as physical too, *provided that* the first-order properties (or states) these analyses quantify over are physical. On this account, the claim that functional properties are physical properties has a contingent status, *if*, as some functionalists insist, it is metaphysically possible for their analyses to be satisfied by non-physical objects, e.g. souls. Those who object to my usage of the term 'physical' may view it as stipulative. The same comments apply *mutatis mutandis* to my usage of 'physicalist' and 'physicalism'.

by persons who have not experienced the relevant qualities. Finally, in Section III, I briefly address certain objections which may appear to threaten the central claim in my defence of physicalism.

## I

The Argument from Possibility is any argument of the following general form. Let '*Q*' be a rigid designator for a given qualitative character, for example, the burning hurtfulness of a particular pain or the bright red look of a particular visual field. Let '*N*' be a rigid designator for a given physical property with which the physicalist wishes to identify *Q*. Then if *Q* is identical with *N*, it is metaphysically necessary that *Q* is identical with *N*. But it is not metaphysically necessary that *Q* is identical with *N*. Hence *Q* is different from *N*.

The crucial claim in this argument is the second one. Why should we accept that it is metaphysically possible that *Q* is not identical with *N*? Different answers to this question, along with different referents for the designator '*N*', result in the different versions of the Argument from Possibility.

Consider first what has come to be called 'the Absent Qualia Argument'.[2] This argument, which is usually directed against functionalist accounts of qualia, has as its major premiss the claim that two beings might be in states which are functionally identical in every respect and yet which are such that one has a phenomenal character and the other lacks it. Thus, it is hypothesized, there might be a robot so designed that it reacted in just the way I react whenever my body is damaged and I feel pain, both with regards to my observable behaviour and with regards to my other mental states, but which was not subject to the awful hurtfulness which characterizes my pain experience—indeed, which was not subject to any feeling in its inner state at all. More generally there might be a robot—an 'imitation man'[3]—whose inner states were causally related to stimuli, other inner states, and behaviour in just the ways all my sensory states are causally related to stimuli, other sensory states, and behaviour but which was entirely insentient, i.e. which never underwent feelings or experiences of any variety.

The Absent Qualia Argument, as sketched above, is a variant upon the general Argument from Possibility for the case where '*N*' designates a functional property. It is metaphysically possible that *Q* is not identical with *N* according to the Absent Qualia Argument, since it is metaphysically

[2] See e.g. Keith Campbell, *Body and Mind*, Garden City, NY, Anchor, 1970, pp. 100–4; Robert Kirk, 'Sentience and Behavior', *Mind*, 1974, pp. 43–60; Ned Block, 'Troubles with Functionalism', in C. W. Savage, ed., *Minnesota Studies in the Philosophy of Science*, vol. 9, Minneapolis, University of Minnesota Press, 1978, pp. 261–325; Colin McGinn, *The Character of Mind*, Oxford, Oxford University Press, 1982, pp. 35–6.

[3] The expression 'imitation man' is Keith Campbell's. See his *Body and Mind*.

possible for $N$ to be instantiated without *any* qualitative character. This latter claim is itself supported by advocates of the possibility of absent qualia by an appeal to imaginability. Since we have no difficulty in imagining an imitation man, it is held that we should have no resistance to the major premiss.

Sometimes the Absent Qualia Argument is extended to cover the case where '$N$' designates a neural property or state.[4] Thus, it is sometimes argued that there is no metaphysical impossibility in your feeling nothing at all while you are in the brain state which, in the actual world, is correlated with your experiencing $Q$. Why is this possible? Well, again it is imaginable. Simply imagine having a cerebroscope attached to your skull which registers the fact that $N$ is tokened in your brain while you yourself lack any sensation or feeling.

Another form of the Argument from Possibility is the Inverted Qualia Argument.[5] This argument is directed against functionalist approaches to qualia. The crucial claim now is the claim that two inner states might be functionally equivalent in every respect yet qualitatively different. Thus, you might have abnormal colour vision which is responsible for your visually experiencing, in the presence of *green* objects, the qualitative character which normal perceivers visually experience in the presence of *red* objects; yet your peculiarity in vision would be completely undetectable in either your linguistic or non-linguistic behaviour, if your colour experiences were systematically inverted in an appropriate manner. Here again we can see the Argument from Possibility at play for the case where '$N$' designates a functional property. This time it is possible that $Q$ is not identical with $N$, since $N$ might be co-instantiated with a different phenomenal quality $P$. This latter claim of possibility is based, as before, upon what is imaginable. Advocates of the Inverted Qualia Argument maintain that cases of inverted functionally-equivalent qualitative states are perfectly imaginable and hence possible.

It is worth noting that a modified version of the Inverted Qualia Argument could be developed for the case where '$N$' designates a neural property. In this development, the crucial claim now becomes that two inner states might be of the same neural (as opposed to functional) type and yet be qualitatively different. For example, there might have been a person whose brain instantiated the neural property, $N$, mine instantiates whenever I see red, say, but whose experience had a different qualitative character (the one I experience whenever I see green). Given this claim, it follows that it is metaphysically possible that $Q$ (in this case, the qualitative character I experience in seeing red) is not identical with $N$.

