THE JOURNAL OF PHILOSOPHY
VOLUME LXXIX, NO. 7, JULY 1982

THE INVERTED SPECTRUM*

As best I can determine, the idea of spectrum inversion made its first appearance in the philosophical literature when John Locke, in the *Essay*, entertained the possibility that "*the same Object should produce in several Men's Minds different* Ideas at the same time; v.g. the *Idea*, that a *Violet* produces in one Man's Mind by his Eyes, were the same that a *Marigold* produced in another Man's, and *vice versa*."[1] It was obviously part of Locke's supposition that the color experiences of the two people differ in such a way that the difference could not manifest itself in their behavior and their use of color words, and we will take this as an essential feature of full-fledged intersubjective spectrum inversion. This "inverted spectrum hypothesis" was revived in the early years of this century, and in the heyday of logical positivism it was a favorite target for applications of the verificationist theory of meaning; there are classic formulations and discussions of it in the writings of C. I. Lewis, Moritz Schlick, Hans Reichenbach, John Wisdom, Max Black, and J. J. C. Smart,[2] and it lurks beneath the surface, and sometimes at it, in many of Wittgenstein's discussions of "private experience."

* Versions of this paper were read at Trinity College, Dublin, at Princeton University, and at the Spring 1981 meetings of The Creighton Club (The New York State Philosophical Association). I have had helpful comments from many people, and owe a special debt of thanks to Jonathan Bennett. I gratefully acknowledge the support of The National Endowment for the Humanities.

[1] *An Essay concerning Human Understanding*, Peter H. Nidditch, ed., (New York: Oxford, 1975), p. 389 (Bk. II, Ch. XXXII, sec. 15).

[2] See Lewis, *Mind and the World Order* (New York: Scribner's, 1929), p. 75; Schlick, "Positivism and Realism," in A. J. Ayer, ed., *Logical Positivism* (New York: Free Press, 1959), pp. 92-95; Reichenbach, *Experience and Prediction* (Chicago: University Press, 1938), pp. 248-258; Wisdom, *Other Minds* (New York: Philosophical Library, 1952), pp. 10/1; Black, *Language and Philosophy* (Ithaca, N.Y.: Cornell, 1949), p. 3ff; Smart, *Philosophy and Scientific Realism* (New York: Humanities, 1963), pp. 66-69.

0022-362X/82/7907/0357$02.50

I

Wittgenstein seems to have been the first to give this idea a new twist by envisioning the possibility of *intra*subjective spectrum inversion:

> Consider this case: someone says "I can't understand it, I see everything red blue today and vice versa." We answer "it must look queer!" He says it does and, e.g., goes on to say how cold the glowing coal looks and how warm the clear (blue) sky. I think we should under these or similar circumstances be inclined to say that he saw red what we saw blue. And again we should say that we know that he means by the words 'blue' and 'red' what we do as he has always used them as we do.[3]

What is imagined here is that there should be a systematic difference between the character of someone's color experience at a certain time and the character of that *same* person's color experience at another time. If Wittgenstein was indeed the first to describe such a case, there is a mild irony in this. For there is a natural line of argument, which we will come to shortly, from the possibility of *intra*subjective inversion to the conclusion that it makes sense to suppose, and may for all we know be true, that *inter*subjective spectrum inversion actually exists—that among normally sighted people, i.e., those who are not color blind, there are radical differences in the way things look with respect to color. And Wittgenstein is associated, probably more than any other philosopher, with the view that this supposition makes no sense. In the midst of the attack on the notions of "private language" and "private objects" in the *Philosophical Investigations*[4] there occurs the following passage:

> The essential thing about private experience is really not that each person possesses his own exemplar, but that nobody knows whether other people also have *this* or something else. The assumption would thus be possible—though unverifiable—that one section of mankind has one sensation of red and another section another (95).

I think it is pretty clear from the tenor of the surrounding passages that Wittgenstein thinks that this "assumption" is in fact senseless or conceptually incoherent and takes it to be a *reductio ad absurdum* of the notion of "private experience" he is attacking that it implies that this "assumption" might be true.

[3] "Notes for Lectures on 'Private Experience' and 'Sense Data'," Rush Rhees, ed., *Philosophical Review*, 3 (July LXXVII, 1968): 284.

[4] G. E. M. Anscombe, trans., 3d ed. (New York: Macmillan, 1958).

I said that there is a natural line of argument from what Wittgenstein seems to admit—the logical possibility of intrasubjective spectrum inversion—to what he apparently denies the meaningfulness of asserting—namely, the possibility that intersubjective spectrum inversion actually exists. One reason why the claim that intrasubjective inversion is logically possible makes a natural starting point for such an argument is that it seems immune from verificationist objections; as Wittgenstein's example shows, it is easy to imagine phenomena we would take as verifying that such a change in color experience had occurred. One can imagine this happening in oneself, one can imagine another person reporting that it had happened to him, and one can imagine nonverbal behavior that would be evidence of such a change. But—and here comes the promised argument—it seems, offhand, that if intrasubjective spectrum inversion is possible, intersubjective inversion must also be possible. For suppose that someone, call him Fred, undergoes intrasubjective inversion at time t.[5] Assuming that others did not also undergo inversion at t, it would seem that either before t or afterward (or both) Fred's color experience must have been radically different from that of others. But if we allow that there can be intersubjective inversion in cases in which there is intrasubjective inversion, it seems that we must allow that there could be intersubjective inversion without intrasubjective inversion; if the color experience of a person can differ from that of others at some point during his career, it should be possible for such a difference to exist throughout a person's career. But if this is a possibility, then it does seem perfectly coherent to suppose, and perfectly compatible with all the behavioral evidence we have about the experiences of others, that, in Wittgenstein's words, "one section of mankind has one sensation of red and another section another."[6]

Suppose one allows the premise of this argument: that intrasubjective inversion is possible; how, if at all, can one resist its conclusion? How is one to reject the inference from the possibility of intrasubjective spectrum inversion to the possibility of intersubjective spectrum inversion? One way is to maintain that the relevant notion of similarity and difference, what I shall call *qualitative* similarity and difference, is well defined only for the intrasubjective case. This allows one to deny the possibility of intersubjective inversion, but does so at the cost of forbidding one from saying

[5] For Fred's first appearance in the discussion of this problem see D. M. Taylor, "The Incommunicability of Content," *Mind*, LXXV, 300 (October 1966): 527-541.

[6] See Taylor, *op cit*, where this point is made.

that one's color experiences are qualitatively similar to those of others. On this view (see section IV), experiences belonging to different persons can be neither qualitatively similar nor qualitatively different.

It does not appear, however, that Wittgenstein could have taken this way out. In the passage in which he seems to allow the possibility of intrasubjective inversion he seems to allow that the case he describes would also be a case of intersubjective inversion; he says that "we should under these or similar circumstances be inclined to say that we saw red what he saw blue." Why then is Wittgenstein not committed to the very thing he seems to deny? He was not unaware of the problem, for in a later passage he wrote:

> We said that there were cases in which we should say that the person sees green what I see red. Now the question suggests itself: if this can be so at all, why should it not always be the case? It seems, if once we have admitted that it can happen under peculiar circumstances, that it may always happen . . . This is a very serious situation ("Notes for Lectures," p. 316).

