Qualia and Private Language

Carl Ginet
Cornell University

When you and I gaze at the sky of a perfect day, each of us has a subjective visual experience that can be described in various equivalent ways: it is an experience as of a blue expanse, or an experience as if seeing a blue expanse, or an experience of being appeared to blue-expanse-ly. This experience is subjective in that its existence, though caused by a blue expanse out there, does not entail anything about the external environment. The experience has intentional content: it represents, among other things, that there is a blue expanse in view. Sydney believes, and so do I, that such an experience has this property of representing blue out there in virtue of a non-representational property it has, on which the representational property supervenes in a way analogous to the way a painting’s representing a blue sky supervenes on the canvas’s having on it patches of blue pigment. (I also believe, as does Ned Block and did Sydney at one time, that this non-representational property of the visual experience is an introspectible property, one that the subject is directly aware of.) It is the basis of the experience’s representing blue in that there is a whole spectrum of contrary non-representational properties such that, had the experience had some property in the spectrum other than the one it has, it would have represented a color other than blue. These alleged non-representational properties of visual experience—and analogous properties for other modes of sense experience—have come to be known as *qualia* (thanks to Sydney’s writing as much as anything).
I count myself a friend of qualia and my objective here is to defend them. But of course I won’t try to defend them against all of the plausible objections that have been or might be raised. I shall consider only what threat to them there may be in (what I take to be) the private language argument and how that threat might be met. Like Sydney, I will stick with the example of the color qualia of normal visual experience, the qualia that, as he and I think, underlie its representing colors.

By “the private language argument,” I mean a certain argument that I extract from some things Wittgenstein says in *Philosophical Investigations* (hereafter referred to as PI). The suspicion that the private language argument may pose a threat arises from a fact about color qualia that Sydney recognizes, namely, that individual qualia are not functionally definable. From this it follows that no amount of similarity in your and my discriminating and classifying responses to external color stimuli can entail that the same stimuli cause us to experience the same qualia, rather than, say, qualia that are spectrum-inverted with respect to each other, or even qualia so different that none that you experience is the same as any that I experience. For example, for all that stimulus-response evidence can show, the quale you experience when looking at a red traffic light is the same as the quale I experience when the light changes to green, or perhaps is quite different from any quale I ever experience.

The worry is that, if this is so, then color qualia are epistemically private objects in that it is impossible for anyone other than the subject of any particular quale-instance to know what that instance is like (to know what quale it instantiates). And the private language argument shows, I believe, that if the fact that qualia are not functionally definable does mean that quale-instances are private objects, then it also means that qualia are impossible.

In PI § 202, Wittgenstein gives an argument against the possibility of “obeying a rule privately.” The conclusion of this argument, that private rules are impossible, together with the premise that if our sense-experience instantiates qualia, then it must be possible for us to have terms that mean those qualia, will give us the consequence that quale-instances cannot be private objects. Briefly stated, the argument is that it is impossible to have a term mean such an epistemically private object because the rule for applying the term which gives the term its meaning—a rule of the form “Apply term T to a case iff it has property P”—would have to be a private rule. What I call “the private language argument” is the argument of PI § 202, plus the argument that uses its conclusion to show that there cannot be private objects. The upshot of the private language argument, as I make it out, is that if the fact that qualia are not functionally definable means that quale-instances are private, then there can be no such things as qualia.

In what follows I will first explain how the fact that individual qualia are not functionally definable seems to mean that quale-instances are private. Then, lest we friends of qualia be tempted to suppose that we can live with the consequence that no one can know what another’s quale-instances are like, I will explain how the private language argument shows that this consequence has the further, unacceptable consequence that there could be no language of quale terms. Finally, in reaction to this, I will consider how it might be argued that, despite their functional indefinability, qualia are not private, that it is possible for one subject to know what another’s qualia are like.

Individual qualia are not definable in terms of their functional roles in the lives of their subjects. This was first noted by Ned Block and Jerry Fodor, and a full and convincing explanation of why it is so has been given by Sydney. Here I can only sketch the explanation briefly. The argument starts with the premise that it is possible for there to be a change in the quale-instances produced in a subject by each sort of external color stimulus which is such that after the change no specific kind of color stimulus produces instances of the same quale as it did before the change; but the subject is disposed to make the same sorting of the external color stimuli into distinct classes and degrees of similarity, and, after a period of readjustment to the linguistic habits of her community, the same verbal classification of them as she did before the change. Call this sort of sweeping, systematic change a radical quale change. If this sort of radical introsubjective quale change is possible, then it must also be possible for there to be a similar radical difference between the qualia experienced by different subjects. And if such radical intersubjective difference is possible, then individual qualia are not functionally definable: they are not definable in terms of the stimuli that are wont to cause them or the effects on other mental states and behavior they are wont to cause. (Sydney usually illustrates the possibility of radical intra- and intersubjective quale difference by talking of spectrum inversion, even though, as he acknowledges, it is not clear that the structure of our color quale-space is such that a spectrum inversion would, after verbal accommodation, make no difference to our reactions to color stimuli. But, as he points out, the possibility of radical quale change in that specific way—by spectrum inversion—is not essential to the argument. It will do equally well if it is possible for there to be sweeping, systematic change that makes every post-change quale different from every pre-change quale.)

The fact that individual qualia are not functionally definable seems to provide reason to think that no one can know what another’s quale-instances
are like. How can I know that you and I experience the same color qualia as we gaze at the blue sky? It seems that all I could possibly have to go on is what stimulus causes your experience and how you are disposed to react to it in behavior that describes it or compares it with other experiences of yours. But if qualia are not functionally definable then all of that stimulus-response evidence could be the same as it is with me and yet your experience instantiates a different quale than mine does. Stimulus-response evidence could not discriminate between the possibility that our quale-instances are the same and the possibility that they are radically different, and so, it seems, it could not give me any reason to believe that one possibility rather than the other obtained.