[4] See e.g. Saul Kripke, 'Naming and Necessity' in D. Davidson and G. Harman, eds., *Semantics of Natural Language*, Boston, Reidel, 1972, p. 335.

[5] See e.g. Keith Campbell, *Body and Mind*, p. 74; Colin McGinn, *The Character of Mind*, p. 35; Sydney Shoemaker, 'Functionalism and Qualia', *Philosophical Studies*, 1975, pp. 291–315.

A third and final form of the Argument from Possibility is the Multiple Realizability Argument.[6] This argument, which is normally put forward against the identification of qualia with neural properties, rests upon the claim that qualia can be multiply realized, that the neural state or property, $N$, which realizes a given quale, $Q$, in me might be different from the neural state or property $M$ which realizes that quale in you or in creatures of other possible species. It is then metaphysically possible that $Q$ is not identical with $N$, since it is metaphysically possible that $Q$ is instantiated without $N$.[7] This latter claim, like the earlier claims in the Absent Qualia and the Inverted Qualia Arguments, is itself supported by an appeal to imaginability. Sitting in our armchairs, we can easily imagine $Q$ instantiated without $N$, since we can imagine feeling the characteristic feeling which '$Q$' designates while viewing an auto-cerebroscope which registers the presence of some neural property other than $N$.

These three versions of the Argument from Possibility—particularly the Absent Qualia and Inverted Qualia Arguments—have received extensive analysis in the recent literature,[8] and I cannot possibly discuss them fully in the present paper. Instead, I shall focus on the question as to whether these arguments refute the claim that qualia are identical with neural properties. What I shall suggest is that this claim remains unscathed, and hence that the three versions of the Argument from Possibility do not establish that qualia cannot be captured in the physicalist's net.

Consider again the general assertion that it is metaphysically possible that $Q$ is not identical with $N$, where $N$ is a neural property. We are urged to accept this assertion *either* on the grounds that we can imagine a possible world $W_1$ in which $Q$ is co-instantiated with some other neural property, $M$, instead of $N$ (as in the Multiple Realizibility Argument) *or* on the grounds that we can imagine a possible world $W_2$ in which $N$ is instantiated without $Q$ or any other phenomenal quality (as in the second form of the Absent Qualia Argument) *or* on the grounds that we can imagine a possible world $W_3$ in which $N$ is instantiated without $Q$ but with some other phenomenal quality (as in a hypothetical, modified version of the Inverted Qualia Argument). Obviously these appeals to what we can imagine will be successful, *only if* imaginability entails possibility. This presupposition is widely accepted, and it strikes me as entirely reasonable. Where a person claims to be imagining something which is (unknown to him) impossible,

[6] See e.g. Saul Kripke, 'Naming and Necessity'; also Colin McGinn, 'Anomalous Monism and Kripke's Cartesian Intuitions', *Analysis*, 1977, pp. 78–80.

[7] In some extreme versions of the Multiple Realizability Argument, it is held that $Q$ could be instantiated without *any* physical property, as, for example, in a disembodied soul.

[8] See e.g. Bernard Harrison, *Form and Content*, Oxford, Basil Blackwell, 1973; Sydney Shoemaker, 'Functionalism and Qualia'; also 'Absent Qualia are Impossible—A Reply to Block', *Philosophical Review*, 1981, pp. 581–99; also 'The Inverted Spectrum', *Journal of Philosophy*, 1982, pp. 357–81; Ned Block, 'Are Absent Qualia Impossible?', *Philosophical Review*, 1980, pp. 257–74; William Lycan, 'Form, Function, and Feeling', *Journal of Philosophy*, 1981, pp. 24–50.

surely we would deny that he has succeeded in imagining what he claims. If, for example, a man without scientific knowledge claims to be imagining that gold has atomic number 80 (rather than its actual 79) what I think we would say he *really* imagines is that some substance with the superficial observable qualities of gold has atomic number 80 (rather than 79), and *that* is something quite different. The question we have to address, then, is whether $W_1$, $W_2$, and $W_3$ are genuinely imaginable.

The imaginability of worlds $W_2$ and $W_3$ is more contentious than the imaginability of world $W_1$, since many philosophers hold that phenomenal properties supervene upon neural properties in the way that aesthetic or moral properties are sometimes held to supervene on naturalistic properties. For these philosophers, it will not be possible (and so not imaginable) that $N$ is instantiated without $Q$. These complications are best avoided in the present context. Let us, therefore, concentrate on $W_1$.

It seems to me that what we really imagine when we claim to imagine $W_1$ is this: we imagine ourselves undergoing experiences which instantiate the felt quality $Q$, and we imagine ourselves (or others) viewing our brains through cerebroscopes which we (or they) take to indicate the presence of neural property $M$ rather than $N$. But does this suffice to imagine $W_1$? I think not. For the cerebroscopes we imagine may be malfunctioning or $N$ may have the appearance in $W_1$ which we associate in the actual world with $M$.[9]

It may now be replied that we can imagine what we imagine above whilst also imagining (a) that the cerebroscopes are functioning normally without outside interference and (b) that $N$ and $M$ appear in $W_1$ exactly as they would in the actual world. The major problem with this reply may be brought out in the following way. Suppose I claim to be able to imagine that Goldbach's conjecture is false. When pressed for details, I say that I can imagine famous mathematicians excitedly reading the print-out tape from an appropriately programmed computer and exclaiming, 'So there are even numbers which are not the sum of two primes!' Does this make it plausible to say that I have succeeded in imagining that Goldbach's conjecture is false? Obviously not. The computer I imagine may be malfunctioning or the scientists in their excitement may have misread the tape.