A full discussion of how Wittgenstein thought he could solve or avoid this problem would take me too far afield—and the answer is none too clear. Briefly, however, I think that it is only in a qualified sense that he allowed that intrasubjective inversion is possible, and only in a similarly qualified sense that he allowed that in a case of intrasubjective inversion we would also have intersubjective inversion. I think he thought that such a case would have to be one in which the person who has undergone inversion describes things as looking "queer," and says, for example, that the clear sky looks warm and that the glowing coal looks cold. Now, if the difference between someone's color experience and ours went with a tendency on his part to describe his experience in such ways, then of course it would not be a difference that could not manifest itself in behavior (since verbal behavior is behavior), and so would fail to amount to a case of "full blown" intersubjective spectrum inversion.

But it would seem offhand that Wittgenstein was mistaken if he thought that any case in which we could know that someone's experiences of colors at one time were radically different from his experiences of the same colors at another time would have to be one in which at one or the other of the times the person describes his experiences as "queer." Suppose that it was thirty years ago that Fred underwent his spectrum inversion. We have monitored him closely since then, and at no time has he reported a "reinversion," or given any behavioral indication of one. He has, however, gradu-

ally become accustomed to the new look of things and to describing the colors of things in the same words others use; for some time now it has come natural to him to say that the sky is blue and daffodils yellow, and, moreover, that the sky *looks* blue and daffodils *look* yellow. Each day during the last thirty years we have asked him how things looked compared with how he remembered their looking the day before, and each day (after the first) he has confidently reported that things look the same with respect to color as they did the day before—although in the early years of this period he also reported that their looking this way was seeming less strange with the passage of time, that he was again finding it natural to describe glowing coals as looking warm rather than cold, and so on. In such a case I think we could have good reason to think that Fred's color experience now is systematically different from what it was before the inversion, even though he does not now, and did not then, describe his experience as "queer."[7] But if this case is possible, then it seems that full-fledged intersubjective spectrum inversion should also be possible.

II

I have so far talked as if we would have a case of spectrum inversion if blue things looked to you the way yellow things look to me, and vice versa. But a little reflection shows that this is not enough. If you and I differed *only* in this way and if all other colors looked the same to us, it is obvious that the difference would manifest it self in behavior, both verbal and nonverbal. One of us would find yellow things more similar to orange things, and less similar to violet things, than blue things are, while the other would find just the opposite. What is required for full-fledged intersubjective inversion is that the color "quality spaces" of the two people should have the same structure, which requires (among other things) that under the same lighting conditions they make the same judgments

[7] Though I think we could have good reason to believe this, I do not claim that the behavioral evidence I have described entails it. As Gilbert Harman and David Lewis have pointed out to me, what I have described of Fred's behavior is compatible with there having been over the thirty years a change in Fred's color experience so gradual that the change from one day to the next would not be noticeable (as Lewis put it, a rotation of the color circle so slow that it took it thirty years to go 180 degrees), the net effect of which was to undo the initial inversion. But we could have reason to discount this possibility if we had evidence that over the thirty years there had not occurred in Fred any physiological change such that, if it occurred suddenly rather than gradually, it would produce noticeable behavioral manifestations of intrasubjective inversion. Further possible behavioral evidence of inversion is described in section II, and the relevance of physiological considerations to questions about inversion is discussed in section V.

of relative color similarity ("*A* is more similar to *B* than to *C*") about the same visually presented objects. If this condition is satisfied, then any difference in how the two people see colors will ramify through all the colors; all, or virtually all, will have to look different to the one person than they do to the other. Pretending, for the moment, that we have only the pure, "saturated," colors to deal with, we might have such a systematic difference if to each of our two persons each color looked the way its complementary color looked to the other.

The same applies, *mutatis mutandis*, to the case of intrasubjective inversion. And this puts us in a position to answer an objection that is sometimes raised against putative cases of intrasubjective spectrum inversion. Suppose Fred claims to have just undergone spectrum inversion. How do we know, it is asked, that what has changed in Fred is his color experience, and not his understanding of color words or his memory of how things looked to him in the past—that he is not the victim of some peculiar sort of aphasia or memory illusion? The suggestion behind such questions is often that the questions are unanswerable and that this undermines the claim that we could know that intrasubjective inversion had occurred. To parry such objections, let me complicate our case slightly. Let us suppose that the relationships between the different colors can be represented by associating each determinate shade of color with a point on the circumference of a circle, the distances between the points along the circumference corresponding to the perceived differences between the shades, and the point corresponding to any shade being opposite the point corresponding to its complementary. Let us label the points on the circle as they are on the face of a clock, with the numerals 1 through 12. I will suppose, indeed, that we have a circle, call it a "color circle," on which the points on the circumference actually have the colors they represent. And now let us imagine our case as follows. At time t_1 Fred was perfectly normal in his use of color words, his discriminatory abilities, and the like. But at time t_2 he tells us that a remarkable change has occurred. Although most things look to him the way they used to, a sizable minority look different. He describes the change by saying that, if he looks at a color circle, it looks the way it would have looked at t_1 if the shades between 12 and 2 had been interchanged with their complementaries (those between 6 and 8), the rest of the circle remaining unchanged. According to this, the structure of Fred's visual color space at t_2 is different from its structure at t_1. And, because of this, we can suppose that Fred's testi-

mony is supported by his nonverbal discriminatory and recognitional behavior—the ease with which he discriminates certain shades that formerly were difficult for him to discriminate (and are so for the rest of us) and the difficulty with which he discriminates other shades that seem to us (and previously seemed to him) very different. At t_3 Fred tells us that another such change has occurred, adding itself to the first one; this time it is the shades between 2 and 4 that have changed places with their complementaries (those between 8 and 10). Again we can suppose there is behavioral evidence to substantiate his claim. Finally, at t_4 he tells us that still another such change has occurred; this time it is the shades between 4 and 6 which have changed places with their complementaries (those between 10 and 12). Again there is behavioral evidence to substantiate his claim. But at t_4, unlike t_2 and t_3, Fred's judgments of color similarity and difference will coincide with ours and those he made at t_1; at t_4 the structure of Fred's color space is the same as it was at t_1. Yet Fred reports that his color experience is systematically different from what it was at t_1; each color looks the way its complementary looked then. And this claim seems to be supported by the behavioral evidence that supported his claims that there were changes in his color experience between t_1 and t_2, between t_2 and t_3, and between t_3 and t_4; for these partial inversions add up to a total spectrum inversion. It does not appear that any sort of aphasia or memory failure could account for the phenomena imagined here.[8]

[8] I used a similar example in a footnote of my paper "Functionalism and Qualia," *Philosophical Studies*, XXVII, 5 (May 1975): 219–315, p. 313/4.