Suppose I reply to this point by insisting that the computer I imagine is functioning normally and that the scientists make no mistakes in reading the print-out tape. Have I now imagined that Goldbach's conjecture is false? If so, then my thought experiment entitles me to believe that Goldbach's conjecture is possibly false and hence that the conjecture is in fact false.[10] But, on the basis of my thought experiment, it is evidently wrongheaded for me to say that I may rationally deny Goldbach's conjecture. Surely

[9] Cf Ned Block, 'Troubles with Functionalism', p. 287.

[10] I assume that mathematical hypotheses always have truth-values, and that their truth-values are the same in all possible worlds.

I have no logical right to an opinion on the matter (assuming that I have no other relevant information).

It seems to me that there is a parallel between the above case and the one involving $Q$ and $M$. Our grounds for thinking that we can imagine $Q$ co-instantiated with $M$ rather than with $N$ are no stronger than our grounds for thinking that we can imagine the property of being an even number co-instantiated with the property of being a number which is not the sum of two primes. And just as the latter grounds are insufficient unless we already have independent evidence that the properties can be co-instantiated so too are the former. Where I suggest we go wrong in our thought experiments is in the belief that if it seems to us that we have imagined things $A$, $B$, $C$, . . . occurring together in some possible world $W_n$ it automatically follows that we have really done so. In some cases we may have failed to imagine one of $A$, $B$, $C$, . . ., instead imagining something different but with the same appearance. In other cases we may have succeeded in imagining all of $A$, $B$, $C$, . . . but not together in a single possible world. This may perhaps seem counter-intuitive. But once we acknowledge the existence of a posteriori metaphysical necessity,[11] we should realize that our a priori intuitions about what is imaginable are no longer always reliable. What a priori seems imaginable may well turn out not to be.[12]

I conclude that variants on the Argument from Possibility do not demonstrate that there are experiential facts which are non-physical. I turn next to the Argument from Knowledge.

## II

Suppose that Jones is an extraordinarily brilliant scientist of the twenty-third century, who has acquired exhaustive knowledge of what goes on in us physically when we see colours and use colour words. Suppose also that Jones is congenitally blind and that, although he has recently agreed to take off enough time from his scientific studies to undergo a corneal transplant, he has not as yet been operated upon. Is there anything Jones does not know that his fellows with normal colour vision do? The natural obvious response to this question is 'He doesn't know what red, green, blue, etc., look like.' But *ex hypothesi* he has all the knowledge there is to have about what is going on in the optic nerves, the brains, the central nervous systems, and the vocal chords of his fellows when light of various wavelengths strikes their eyes and they respond by saying 'This is red', 'That is green', 'That is blue'. It follows that there is knowledge of experiential facts which lies beyond the reaches of any physicalist theory.

[11] For powerful arguments in support of such necessity, see e.g. Saul Kripke, 'Naming and Necessity'; also Hilary Putnam, 'Meaning and Reference', *Journal of Philosophy*, 1973, pp. 699–711.

[12] For a lengthy general discussion of imaginability and possibility (one with which I find a good deal to agree), see George Seddon, 'Logical Possibility', *Mind*, 1972, pp. 481–94.

This, in brief, is what I am calling 'the Argument from Knowledge'. It has, I think, substantial intuitive force, since it is very hard to deny that the blind Jones does not know what red, green, blue, etc., look like. After all, when Jones has recovered from the operation and acquired normal sight he will surely *learn* something new about visual experience, about how the colours appear.

The argument I have sketched has a long history. Versions of it are to be found in the early Russell,[13] where they are associated with the famous doctrine of knowledge by acquaintance and the thesis that phenomenal colour words, for example, 'red' in statements of the form 'Person, *P*, has a visual sensation of red' have meanings which cannot be grasped by persons who have not experienced the relevant subjective qualities. In the more recent past, the argument, or something very like it, is to be found in the writings of numerous philosophers.[14] In some versions of the argument, the primary concept is that of understanding. Thomas Nagel, for example, in his much-discussed paper 'What is it like to be a bat?'[15] presents a version of the argument which may be reconstructed as follows: to undergo an experience is for there to be something it is like for the subject of that experience. Now what it is like to have a given set of experiences can only be understood from a single (type) of point of view, that conferred by being oneself the subject of a similar set of experiences. So, for example, what it is like to have the sorts of experiences a bat has can only be understood from a bat's point of view, which certainly is not our point of view. But physical facts can be understood from *any* point of view, irrespective of the phenomenology of the experiences of the creatures occupying the points of view. Hence, there are facts about experience which are not physical facts.

It is obvious that this argument is very closed related to the one I stated earlier. Indeed Nagel's version can be restated (without loss of content) as an argument from knowledge in the following manner: we cannot (even in principle) know what it is like to undergo the experiences a bat undergoes since we ourselves, given our neurophysiology, cannot be visited by similar experiences. But we can (in principle) know all the physical facts (including functional ones) about bat brains, bat bodies, and bat sonar. Physicalism, therefore, is incomplete.

Stated in this manner, it appears that there is no significant difference between Nagel's argument and the version which makes reference to Jones.

[13] See e.g. 'The Philosophy of Logical Atomism', in R. C. Marsh, ed., *Logic and Knowledge*, London, George Allen and Unwin, 1956, p. 194.

[14] See e.g. Paul Meehl, 'The Compleat Autocerebroscopist' in P. Feyerabend and G. Maxwell, eds., *Mind, Matter, and Method*, Minneapolis, University of Minnesota Press, 1966, pp. 151–60; M. T. Thornton, 'Ostensive Terms and Materialism', *Monist*, 1972, p. 193; Thomas Nagel, 'What is it like to be a bat?', *Philosophical Review*, 1974, pp. 435–50; Colin McGinn, 'Philosophical Materialism', *Synthese*, 1980, pp. 182–3; Frank Jackson, 'Epiphenomenal Qualia', *Philosophical Quarterly*, 1982, pp. 127–32.

[15] See Thomas Nagel, 'What is it like to be a bat?'

For instead of saying that the blind Jones does not know what the various colours look like it appears that we could equally well say that the blind Jones does not know what it is like to undergo the experiences characteristic of seeing the various colours. Let us, then, return to the original argument, which has the benefit of restricting itself to familiar experiences of a possible member of our own species, and let us begin our response by trying to say what it is to know what a given sort of experience is like.

It might be proposed that the only salient thing Jones lacks is the capacity to undergo the visual sensations or experiences which normal perceivers undergo when they see objects of given colours. This leads to the thought that statements of the type

(1) Person, $P$, knows what it is like to have an experience with a certain phenomenal quality $Q$

may be analysed as

(1a) $P$ has the capacity to undergo experiences with $Q$.

There is an immediate difficulty, however, with this very straightforward analysis. Visual sensations are non-cognitive episodes which happen to young children and adults alike, assuming their receptor systems are matured and in normal working order. They can and do occur whether or not their subjects are introspectively aware of them and they normally carry much more information than can be extracted and exploited by the cognitive centres.[16] Thus, a person who satisfies (1) must have *more* than just the capacity to undergo the relevant experiences. Intuitively, he must also stand in the appropriate cognitive relationship to those experiences either via present introspective awareness or via memory.

These reflections suggest the following analysis as a replacement for (1a):

(1b) Either $P$ is presently undergoing an experience with $Q$ and $P$ is introspectively aware that his experience has $Q$, or $P$ can remember having an experience with $Q$.

Unfortunately (1b) is still not quite right. For (1b) requires $P$ to have correctly identified, either presently or in the past, some experience of his as having $Q$ on the basis of introspection (since $P$ surely cannot remember having had an experience with $Q$ unless he was at some past time introspectively aware that an experience of his had $Q$). Yet (1) requires no such thing. If $Q$ is a complex phenomenal quality, $P$ may know what it is like to undergo an experience with $Q$ by imagining such an experience on the basis of what he remembers of past experiences having (separately) $Q$'s components. Alternatively, if $Q$ is phenomenally simple, for example, a phenomenal colour shade quality, $P$ may perhaps know what it is like to

[16] For more on this conception of visual experience, see Fred Dretske, *Knowledge and the Flow of Information*, Cambridge, Mass., M.I.T. Press, 1981, pp. 135–53.

undergo an experience with *Q* by imaginatively extrapolating from his memories of very similar colour experiences.

These objections can be answered by adding a further disjunct to (1b) as follows:

(1c) *Either P* is presently undergoing an experience with *Q* and *P* is introspectively aware that his experience has *Q*, *or P* can remember having an experience with *Q*, *or P* can remember having experiences with qualities either phenomenally similar to *Q* or phenomenally constitutive of *Q* and *P*, on the basis of what he is here able to remember, can imagine having an experience with *Q*.

On this analysis, correct identification via introspective awareness remains a central requirement in knowing what it is like to have an experience with a given phenomenal quality. Since such identification yields knowledge *that* (i.e. knowledge of fact), it is clear that knowing what it is like, according to (1c), is grounded upon factual knowledge which is obtained in the appropriate manner, namely by introspection. It is also worth noting that (1c) draws a conceptual link between knowing what it is like and the possession of certain abilities. The analysis we have reached, therefore, is really a *hybrid* one, and it has, I suggest, considerable intuitive plausibility. We are now ready to pursue further the case of Jones.

The problem Jones presents for physicalism is supposedly that Jones does not know what it is like to undergo the characteristic experience of seeing red (or blue or green, etc.). But the primary reason that Jones does not know this is that he has never undergone the appropriate experiences. When he *does* undergo these experiences and he identifies them as the experiences they are via introspective awareness, *then* he will know what it is like to have the experience of seeing red, blue, green, etc. So after the operation he will indeed *learn* something. But in learning something he will not come to know facts of a new sort different in kind from those he knew before, or so the physicalist can insist. Rather he will come to undergo experiences of a sort *he* has not undergone before, and by introspectively responding to those experiences, he will come to know facts of an old sort, facts just like those he already knew about the experiences of his sighted fellows, but in a new way.[17] Before the operation he only knew facts of that sort by external physical observation of his fellows' bodies and brains; after the operation he knows them by introspective awareness. The point