It might be objected that it is compatible with the behavior I have imagined on the part of Fred that between t_1 and t_2 he underwent a series of experience *cum* memory changes that resulted in his color experience at t_4 being the same as his color experience at t_1, despite his seeming to remember it as being different. I do not deny this, but do not think it follows that in the case as described we would not be justified in thinking that Fred at t_4 was spectrum-inverted relative to Fred at t_1. In general, a memory change having a certain behavioral effect seems far less likely than an experience change having the same effect (where either could produce the effect); on any plausible assumption about how perception and memory are realized physically, a change that alters the way specific kinds of stimuli are linked with specific color qualia (as happens when someone puts on tinted spectacles) seems far more likely than one that systematically modifies all a person's memories of how things looked with respect to color prior to a certain time, leaving these memories otherwise unchanged. And, quite apart from this, in the absence of overriding physiological evidence (and I do not deny that there could be such), the hypothesis that someone has undergone an experience change that would produce a certain behavioral effect is obviously to be preferred, on grounds of simplicity, to the hypothesis that the person has undergone an experience change *and* a memory change which jointly would have the same effect.

Returning to the case of intersubjective spectrum inversion, I suspect that the main reason why many philosophers are hostile to the claim that such spectrum inversion is possible is that they suspect that admitting that claim will put one on a slippery slope which will eventually land one in skepticism about other minds. If I cannot know from the behavior of others that their color experiences are like my own, neither can I know that when they are cut or burned they have experiences phenomenally or qualitatively like my pains; and if I cannot know even that much, it is natural to suppose, then I cannot know anything about their minds. A common counterargument, which argues from the falsity of skepticism to the impossibility of spectrum inversion, goes as follows. According to our ordinary standards of evidence, if two individuals Jones and Smith make the same color discriminations, agree in their judgments of color similarity and difference, and apply color words in the same way, this is sufficient evidence that they mean the same by their color words. And if in addition they have learned to use 'looks', 'appears', etc. in the same ways in the same objective circumstances—for example, both say of what they know to be a white wall illuminated by red light that it "looks" red—this is sufficient evidence that they mean the same by expressions like 'looks red'. But if Jones says, truthfully, that an object looks red to him, and Smith says the same of the object, and if they mean the same by the expression 'looks red', this surely shows that with respect to color the object looks pretty much the same to them and that their experiences of it are similar. Yet, according to the view that intersubjective spectrum inversion is possible, this information about Jones's and Smith's use of color words, discriminatory abilities, and so forth would leave it an entirely open question whether red things look alike to them. So we must choose between rejecting our ordinary standards of evidence concerning such matters and rejecting the view that intersubjective spectrum inversion is possible. But we cannot abandon our ordinary standards of evidence here without accepting an absurd general skepticism about other minds. So, the argument concludes, we must reject the claim that spectrum inversion is possible.

I will not deny that if it is possible to have some knowledge of other minds, it must be possible to know whether others mean the same as we do by their color words. Nor will I deny that if we can know what someone means by expressions like 'red' and 'looks red', we can also know that something looks to him the way it looks to us. What I dispute in the argument just given is the claim

that this conclusion, *in the sense in which it is true,* contradicts the claim that spectrum inversion is possible. What emerges here is that expressions like 'looks the same' are potentially ambiguous. And this can be seen from further reflection on Fred and his intrasubjective spectrum inversion.

We have supposed that Fred eventually accommodates to the change in his color experience, and says that an object looks yellow in just those objective circumstances in which others would say this—even though the way something looks to him when he says it looks yellow is the way things formerly looked to him when he said they looked blue. But this accommodation will cause him some difficulty in the use of such expressions as 'looks the same' and 'looks similar in color to'. Others will assume that from the premises "X looked yellow to Fred at time t_1" and "Y looked yellow to Fred at time t_2" we can infer the conclusion "The way X looked to Fred at t_1, with respect to color, is the way Y looked to Fred at t_2, with respect to color." But if t_1 is a time before Fred's spectrum inversion and t_2 is a time after he has accommodated to it, then that conclusion will be false on one interpretation of it, although true on another, if the premises are true. Fred needs, and we need, a distinction between different senses of the expression 'looks the same', and of related expressions. If Fred's house looked yellow to him at both t_1 and t_2, then with respect to color his house "looked the same" to him at those two times in the sense that his experiences of it on those two occasions were *of* the same objective color, or had the same color as their "intentional object." Call this the *intentional* sense of 'look the same'. But in another sense his house did not "look the same" to him at the two times; call this the *qualitative* sense of that expression. That Fred has undergone spectrum inversion requires that things look different to him than they did before in the qualitative sense, but not (once he has accommodated to the change) that they look different to him in the intentional sense.

Applying this distinction to the case of Jones and Smith in the anti-inversion argument, we can say that it is only in the intentional sense of 'looks the same' that the information about Jones and Smith—their use of color words, their abilities to distinguish and recognize colors, and so forth—establishes that red things look the same to them. What this information does not establish, by itself, is that red things look the same to them in the qualitative sense—that their visual experiences of redness are qualitatively as well as intentionally similar. So long as our ordinary standards of evidence

are taken as standards for establishing intentional similarity of experiences, the adherence to these standards is perfectly compatible with allowing the possibility of spectrum inversion. And, I think, it is only if the standards are taken in this way that it is plausible to maintain that abandoning them would lead to general skepticism about other minds.

It may appear that I have conceded part of the claim that allowing the possibility of spectrum inversion leads to skepticism, namely, that it leads to skepticism about our ability to have knowledge of the qualitative character of the experiences of other persons. But I have not conceded that; on the contrary, whether this is so will be one of the main issues under consideration in the rest of this paper. What I have conceded is that it follows from the possibility of spectrum inversion, and is indeed true, that the *behavioral* evidence that establishes intentional similarities and differences between experiences of different persons is not by itself sufficient to establish qualitative similarities and differences between such experiences. And it was such behavioral evidence I had in mind in speaking of our "ordinary standards of evidence." It remains to be considered whether other sorts of evidence, e.g., evidence of physiological similarities or differences of some kind, might enable us to make intersubjective comparisons of the qualitative character of experiences.

III

The last objection to the possibility of spectrum inversion which I shall discuss (and there are many I shall have to ignore) is empirical. A brief consideration of it will help me bring into focus some of the problems raised by the possibility of spectrum inversion.

If spectrum inversion is to be possible, there must be a mapping which maps every determinate shade onto some determinate shade and at least some onto shades other than themselves, which preserves, for any normally sighted person, all the "distance" and "betweenness" relationships between the shades, and which maps primary colors onto primary colors. Now as long as we restrict ourselves to the pure saturated colors, various such mappings seem to be possible; one is the mapping of shades onto their complementaries, and others can be got by rotating the "color circle" in different ways. It has been questioned whether even these mappings satisfy the condition that primaries be mapped onto primaries.[9] But it

[9] See Bernard Harrison, "On Describing Colors," *Inquiry*, x, 1 (Spring 1967): 38-52.

seems even more questionable that we can get a mapping that satisfies these conditions once all the unsaturated colors (beige, olive, rust, etc.) and nonchromatic colors (black, white, and the various shades of grey) are taken into account. Obviously the simple color circle is inadequate for the representation of the relationships of these to one another and to the "pure" colors.

The question of whether our color experience does have a structure that allows for such a mapping—whether it is "invertible"—is an empirical question about our psychological makeup. And it is one I intend to by-pass. Even if our color experience is not invertible, it seems obviously possible that there should be creatures, otherwise very much like ourselves, whose color experience does have a structure that allows for such a mapping—creatures whose color experience *is* invertible. And the mere possibility of such creatures is sufficient to raise the philosophical problems the possibility of spectrum inversion has been seen as posing.

Let me be more explicit about what these problems are. One problem, call it the *metaphysical* problem, is about the nature of "qualia"—the qualitative or phenomenal features of sense experience, in virtue of having which they resemble and differ from each other, qualitatively, in the ways they do. If spectrum inversion is so much as a logical possibility—whether or not it is a possibility *for us*, as we are currently constituted—then it is clear that no behavioristic account of qualia will do. For what the possibility of interpersonal spectrum inversion comes to is that two people might be behaviorally indistinguishable, might share all the same behavioral dispositions, even though their color experiences were radically different in qualitative character—and this means that there might be a psychological difference between people who would have to be psychologically identical if behaviorism were true. This is perhaps not such a great problem, since there are plenty of independent reasons for regarding behaviorism as false. But it also appears that the possibility of intersubjective spectrum inversion is incompatible with what many regard as the most respectable descendant of behaviorism, namely functionalism, where this is understood as the view that mental states are definable in terms of their causal relations to sensory inputs, behavioral outputs, and other mental states. Now if spectrum inversion is a logical possibility, then the quale currently involved in my perception of blue things cannot be defined by the functional role it plays in the likes of me, since in someone whose spectrum was inverted related to mine a different quale (perhaps the quale involved in my perception of yellow

things) would play that functional role, and this quale would play a different role. This is a version of what has been called the "inverted-qualia objection" to functionalism—spectrum inversion being a special case of "qualia inversion."[10]

The other main problem posed by the possibility of spectrum inversion is what I shall call the *epistemological* problem—the problem of how we can know about the qualitative states of other persons. Now if it turns out that our color experience is not invertible, there is one epistemological problem we do not face: we do not have to worry about the possibility that the color experience of others is inverted relative to our own. But once it is clear that qualia are not behavioristically or functionally definable, other possibilities have to be contemplated. For example, it seems compatible with two creatures' having color quality spaces with the same structure that none of the color experiences of either creature should bear any qualitative similarity to any of the color experiences of the other—e.g., the sensations of red of the one are not only not like the sensations of red of the other, but also are not like the other's sensations of green or his sensations of any other color. So we must consider what grounds, if any, we have for thinking that we do not differ from our friends and neighbors in this radical way. And the fact, if it is a fact, that our color experience is not invertible, i.e., that its structure does not yield mappings of the sort described above, is no reason whatever for thinking this.

I conclude that there is no fundamental epistemological or metaphysical problem here which is solvable on the assumption that our experience is not invertible but not solvable on the assumption that it is. So it will do no harm—and will make exposition easier—to assume from now on that it is.

IV

I now want to take up the view, mentioned earlier, that the relationships of qualitative similarity and difference are well defined only for the intrasubjective case. This view solves the epistemological problem about the qualia of others, not by offering an account of how we know which color experiences of others are qualitatively similar to our own, but by denying that there is anything of this sort for us to know.

A view something like this is suggested by remarks of Frege in his essay "The Thought." Frege discusses the case in which he and a color-blind companion are looking at a strawberry field, and he

[10] See Ned Block and Jerry Fodor, "What Mental States Are Not," *Philosophical Review*, LXXXI, 2 (April 1972): 159-181, pp. 172-174.

declares to be "unanswerable, indeed really nonsensical," the question "does my companion see the green leaf as red, or does he see the red berry as green, or does he see both as one colour with which I am not acquainted at all?"[11] He goes on to say that "when the word 'red' does not state a property of things but is supposed to characterize sense-impressions belonging to my consciousness, it is only applicable within the sphere of my consciousness." Putting this in my terminology, Frege is saying that insofar as he uses 'red' as applying to experiences qualitatively similar to certain experiences of his, it will be applicable only to experiences of his. This is not a solipsist claim that assigns a special status to his sense experiences; it merely asserts that the relationship of qualitative similarity can hold only intrasubjectively. A similar view was held by Moritz Schlick in "Positivism and Realism," on explicitly verificationist grounds: "The proposition that two experiences of different subjects not only occupy the same place in the order of a system but are, in addition, qualitatively similar has no meaning for us. Note well, it is not false, but meaningless: we have no idea what it means" (93).[12]

This view, call it the *Frege-Schlick view,* does not fly as violently in the face of common sense as it may initially seem to do. It does not imply that we are talking nonsense, or that our remarks are without truth value, when in ordinary circumstances we speak of

[11] Gottlob Frege, "The Thought: A Logical Inquiry," *Mind,* LXV, 259 (July 1956): 289-311, p. 299.

[12] A more recent expression of this view (minus the verificationism) is that of Thomas Nagel: "A type of relation can hold between elements in the experience of a single person that cannot hold between elements in the experience of distinct persons: looking similar in color, for example. Insofar as our concept of similarity of experience in the case of a single person is dependent on his experience of similarity, the concept is not applicable between persons" ["Brain Bisection and the Unity of Consciousness," in *Mortal Questions* (New York: Cambridge, 1979), fn 10, p. 160/1]. Nagel informs me that he is now doubtful about this view. I suggested such a view myself in my "Critical Notice: *Myself and Others* by Don Locke," *Philosophical Quarterly,* XIX, 76 (July 1969): 272-279, pp., 276-278.

As I am interpreting this view, it regards qualitative similarity as well defined for diachronic intrasubjective comparisons as well as for synchronic ones. But another version of the view is possible. Reichenbach held that "if we call the impressions of two persons incomparable, we are obliged to call the impressions of one person at different times incomparable as well" (*op. cit.,* p. 252), and this could lead (although I do not find that it did in Reichenbach—he uses his claim to attack the very notion of a quale) to the view that qualitative similarity and difference are well defined only for synchronic intrasubjective comparisons. On such a view intrasubjective inversion is no more possible than intersubjective inversion. I shall not consider this view, partly because it seems to me to have little intrinsic plausibility (especially the sharp epistemological distinction it has to make between synchronic and diachronic intrasubjective comparisons), and partly because I am investigating what options are open to someone who allows the possibility of intrasubjective inversion.

different persons having similar or dissimilar experiences, or when we say that something looks the same, or different, to two different people. For it is open to us to construe such remarks as about the *intentional* similarity or difference of experiences, rather than as about the *qualitative* similarities and differences; and the view in question does not of course deny that the relationships of intentional similarity and difference are well defined for the intersubjective case.