[17] Sydney Shoemaker makes this point in 'The Inverted Spectrum'. However, his comments, which appear in a footnote, are extremely brief. They are also associated with an analysis of knowing what it is like to which counter-examples can be constructed. Moreover, there are difficulties brought out below for the stated response to the Argument from Knowledge which Shoemaker does not discuss. Terence Horgan also takes a similar line in his 'Jackson on Physical Information and Qualia', *Philosophical Quarterly*, 1984, pp. 147–51. Horgan's comments are not quite as brief as Shoemaker's but they are still too terse to do full justice to the Argument from Knowledge. Furthermore, there is a significant difficulty for Horgan's discussion. In this connection, see n. 19 below.

here is a subtle one and it requires further defence. I propose to support it by considering a new version of the Argument from Knowledge which appears still harder for the physicalist to answer.

Suppose that Smith is a normal perceiver in the twenty-third century who, as it happens, is facing a red object in good light. Suppose also that Smith is introspectively aware of the visual experience (call it '*e*') which he is presently undergoing and that he believes truly of *e* that it has a certain phenomenal colour content which he rigidly names '*R*'. Then clearly Smith knows of *e* that it has *R*. The knowledge Smith has here is knowledge *of* one particular experience. It is not identical with knowledge of what it is like to undergo *an* experience with content *R*, since the latter knowledge, unlike the former, can be possessed by persons who hold no beliefs at all about the particular experience *e*.

Jones, it may be suggested, cannot know of *e* that it has *R* no matter how much he knows of Smith's body, brain, and behaviour. This is not to say, of course, that had Jones not been blind he could never have known that fact. If, for example, Jones had not needed a corneal transplant and if he had been located next to Smith, facing the very same red object, then he would have known of his own experience that it had content *R* on the basis of introspection, and hence he would also have known of the experience Smith was undergoing that it had that content too (unless he had some reason to think that his colour vision was unlike that of his fellows). The point is that there is something Smith knows which Jones, given his blindness, does not, *despite* his exhaustive knowledge of the physical world. That this is so, it may be argued, is clearly shown by the fact that after the operation Jones will surely *learn* or *discover* something about *e* in particular and its phenomenal content. Indeed how can this be denied? By finding out what he himself experiences facing a red object in good light and by checking that he and Smith are alike physically, Jones will come to know of *e* that it had a certain phenomenal content—the one which Smith called '*R*' and which he now, like Smith, typically experiences in the presence of red objects. This is new knowledge, knowledge he did not have before. It is knowledge of what *e* was like, how *e* appeared, knowledge of *e* that it had specific phenomenal content *R*. The conclusion we reach, then, is that *there are* facts which physicalism leaves out.

It seems to me that this is a powerful argument, and great care must be taken in disarming it. It is also important to notice that this version of the argument, if sound, entails not just that after the operation Jones will discover a *new* fact about one particular experience of Smith's but also that he will discover *new* facts of a more general sort. For if Jones before the operation does not know of *e* that it has *R*, then by similar reasoning Jones for that period does not know of *any* experience which takes place that it has *R*. It must be true, then, that Jones, after the operation, *also* comes to learn various general facts, like the fact that there are experiences having

phenomenal content *R* and the fact that normal perceivers undergo experiences having *R* in the presence of red objects. Hence, Jones *does* discover facts of a new sort, both general and particular, when he gains his sight. Hence, my earlier claim that Jones merely comes to know facts of an old sort in a new way is now completely undermined.

I believe that the way the physicalist should answer this argument is by claiming that the blind Jones *does* know of Smith's experience *e* that it has phenomenal content *R*. This claim is, on the face of it, counterintuitive. Let us see, then, what the physicalist can say to make it plausible.

Consider first the suggestion that Smith and Jones, in knowing of *e* that it has *R*, both know of *e* that it has the content which, in normal perceivers, typically results from viewing red objects in standard circumstances. This suggestion obviously cannot be correct. While Jones, given his wealth of functional knowledge, does indeed have the knowledge cited above, this is not knowledge of *e* that it has *R*. After all, Smith can know of his experience that it has phenomenal content *R* without possessing either the concept of physical object redness or the concept of a normal perceiver, and so without knowing of *e* that its content fits the above causal description. Furthermore, had the world been different in certain ways, content *R* would not have fitted that description. In these circumstances Smith could have known what he actually knows with respect to *e*, namely that it has content *R*, but obviously he could not have known of *e* what is false, namely that *e* has the content which (in normal perceivers) typically results from viewing red objects in standard conditions. We have, then, at least two reasons for distinguishing what Smith knows from what it is claimed he knows above.

Similar reasoning will undercut the suggestion that Smith and Jones, in knowing of *e* that it has content *R*, know of *e* that it has the content which typically causes normal perceivers to believe of objects in their environment that they are red (or to group such objects as being alike in colour). Just what is it, then, that Jones and Smith both know?

Suppose it is said that Jones and Smith both know of *e* that is it is a token of such-and-such a physico-chemical type.[18] Smith knows this fact about *e* via introspection; Jones knows it via information transmitted to him from a cerebroscope. Is *this* claim at all plausible? Surely not. Smith may not have the foggiest idea what is going on in his brain. So, in knowing of *e* that it has *R*, he cannot possibly know of *e* that it is a token of a given physico-chemical type.

Suppose the physicalist now suggests that while *R* is indeed a physico-chemical property instantiated within the brain, no reductive analysis is possible of what Smith and Jones both know. In knowing of *e* that it has *R*, Smith and Jones know just that. On this proposal, the crucial difference between Smith and Jones with respect to *e* is merely that Jones

[18] Here and below I use the expression 'physico-chemical type (property)' narrowly to mean 'physical property of a sort expressed by a first-order predicate of some true theory of physics or chemistry'.

alone knows the nature of the physico-chemical property which is identical with *R*.

It may be instructive here to consider a possible parallel. Suppose Brown is an ordinary man who has no scientific knowledge at all but who reliably discriminates water from other substances. Clearly there will be occasions on which it is true to say that Brown knows of the liquid in a given container that it is water. But Brown will not know of the liquid in the container that it is $H_2O$, since Brown does not know the first thing about the micro-structure of water. Since 'water' is, let us agree, a rigid name for $H_2O$, just as '*R*' is a rigid name for a property which is, according to the physicalist above, physico-chemical, the two situations seem parallel. And just as Brown's ignorance in the latter situation presents no problem for physicalism so, it may be suggested, neither does Smith's ignorance in the former.

I conclude that Smith can know of *e* that it has *R* without knowing of *e* that it has such-and-such a physico-chemical property, even if *R* is a physico-chemical property. But are there any positive grounds for saying that *Jones* knows of *e* that it has *R* in these circumstances? I believe that there are. Indeed, I believe that it can be shown that Jones knows of *e* that it has *R* even without the assumption that *R* is a physico-chemical property.

To see this point, consider the following parallel case. Suppose that someone we shall call 'Johnson' knows of a given beaker of liquid, *b*, that it contains the colourless, tasteless liquid that comes out of taps. Suppose further that 'water' is a rigid name with which Johnson is acquainted and that he knows a description which fixes its referent, namely 'the colourless, tasteless liquid that comes out of taps'. Then Johnson knows that water itself is the colourless, tasteless liquid that comes out of taps. This knowledge together with the knowledge above of *b* clearly suffices for Johnson to know of *b* that it contains water.

Jones is in the same epistemic situation with respect to *e* and *R* as Johnson is with respect to *b* and water. One of the things Jones knows of *e* is that it has the phenomenal content which is typically caused in Smith by his viewing red objects. Jones is also aware of Smith's rigid name, '*R*', and he knows a description which fixes its referent, namely 'the phenomenal content which is typically caused in Smith by his viewing red objects'. Hence, Jones knows that *R* is the phenomenal content which . . . etc. Hence, Jones must surely know of *e* that it has *R*.[19]

[19] And, relatedly, he must also know both, for example, that there are experiences having *R* and that normal perceivers undergo experiences having *R* in the presence of red objects.

It is convenient, at this time, to make some comments on the position Terence Horgan adopts in his 'Jackson on Physical Information and Qualia'. Horgan agrees with Jackson that a person who, for the first time, is introspectively aware of the characteristic experience of seeing red gains new knowledge about, for example, what it is (phenomenally) like to see ripe tomatoes. But this new knowledge, according to Horgan, is expressible in the sentence

(S) Seeing ripe tomatoes has this property,

where 'this property' is used to designate the colour-quale instantiated in the person's present

The situation in the Smith–Jones case, then, may be summarized from the physicalist perspective as follows: factual knowledge of *e* and its phenomenal content is either knowledge of *e* that it has *R*, where '*R*' is a rigid designator for the phenomenal content, or it is knowledge of *e* that it has the *F*, where 'the *F*' is a description for the content linking it to various causes or effects. Jones even before he gains his sight has both the former knowledge and the latter, whatever '*F*' signifies. Hence, there is no fact Smith knows about *e* and its phenomenal content of which Jones is ignorant.

This conclusion will, no doubt, still strike some philosophers as unsatisfactory, since it appears to be inconsistent with the obvious fact that after the operation Jones will *find out* something about *e* in particular and its phenomenal content. But there is no inconsistency. Jones will certainly learn something about *e* when he gains his sight. He will learn or discover what *e* was (phenomenally) like. This is new knowledge, knowledge of one particular experience, but it is *not* knowledge of any new facts. This needs a little explanation.

Intuitively, knowing what *e* is like is a matter of knowing what it is like to undergo experiences of *e*'s phenomenal type, that is, *R*, and knowing of *e* in particular that it has *R*. The latter knowledge links the former to *e* and thereby ties down general knowledge of what experiences with content *R* are like to the token event *e*. The analysis, I propose, then, of knowing what a given token event, *x*, is (phenomenally) like is as follows:

(2) *P* knows what *x* is like = df. There is a phenomenal content *C* such

experience. And the knowledge expressible in (S), so Horgan argues, presents no problem for physicalism.

This response to the Argument from Knowledge is unsatisfactory. The knowledge expressible in (S) is equally well expressible in

(S*) Seeing ripe tomatoes has *R*,

if the given person is presently experiencing *R*. Since our man Jones, prior to gaining his sight, *already* knows that seeing ripe tomatoes has *R*, it follows that knowledge of what it is (phenomenally) like to see ripe tomatoes is not the same as the knowledge expressible in (S).