One attraction of the Frege-Schlick view is that it solves, or dissolves, the epistemological problem about our knowledge of the qualitative states of others. But this view also suggests an answer to the "inverted-qualia objection" to behaviorism and functionalism, and thus a solution to the metaphysical problem. Part of the motivation for holding that qualitative similarity is well-defined only for the intrasubjective case is that it is only in the intrasubjective case that we can have direct behavioral evidence of the holding of this relationship. It is this, after all, which makes *intra*subjective spectrum inversion, but not *inter*subjective inversion, behaviorally detectable. Now if intrasubjective qualitative similarity is the only sort of qualitative similarity there is (as the Frege-Schlick view holds), and if it is behaviorally detectable, then for all that has been shown it is behaviorally *definable* as well. But a much more promising view that is suggested by the same considerations is that this relationship is *functionally* definable, i.e., definable in terms of how the holding of the relationship between experiences is causally related to sensory inputs, behavioral outputs, and other mental states. It would be central to the functional account that when this relationship holds between different visual experiences of a person it tends to produce in that person the belief that there are objective similarities in the things he is seeing—to put it roughly, similarity of color qualia tends to produce belief in similarity of seen colors. And, via their effects on the person's beliefs, the qualitative similarities between his experiences would affect his behavior, in particular his recognitional and discriminatory behavior. But in addition, the holding of this relationship between different experiences of a person will produce in that person the belief that the experiences themselves are similar and, via this belief, will affect his verbal behavior—thus it is that Fred's saying that marigolds look today the way violets looked yesterday is taken as evidence that he has undergone inversion. Similar remarks apply to the relationships of qualitative identity (a special case of qualitative similarity) and qualitative difference. It is obvious that it is only when these relationships hold intrasubjectively that the holding of them can play such causal roles. But if, as the Frege-Schlick view holds, the relation-

ships *can* hold only intrasubjectively, then there seems to be no reason why they should not be functionally definable in terms of such causal roles. Given this, there seems no reason why we should not be able to define in functional terms what it is for a state to have a qualitative character—a state will have a qualitative character if it is qualitatively similar to or different from some other state or states, and we are supposing that we already have a functional account of qualitative similarity and difference. And this clears the way for making states like pain functionally definable; at any rate, the fact that pain necessarily has a qualitative character, one that is unpleasant and distracting, is no bar to its being functionally defined. If individual qualia are not functionally definable, then of course no individual quale can be mentioned in the functional definition of pain; but if the similarity and identity conditions of qualia are functionally definable, it will still be possible to quantify over qualia in such a definition, and that is all that seems to be required. If such an account can be made to work, functionalism has nothing to fear from qualia.[13]

It is natural to suppose that a functional account of qualia of the sort just sketched requires the Frege-Schlick view. And it may seem unpromising for just this reason. For the Frege-Schlick view clashes with strongly felt intuitions. Most of us, I suspect, cannot help feeling that a visual experience of mine can be like a visual experience of yours in exactly the way it can be like another visual experience of mine, even though this intersubjective similarity differs from intrasubjective similiarity in not being directly experienceable or rememberable by anyone. Call this the *common-sense view*.[14] I shall now try to show that, although the common-sense view is of course incompatible with the Frege-Schlick view, never-

[13] See My "Functionalism and Qualia," *op. cit.*

[14] It is worth observing that if, as has often been supposed in recent discussions of personal identity [e.g., Derek Parfit's "Personal Identity," *Philosophical Review*, LXXX, 1 (January 1971): 3–27], "fusion" of persons is a logical possibility, then the Frege-Schlick view must be false. For if persons *A* and *B* fuse to form *C*, and *C*'s subsequent mental life is psychologically continuous with the past lives of *A* and *B* (and so, among other things, contains memories of them), then, since the pre-fusion experiences of *A* and *B* must be qualitatively comparable with the post-fusion experiences of *C* (must stand to them in determinate relationships of qualitative similarity and difference), they must be qualitatively comparable to each other. But suppose that *A* and *B could* fuse at time *t*, but do not in fact do so. It would seem that, since *A*'s and *B*'s experiences prior to *t* will be qualitatively comparable in the case in which fusion subsequently occurs, they must also be qualitatively comparable in the case in which it doesn't occur; it can scarcely be the case that whether experiences occurring before *t* are qualitatively comparable depends on what happened at *t* or afterwards. It thus appears that, if creatures are "fusible," their experiences are qualitatively comparable. But this supports the common-sense view only if human beings are fusible—and that seems rather questionable.

theless, the functional account of qualia, which is suggested by the Frege-Schlick view and may seem to imply it, is in fact not only perfectly compatible with common-sense view but can be used to defend it.

V

Let us begin by seeing why the functional account of qualitative similarity does not imply the Frege-Schlick view. The word 'qualia', it will be remembered, is intended to refer to those features of sensory states in virtue of which they stand to one another in relationships of qualitative similarity and difference. States will be qualitatively similar in virtue of having identical or similar qualia—so our functional account of qualitative similarity must say what it is for qualia to be qualitatively similar. The first point to be emphasized is that qualia are properties and, therefore, universals. Suppose, then, that $Q1$ and $Q2$ are two different qualia. What the functional analysis of qualitative similarity tells us is that $Q1$ and $Q2$ are similar to a certain degree if it is the case that when $Q1$ and $Q2$ characterize two different experiences belonging to one and the same person, this tends to have certain effects on that person's beliefs and behavior. There is nothing in this to say that $Q1$ and $Q2$ can characterize only experiences belonging to one and the same person. And, if they characterize experiences belonging to different persons, those experiences will be similar to the appropriate degree, even though this pair of instantiations of $Q1$ and $Q2$ will not have, or tend to have, the effects definitive of qualitative similarity; the similarity will hold in virtue of the fact that $Q1$ and $Q2$ are such that, if they *were* to be instantiated in experiences of the same person, this *would* tend to have these effects.

But these remarks presuppose that one and the same quale can be instantiated in the experiences of different persons, and it may be objected that no sense can be made of this, given the sort of functional account of qualitative similarity and identity I have suggested. The answer to this is, in brief, that qualia can be shared by experiences of different persons in virtue of their being "realized" in other properties, presumably physical properties, that can be shared by experiences of different persons. Here we must remember that although there is a sense in which qualia are not functionally definable (if qualia inversion is a possibility), there is also a sense in which they are—their similarity and identity conditions, I have claimed, are functionally definable. Now functional states and properties can be said to have physical "realizations." A physical state or property *realizes* a functional state or property in a particular creature if in the workings of that creature it plays the "causal role" definitive of that functional state or property, i.e., if

it interacts causally in the required ways with inputs, outputs, and other internal states of the creature. Given the sense in which they are functionally definable, qualia too can be said to have physical realizations, which in principle we could discover by physiological investigations. The physical properties that realize qualia will be properties that can be instantiated in different people. This makes it possible for the same qualia to be instantiated in different people and, thus, for experiences of different people to be qualitatively similar and different in all the varying degrees. Moreover, it makes it possible that such similarities and differences should be discovered.

I am not saying that it is automatically true that, for any two creatures who have experiences having qualitative character, the color experiences of each will stand to those of the other in determinate relationships of qualitative similarity and difference—or, to abbreviate this, that the color experiences of the two creatures will be "qualitatively comparable." The color experiences of two different creatures will be qualitatively comparable only if those creatures are capable of having states having the same qualia, and on my account this in turn will be true only if, for at least one color quale, both creatures are capable of having states that share at least some of the physical properties that are realizations of that quale. Suppose, to invoke a favorite functionalist fantasy, that we come across a race of Martians who are behaviorally indistinguishable from us and have a "psychology" isomorphic with ours, but whose internal physical makeup—their neurophysiology and biochemistry—is utterly different from ours. These Martians are to be creatures who share our mental states, at least on a functionalist view, but in whom the physical realizations of these states are as different as they could possibly be from their realizations in us. On my functional account of qualia, these Martians would have states having qualia—there would be something it would be like for them to have these states. But their experiences would not share any of the qualia our experiences have; for I am assuming that none of the properties that realize qualia in us could be instantiated in them. When it comes to comparing Martian experiences and ours, something like the Frege-Schlick view holds: their experiences and ours are not qualitatively comparable. But I do not say, with Schlick, that it is meaningless to assert that our experiences are qualitatively similar to those of the Martians; on my view, that our experiences are not qualitatively comparable with theirs would be something to be discovered empirically, by discovering the physiological differences between them and us.

Now let me fill this account out a bit. Let *Q-BY* be the quale

currently involved in my perception of the color blue, and let *Q-YB* be the quale currently involved in my perception of the color yellow. And let us make the simplifying assumption that only one sort of spectrum inversion is possible for humans and that this involves *Q-BY* playing the causal role in the visual perception of one person, or at one time, which *Q-YB* plays in another person, or at another time, and vice versa. As I have already said, it is not possible (if spectrum inversion is possible) to give a purely functional characterization of either *Q-BY* or *Q-YB*. But if the notion of a quale can be functionally defined (i.e., if its similarity and identity conditions can be), then the following will be a functional description: 'pair of qualia such that, at any given time, one member of that pair characterizes perceptions of blue while the other characterizes perceptions of yellow'. And, given our simplifying assumption, this description will pick out a unique pair of qualia for each creature having a color quality space with the same structure as our own, although the unique pair may be different in different creatures (e.g., it will be different in a Martian than in one of us). And we can ask how the satisfaction of this description is physically realized in the case of a particular creature at a particular time. In principle we can go hunting in the physiology of a creature for a pair of physical properties which play the functional role of qualia, which are involved only in the perception of blue and yellow, and which are such that, as long as one of them characterizes perceptions of blue the other can characterize only perceptions of yellow, and vice versa (given that the structure of the total color experience is normal). If the creature is me, one member of any such pair will be a realization of *Q-BY* and the other will be a realization of *Q-YB*. There may in fact be a number of different realizations of each of these qualia. To be realizations of the same quale, different properties must be qualitatively identical in the following sense: the experience someone has in virtue of being in a state having the one has the same color quale as the experience he has in virtue of being in a state having the other. These qualia can, in fact, be identified with the disjunctions of the properties in their respective realization classes. What realizes in me the above functional description of a "blue-yellow invertible pair" is the pair of disjunctive properties consisting of the disjunction of the realizations of *Q-BY* and the disjunction of the realizations of *Q-YB*.

Having discovered that in a certain creature a certain pair of disjunctive properties realize this description, we could of course discover which of the properties is, at a given time, involved in a creature's perception of blue and which is involved in the creature's

perception of yellow. If in another creature we find that the same members of this pair of properties are involved in the perception of the same colors, we know that blue and yellow look qualitatively the same to these two creatures. If we find the reverse, we know that we have a case of spectrum inversion. But for blue to look to me as it does to you, it is not essential that we be in physically similar states. *Q-BY*, I have said, can be identified with a disjunction of different properties, and it may be that my experience has *Q-BY* in virtue of having one of these properties while yours has *Q-BY* in virtue of having a different one. But unless you and I are enough alike physically that there is some physical realization of *Q-BY* that *can* be instantiated in both of us, it is impossible for us both to have experiences having *Q-BY*. What I am supposing about my Martians is that the sets of possible physical realizations in them of their qualia do not overlap at all with the sets of possible realizations in us of our qualia—and from this it follows that none of our qualitative states are qualitatively similar, to any degree at all, to any of theirs.

But now let us confront a problem. Let us suppose that in me having visual experiences is realized in *X*-fiber firings, and that my visual experience's having a certain color quale, say *Q-BY*, is realized in the *X*-fiber firings occurring in a certain pattern. And so it is with you too. In the Martians, let us suppose, the having of visual experiences is realized in having *Z*-fibers firing, and particular qualia are realized in certain properties of the patterns of *Z*-fiber firings. Let *X*1 be a realization in me of the quale *Q-BY*, and let *Z*1 and *Z*2 be realizations in the Martians of a pair of qualia that constitute a blue-yellow invertible pair. Now there are various physical properties that can be instantiated in both us and the Martians, and among these are the disjunctive properties *X*1-or-*Z*1 and *X*1-or-*Z*2. Why shouldn't one of these be a physical realization of the quale *Q-BY*? Of course, it is not possible that *both* should be, since in the Martians *Z*1 and *Z*2 realize different and incompatible qualia. And I am supposing that the physical differences between the Martians and us are such that there would be no reason for picking one of these rather than the other as a realization of *Q-BY*. But what is it that disqualifies these disjunctive properties as realizations of qualia? The answer cannot be just that they are disjunctive. For any property that can be shared by things that are different in any way can be construed as disjunctive. You and I will be physically different in a variety of ways; at the very least, our DNA will be different. So the property *X*1, which we share, can be construed as the following disjunctive property: (having *X*1 and being the experience

of someone having my DNA) or (having $X1$ and being the experience of someone having a DNA different from mine). But this suggests yet another problem. How could we know, and what could make it true, that what realizes *Q-BY* in me is $X1$ and not a more specific property (a disjunct of $X1$) which cannot be instantiated in you—perhaps one that can be instantiated only in creatures having my DNA?

The problem raised is one about the individuation of qualia realizations.[15] What is it that makes it appropriate to say that $X1$, but not the disjunctive property $X1$-or-$Z1$ and also not the more determinate property of having $X1$ and being an experience of someone having my DNA, is a particular physical realization of a quale? I think the answer is given by the following rule. Property P realizes a quale if (1) it is a consequence of functional definitions and causal laws that, whenever different states of the same person share P, they are qualitatively identical in some respect (e.g., with respect to color qualia); (2) P is not equivalent to conjunctive property $P1$-and-$P2$ such that (i) $P1$ satisfies condition (1) and $P2$ doesn't, and (ii) P satisfies condition (1) because, and only because, $P1$ does; and (3) if P is equivalent to a disjunctive property *Pa*-or-*Pb* such that *Pa* and *Pb* both satisfy conditions (1) and (2), then it must be that *Pa* and *Pb* satisfy (1) and (2) because P satisfies them, and not that P satisfies them because *Pa* and *Pb* satisfy them.[16] Let me illustrate this with an example.