Note that I am not claiming here that (S) and (S*) have the same meaning. Obviously there could be occasions when (S) expressed a truth and (S*) did not, as, for example, when a person with inverted colour vision saw ripe tomatoes for the first time and uttered (S). My point is that if 'this property' is used to designate *R*, then (S) and (S*) convey exactly the same factual information, information which is already in the blind Jones's possession.

Horgan insists that a hearer of (S) cannot obtain the information which the speaker expresses by (S), unless he, the hearer, experiences *R* himself and knows that 'this property' designates the same property that he experiences. I concede that the hearer must know that 'this property' designates *R* but I deny that he must experience *R* himself. A requirement of this latter sort is not needed in other indexical contexts (for example, I can know what you are saying when you utter the sentence 'This book is interesting', even if I do not perceive, and have never perceived, the book you are referring to, if, say, I already know, on other grounds, that it is Iris Murdoch's latest novel). Why, then, should such a requirement be imposed here? What, I suggest, Horgan should have said is the following: a speaker cannot express his knowledge that seeing ripe tomatoes has *R*, *by* uttering (S), unless he knows the fact that 'this property' designates *R in a certain way*, namely via introspective awareness of the presence of *R* in his experience and the application of 'this property' to *R*. Still the fact he knows can be known in *other* ways by persons who have not experienced *R*.

that (i) $P$ knows what it is like to undergo experiences with $C$, and (ii) $P$ knows of $x$ that it has $C$.

Given this analysis, the blind Jones does not know what Smith's colour experience $e$ is like, since for reasons given earlier he fails to satisfy condition (i). But he *does* satisfy condition (ii)—the condition which imposes *de re* knowledge of $e$—and hence he knows just as many facts about $e$ as Smith. After the operation, Jones comes to know what it is like to undergo experiences with content $R$ by undergoing the appropriate experiences and introspectively identifying them. He, then, satisfies both (i) and (ii) in (2)'s *analysans*. Hence, he discovers or comes to know what $e$ in particular was like. I conclude that the case of Smith and Jones presents no difficulty for the physicalist.

There is one further complication left to consider. Suppose that there is a congenitally blind person who, like Jones, has exhaustive knowledge of what is going on physically in his fellows when they see colours and use colour words, but who knows much more than Jones. In particular, this person—Super-Jones, as I shall call him—knows every physical fact there is to know about his own past, present, and future. Now if $R$ is a physical property, then Super-Jones, prior to the operation which gives him sight, already knows of each and every future colour experience of his which has $R$ that it has $R$. For he has complete knowledge of what will go on in his brain after the operation, and he will, therefore, already know of various future states of his brain that they have the appropriate physical property. Super-Jones, then, when he recovers from the operation cannot *come to know* of any of his own new colour experiences that they have $R$. So Super-Jones will not *learn* anything about these colour experiences. Yet this seems quite wrong. Surely when Super-Jones is first introspectively aware that he is having an experience with phenomenal content $R$ he will make a significant discovery.

Super-Jones will make a discovery, sure enough. When, for the first time, he becomes aware, via introspection, that a particular experience of his (call it '$f$') has $R$, he will come to know what $f$ is like. He did not know this at any earlier time because he did not know until then what it was like to undergo an experience having $R$. And he did not know what it was like to undergo an experience having $R$ because, before the operation, neither had he undergone an experience having $R$ nor had he identified any experience as having $R$ solely on the basis of introspection.[20] So, after the operation, Super-Jones does learn things about his colour experiences. For he acquires knowledge as to what they (both token and type) are like. But, according to the physicalist, he does not acquire knowledge of any new facts. Rather by introspection he acquires a new way of knowing certain facts he already knew by other means.

[20] I ignore here the complications introduced by the third disjunct in (1c).

In closing this section, I want now to show *very* briefly that physicalism is consistent with the empiricist thesis that some linguistic terms cannot be fully understood by persons who have not experienced their referents. Consider again '*R*'. Empiricists would maintain that Jones before the operation does not fully understand '*R*'. Yet he clearly has some understanding of '*R*'. He can formulate descriptions which '*R*' uniquely satisfies and he has no difficulty in identifying occasions on which *R* is present in Smith's experiences. So what does Jones lack? The obvious answer is that Jones, unlike Smith, does not know what it is like to have an experience with *R*. *This* is why he does not fully understand '*R*'. But according to the analysis I have given, Jones does not know what it is like to have an experience with *R* (at least in part) because he has not experienced *R* (or any phenomenally similar quality). The physicalist, then, can certainly accept the following empiricist thesis: full understanding of the meaning of a rigid name for a phenomenally simple quality presupposes experience of that (or some phenomenally similar) quality.

## III

Central to my argument so far has been the following claim: a person who knows what it is like to undergo the characteristic experience of seeing red does not thereby know any further fact than a person who has all the relevant physical information. It may be suggested that this claim leads to certain problems which my discussion does not yet fully answer. Suppose, for example, that a supernatural being has created an entire material world just like our material world, but that he still has to decide what it will be like for the humans in this world to see red. According to my defence of physicalism, prior to making any decision, this being, in knowing all the physical facts, knows *all* the facts. So, counterintuitively, he cannot really have anything left to decide after all. Alternatively, consider again the case of Super-Jones. How, it may be asked, can this remarkable individual *really* learn anything about colour experiences, on my account, when he gains his sight? Admittedly, as I say, Super-Jones will acquire new epistemic access to what he already knows. But, contrary to my claims, he will not thereby learn anything. For surely to learn something is to acquire *genuinely new* knowledge.

In my view, the first of these apparent difficulties, like others of its ilk, rests upon an illegitimate appeal to imaginability of the sort I criticized in Section I. That is to say, the described counterfactual situation is classified as possible presumably because intuitively it seems imaginable. But the fact that it seems to us that we can imagine a supernatural being who has created a physical world in every respect like ours and yet who remains in a position to decide what it will be like for humans in that world to see red, say, does not entail that we have succeeded in imagining any such thing. Perhaps the physical world we imagine is not exactly like ours (have we, for example,

really imagined all the relevant neural details?). Alternatively, perhaps we have imagined the right physical world and even a being who has created it, but one who *erroneously* believes that there is still something he can decide about what human experiences are like.

This reply does not strike me as counterintuitive (especially when it is seen against the background of my earlier discussion of imaginability). Furthermore there is, I suggest, separate reason to think it correct. For if, as is widely accepted, the physical facts determine all the facts, then, whether or not there are any non-physical facts, necessarily once the physical facts are fixed all the phenomenal facts are automatically fixed too. So, after the creation of a complete physical realm, the supernatural being is *not* in a position to choose which subjective experiences the inhabitants of that realm undergo.

Let us now take up the second problem. Does Super-Jones really learn anything, on my account, after he has the operation? The earlier analyses (1c) and (2) entail that, before the operation, Super-Jones does not know what certain experiences, token and type, are like, whereas, after the operation, he does. Given the following commonsensical definition:

(3) *P* discovers at time *t* what an experience *E* (token or type) is like = df (i) *P* does not know, prior to *t*, what *E* is like; (ii) *P* knows at *t* what *E* is like,

together with (1c) and (2), it should be clear that, on my approach, Super-Jones makes a discovery: he discovers what certain experiences are like. Hence, in one sense of the term 'learn', he does indeed learn something. Of course, as I have repeatedly emphasized, if physicalism is true, then he does not learn any new facts. But this presents no real difficulty. I maintain that discoveries can be made even when no new facts are discovered. This, it seems to me, is an independently plausible position. Consider, for example, the case of a man who, by experiment, *finds out* how to balance a pencil on the end of his nose. This man obviously discovers something, but what he discovers is not a new fact. I claim that, in this respect, Super-Jones and he are in the same boat.

On my view, then, a conceptual distinction can be drawn between discovering a new fact and discovering what a new experience is like. This is not to say that no discoveries of the latter type are ever discoveries of the former type. On the contrary, in typical cases a person who finds out what a certain experience is like comes to know a fact about that experience which he did not know before. My point is that (1c), (2), and (3) do not *require* for their truth that a new fact be discovered. Thus, there will be possible cases, such as that of Super-Jones, where the conditions laid down in (1c), (2), and (3) are satisfied, thereby ensuring that a discovery is made, even though, *according to the physicalist*, no new fact is uncovered.

I believe that the above critical examination renders the Argument from Knowledge entirely ineffective. Since I know of no other arguments which succeed in establishing that there are facts about our sensory experiences which cannot be captured in the physicalist's net, I see no reason to suppose that the subjective character of experience poses any special problem for physicalism.[21, 22]

*Department of Philosophy*
*Northern Illinois University*
*Dekalb, Illinois 60115*
*U.S.A.*

MICHAEL TYE

[21] It remains to be shown just where qualia are best located within the general physicalist scheme. I am presently inclined to favour the view that qualia are identical with physico-chemical properties. Admittedly this view places restrictions on the multiple realizability of qualia, but it does not entail that creatures whose physical constitution is different from ours *must* have different qualia (any more than the identification of temperature with mean molecular kinetic energy entails that humans and electro-mechanical robots cannot share the same temperature). Cf. Jaegwon Kim, 'Phenomenal Properties, Psycho-physical Laws, and the Identity Theory', *Monist*, 1972, pp. 177–92.

One other point should be added. I have repeatedly spoken as if qualia are properties of token experiences. I do not, in fact, accept this view except as a convenient way of speaking. I have argued elsewhere that ordinary talk which appears to refer to or quantify over token experiences can be analysed in a way which avoids that reference or quantification. See here my 'The Adverbial Approach to Visual Experience', *The Philosophical Review*, 1984, pp. 195–225; 'Pain and the Adverbial Theory', *American Philosophical Quarterly*, 1984, pp. 319–28; 'The Debate about Mental Imagery', *The Journal of Philosophy*, 1984, pp. 678–91; 'Against the Token Identity Theory' (with Terence Horgan), forthcoming in E. Le Pore and B. McLaughlin, eds., *The Philosophy of Donald Davidson: Perspectives on Actions and Events*. Within the adverbial account of experience which emerges from these papers, qualia are best seen as properties of persons.

[22] This paper was written during a visit to Oxford University. I am grateful to the fellows of Hertford College for their hospitality. I am also indebted to Simon Blackburn for helpful criticism.