[15] I am grateful to John Bennett and Richard Boyd for making me aware of this problem and of the inadequacy of earlier attempt of mine to solve it.

[16] In an earlier version of this paper clause (ii) of condition (2) did not contain the word 'only', and condition (3) did not refer to condition (2). That version of the rule was open to the following counterexample (due to Mr. Mark Johnston). Suppose that the brain contains a "backup system" in which qualia are realized quite differently than in the primary system. In the primary system a particular quale is realized by property *Pa*, and in the backup system it is realized by *Pb*. If at a particular point t in a person's life the backup system takes over from the primary system, then the person's pre-t experiences characterized by *Pa* will be color-qualia-identical to the person's post-t experiences characterized by *Pb*. It would seem offhand that the disjunctive property *Pa*-or-*Pb* should count as a realization of this quale. But though this disjunctive property can be presumed to satisfy conditions (1) and (2), it will fail condition (3) (as originally formulated); for it will satisfy (1) because *Pa* and *Pb* do, whereas, if it is to satisfy (3), it would have to be because it satisfies (1) and (2) that *Pa* and *Pb* do.

Actually, the objection as just stated does not work even against the original formulation, for the disjunctive property *Pa*-or-*Pb* will not in fact be a realization of the quale (although a closely related property will be). If before t an experience realized in the primary system has property *Pa* and after t an experience realized in the backup system has property *Pb*, this will not amount to those experiences' being color-qualia-identical unless the backup system and the primary system are connected in an appropriate way (e.g., so that they will contribute jointly to the person's recognizing after t things he had seen before t). Let C be a property an experience has in virtue of belonging to a brain in which such a primary system and

Suppose the property *X*1, which can be instantiated in both you and me, satisfies condition (1); more specifically, it is a consequence of functional definitions, including the functional definition of qualitative identity, that whenever different states of the same person share *X*1 they realize experiences identical with respect to color qualia. Suppose, further, that *X*1 satisfies condition (2) as well. This means, among other things, that my having the particular DNA I do have plays no essential role in the physiological explanation of the behavior that manifests the existence of my qualitative states and the qualitative similarities and differences between them. This might be because what matters, for purposes of such explanation, is the organization of the brain at the neuronal level, so that, as long as the neurons are such as to interrelate causally in certain ways, differences in their microstructure are irrelevant. On this supposition, of course, the conjunctive property of having *X*1 and being an experience of someone with my DNA is ruled out, as a realization of a quale, by condition (2).[17] Both *X*1

---

such a backup system are appropriately connected. What will satisfy (1) in such a case is not *Pa*-or-*Pb* but rather (*Pa*-or-*Pb*)-and-*C*. But the original formulation of the rule is nevertheless in trouble. On a natural interpretation of (2), (*Pa*-or-*Pb*)-and-*C* fails to satisfy condition (2); for its first conjunct satisfies (1) but its second conjunct doesn't, and it is at least plausible to say that the property as a whole satisfies (1) because the first conjunct does. Moreover, (*Pa*-or-*Pb*)-and-*C* is equivalent to the disjunctive property (*Pa*-and-*C*)-or-(*Pb*-and-*C*), and this might be said to fail condition (3) on the grounds that it satisfies condition (1) because its disjuncts do.

The amended version of the rule in the text avoids these difficulties. The inclusion of 'and only because' in (2) enables (*Pa*-or-*Pb*)-and-*C* to satisfy (2). For it is not the case that the latter satisfies (1) only because its first conjunct does; sometimes (when the similarity is between experiences realized in different subsystems) the second conjunct, i.e., *C*, plays an essential role. And (*Pa*-or-*Pb*)-and-*C* [or, equivalently, (*Pa*-and-*C*)-or-(*Pb*-and-*C*)] does not fail the amended version of (3), since it is not the case that it satisfies condition (2) because its disjuncts do, rather than vice versa. For *Pa*-and-*C* and *Pb*-and-*C* satisfy condition (2) because of the essential role played by *C*, a common conjunct of both of them, in making *Pa* experiences realized in the primary system color-qualia-identical to *Pb* experiences realized in the backup system. And this seems to me tantamount to saying that they satisfy (2) because (*Pa*-or-*Pb*)-and-*C* satisfies it, and not vice versa.

[17] It is worth noting that, if condition (2) did not include clause (ii), *X*1 would be ruled out as a quale realization as well, contrary to what we want. For let *Z*1 be (as before) a property of Martian states which satisfies condition (1), and let *Q* be any property of states which is independent of *X*1 and *Z*1 and is such that possession of *Q* in the absence of *X*1 or *Z*1 is not enough to satisfy (1). Then *X*1 is equivalent to the complex conjunctive property (*X*1-or-*Z*1)-and-[*X*1-or-(*Q*-and-not-*Z*1)]. (Here I am indebted to John Bennett.) But whereas the first conjunct of this, *X*1-or-*Z*1, satisfies condition (1), it surely cannot be said that it is *because* this conjunct of *X*1 satisfies (1) that *X*1 satisfies it; on the contrary this conjunct satisfies (1) because both *X*1 and *Z*1 satisfy it. Whereas it plainly is true that the conjunctive property expressed by 'has *X*1 and is an experience of someone with my DNA' satisfies (1) because *X*1 satisfies it, given what we are assuming about the case. Thus the latter property, but not *X*1, is ruled out by condition (2).

and the Martian property $Z1$, we can suppose, satisfy conditions (1) and (2). But the disjunction of $X1$ and $Z1$, although it satisfies conditions (1) and (2), is ruled out by condition (3)—clearly this disjunction satisfies (1) because its disjuncts do, and not vice versa. We can coherently suppose, however, that $X1$ itself satisfies (3) (and, if we like, that $Z1$ does also). Though it is not coherent to suppose that $X1$ is not equivalent to a disjunctive property, it is coherent to suppose that it is not equivalent to any disjunctive property that does not violate condition (3). And if this supposition is true, then $X1$ will be a qualia realization of the sort we are looking for.

Conditions (2) and (3) both make use of the notion of something's being the case *because* something else is the case, where this is not simply a matter of the latter thing's being a logically or nomologically sufficient condition of the former. I have no analysis of this notion to offer; but it seems to me that it is clearly a notion we do have, and one that has application. The word 'because' of course signals that something is being said to be explanatory of something. And, on my account, what singles out certain properties as realizations of qualia is the fact that they are suited, and have the right degree of specificity, to play a certain explanatory role. Some properties will be ruled out as too specific because they are analyzable into conjunctions having conjuncts irrelevant to the explanatory role. Others will be ruled out as not specific enough, in that they are analyzable into disjunctions of properties, each of which plays the relevant causal role by itself. But if I am not mistaken, those which satisfy conditions (1)-(3) are just those which are capable of playing the causal role and have just the right degree of specificity.

VI

What I have just been saying was addressed to what I earlier called the "metaphysical problem." Where does all of this leave us with respect to the epistemological problem? On the account I have suggested, there is no reason in principle why we should not be able to discover whether the color experiences of different human being are qualitatively comparable and, if they are, whether they are qualitatively similar in similar circumstances or, on the other hand, are spectrum-inverted relative to each other. We could discover this by finding how qualia are realized in the brain and by determining whether the relevant physiological similarities hold between the brains of different human beings.[18] But, given that

[18] It may be thought that, if Martian color experiences are not qualitatively comparable with our own, there is one thing that is in principle unknowable—we can-

such physiological investigations are still far in the future, what are we to say about our present epistemological situation? Do I have good reason right now to think that my experiences of violets and marigolds are similar to those of others, or even that my color experiences and those of others are qualitatively comparable?

If we do have good grounds for such beliefs, I think it will be an essential part of these grounds that the creatures whose experiences are being compared are members of a single species and, therefore, can be presumed to share a genetic endowment. By and large, different members of our own species have color-quality spaces having the same structure—i.e., they make the same color discriminations, see the same similarity relationships between objects, and so forth. Even in the exceptional case of color-blind people, the breakdown of this structural uniformity is very limited. The existence of this uniformity can scarcely be a coincidence, and it calls out for an explanation in terms of our shared genetic endowment. The situation would be altogether different if the structural similarity in question were that between human color-quality spaces and those of my hypothetical Martians, supposing them to exist. The existence of such an interspecies similarity could very well be a coincidence; and in any case, it could not be explained by a shared genetic endowment, since, *ex hypothesi,* the Martian evolutionary history would be entirely independent of our own. The similarity between different Martians could be explained in terms of their genetic endowment, and the similarity between different humans could be explained in terms of ours; but the similarity between humans and Martians would have to be explained in some quite different way, if at all.

---

not know "what it is like" to have experiences possessing the Martian color qualia (cf. Thomas Nagel, "That Is It Like to Be a Bat?" in his *Mortal Questions*). There is a sense in which this is a true, but it does not imply that there would be facts about Martian experiences which would be unknowable by us. If, as I have suggested, qualia can be identified with disjunctive physical properties (the disjunctions of their possible realizations), there is no reason in principle why we should not be able to pick out the Martian qualia, assign names to them, and know which of them characterize the experiences of a Martian on a given occasion. Granted, this would not be to know what it is like to have experiences characterized by these various qualia. But what would it be to know this? I suggest that to know what it is like to have an experience having a certain quale is (a) to have such experiences in one's own repertoire of possible experiences, and (b) to be able to recognize such experiences as such "introspectively," i.e., simply by having them. Where we fail to satisfy condition (b) but not condition (a), there is no reason to suppose that our failure to "know what it is like" is irremediable. And our failure to satisfy (a) in the case of Martian color experience, which presumably would be irremediable, would not as such preclude us from knowing any facts; it would merely preclude us from knowing certain facts in a certain way [namely that way which involves the satisfaction of (b)].

What would seem the simplest explanation of the uniform structure of human color-quality spaces is that, as part of our shared genetic endowment, we are all "wired" in such a way that the same environmental stimuli give rise to the same color qualia in our visual experiences. This, of course, would guarantee both that color experiences of different human beings are qualitatively comparable and that, under similar conditions, they have qualitatively similar color experiences—i.e., that there is no intersubjective spectrum inversion among human beings. But if we suppose that our experiences are not qualitatively comparable or that some of us are spectrum-inverted relative to others, then, in order to explain this uniformity of structure, we must suppose there is something in our genetic endowment that compensates for the differences there are in our qualia, so as to make the differences cancel out so far as their effects on behavior are concerned. This is not impossible, but presumably it would call for more complex neural mechanisms than the first arrangement; and from an evolutionary standpoint it seems unlikely that we would have the more complex rather than the less complex arrangement. Obviously no such considerations could be used to support the claim that our experiences are qualitatively comparable with, or similar to, those of my Martians. And this is as it should be; if we know that Martian evolutionary history is completely independent of our own, then even before we discover the physiological differences between them and us, it will not be reasonable to think that our color experiences and theirs are qualitatively similar, or even qualitatively comparable.

It seems to me that if we are indeed entitled to think that, as John Locke put it, "the sensible *Ideas*, produced by any object in different Men's Minds, are most commonly very near and undiscernibly alike" (*loc. cit.*), our entitlement will rest on an empirical argument along the lines of that just sketched. I shall not attempt to elaborate the argument or to rebut the various objections that might be made to it; for part of the point I want to make is that the argument is at best fairly weak. It gives its conclusion—that our experiences are qualitatively comparable with those of others and that spectrum inversion does not occur—the status of a hypothesis which, although perhaps reasonable on the basis of current evidence, could easily be overthrown by future discoveries about how mental states are realized in the brain. We do not, if this is so, know with certainty at the present time that this conclusion is true. If one finds this counterintuitive, one should remember a point I made earlier: given the distinction between qualitative similarity and intentional similarity, our everyday claims about the experiences of

others need not be taken as implying anything specific about the qualitative character of those experiences, and thus the certainty we would like to ascribe to some of our everyday claims need not be undermined by the admission that we lack certain knowledge of the qualitative character of the experiences of others.

A more expeditious, if not wholly satisfying, treatment of the epistemological problem is that of John Locke:

> I am nevertheless very apt to think, that the sensible *Ideas*, produced by an Object in different Men's Minds, are most commonly very near and undiscernibly alike. For which Opinion, I think, there might be many Reasons offered; but that being besides my present Business, I shall not trouble my reader with them, but only mind him, that the contrary Supposition, if it could be proved, is of little use, either for the Improvement of our Knowledge, or Conveniency of Life: and so we need not trouble ourselves to examine it (*loc. cit.*).

Locke was not always so anxious not to "trouble his reader," and one could wish that he had not been so here.

SYDNEY SHOEMAKER

Cornell University

NOTE ADDED IN PROOF

Saul Kripke has called my attention to a mistake in what I say in section V about the realization of qualia. Suppose that in person A quale Q_1 has the properties P_1 and P_2 as its only possible realizations, that in person B the quale Q_2 has properties P_2 and P_3 as its only possible realizations, and that in person C the quale Q_3 has properties P_3 and P_4 as its only possible realizations. Suppose further that there is no possible creature in which P_1 and P_4 can both be instantiated. According to what I say in section V, the last supposition implies that P_1 and P_4 are not "qualitatively identical," and thus that Q_1 and Q_3 are not the same quale. Yet plainly Q_1 and Q_3 would have to be the same quale, by the transitivity of identity. Let R be the relationship that holds between two qualia realizations when (a) they can be instantiated in the same creature (i.e., it is possible for there to be a creature in which both are instantiated), and (b) they are such that when they are so instantiated the experiences that have them are qualitatively identical in some respect. Then the relation "being realizations of the same quale" should be equated not with R but with its ancestral R^*; and what I must stipulate about my imaginary Martians is not only that they do not share any qualia realizations with us, or have any that stand in R to any of ours, but also that they do not have any qualia realizations that stand in R^* to any of ours. I believe, however, that an elaboration of the considerations adduced in footnote 16 will show that it is to a very limited extent, if at all, that qualia realizations can be related by R^* without being related by R.

S.S.