Consider the position I would be in were you to experience a radical quale change and report the fact to me. Then, it seems quite clear, no stimulus-response evidence could tell me whether your qualia before the change or your qualia after the change were the same as mine, and so I wouldn't be able to know which is the case. That is, I would not then know that your qualia before the change were the same as mine. It seems to follow that I also couldn't have known before the change that your qualia then were the same as mine (even if they were). And if I failed to know this before the change, this failure does not seem to depend on the radical change's actually occurring but only on its being possible, since it does not depend on my having reason to think that any such change was going to occur. That is the sort of reasoning that leads one to think that if qualia are not functionally definable, then one cannot know by means of stimulus-response evidence whether another's quale-instances are the same as one's own.

II

If we find this reasoning persuasive, should we friends of qualia simply accept the consequence that quale-instances are epistemically private? I think not. I think it is too much to swallow, because I think that the private language argument, properly understood, shows that if qualia were epistemically private then we could have no terms that mean them. And I am most reluctant to accept that we can be acquainted with properties that we cannot name.

The private language argument has as its main lemma the proposition that it is impossible to obey a rule privately. Wittgenstein gives a short argument for this proposition at PI § 202: "to think one is obeying a rule is not to obey a rule. Hence it is not possible to obey a rule 'privately'; otherwise thinking one was obeying a rule would be the same thing as obeying it." What is meant by this talk of obeying a rule privately? Later, in PI § 243, where Wittgenstein introduces for consideration the idea of a private language, he says, "The individual words of this language are to refer to what can only be known to the person speaking; to his immediate private sensations. So another person cannot understand the language." This leads me to suggest that to say that a subject S obeys a particular rule R privately means the following:

S understands R, and follows R correctly in particular cases, and it is not possible that there be a subject distinct from S who understands R.

The argument of PI § 202 has two premises:

If it is possible to obey a rule privately, then thinking one is obeying a rule is the same thing as obeying it.

But it is possible to think one is obeying a rule when one is not.

Why should we accept these premises?

Let us consider the second one first. What possibility does it assert? Certainly at least the following:

P-2 For any (subject) S and (rule) R: it is possible that S thinks she understands R but she does not (but instead either follows some other rule she confuses with R or follows no rule at all).

So construed, the second premise looks hard to deny. If we express the conclusion inferred from these two premises as follows:

P-3 It is necessary that, for any S and R: it is not the case both that S understands R and that it is impossible that there be a subject distinct from S who understands R.

and we take the first premise to be spelled out as follows:

P-1 It is necessary that for any S and R: if S understands R and it is impossible that there be a subject distinct from S who understands R, then it is impossible that S thinks she understands R when she does not.

then the conclusion follows from the two premises by modus tollens.

Now what argument is there for P-1? I suggested an argument for it in my 1970 article, "Wittgenstein's Argument That One Cannot Obey a Rule Privately." There, I spoke of the totality of the potential users of the rule and suggested that Wittgenstein infers P-1 from the premise that the agreement of the overwhelming majority of the potential users of the rule as to what accords with it determines what accords with it (where "determines" means is a sufficient truth condition for). Call this the "majority-agreement criterion." You see that P-1 follows from this as soon as you note that the antecedent of P-1 entails that the potential users of R would be restricted solely to S.
I attributed to Wittgenstein the view (with which I was then strongly inclined to agree) that it is an illusion that there could be any other sort of definitive criterion, i.e., sufficient truth condition, for what accords with a rule. I am no longer inclined to agree with this view. I am now inclined to think the following:

(1) this sort of argument for the majority-agreement criterion is not sound because it depends on overlooking a certain important distinction;
(2) there are reasons for rejecting this criterion; but
(3) those reasons suggest another way of supporting P-1.

Let me explain each of these points in turn.

(1) We must distinguish the following two things: what constitutes a person’s understanding what it is for a case to be in accord with a rule (this is a state of the person who has the property that in some way contains all the possible applications of the rule) and what constitutes a case’s being in accord with a rule. Wittgenstein has persuaded me that the first—understanding a rule—could not be constituted just by having in mind an expression of the rule, by contemplating a paradigm, or by having an image of a paradigm (or even contemplating a Platonic Form), or indeed by anything other than an enormous number of counterfactual conditional (that is, dispositional) facts as to what a person would judge about particular cases. But it does not follow that the second—a case’s being in accord with a rule—could be nothing other than such counterfactual conditional facts.

That there is in common to all the cases that accord with the rule a property, perhaps a quale, which is the necessary and sufficient condition for the cases all being in accord with the rule, may be quite independent of any facts as to what anyone would judge about the cases. What is not independent of such facts about judgments is only someone’s being able to detect this property, knowing how to recognize it.

(2) Now let me give positive reasons for rejecting the idea that what the judgments of the majority of potential users of a rule agree on defines what accords with the rule.

Let’s start with the fact that we are intuitively inclined to suppose that for no rule are a single individual’s judgments—all those she does or would make in particular cases—the definitive criterion of what accords with the rule. Why? Well, because it seems so clear, with respect to any rule and any individual whose judgments about the application of that rule are in fact correct, that her judgments could have been partly (or even wholly) incorrect. For example, anyone could have been the quirky individual in W’s example who goes astray after 1,000 in following the rule for continuing a series, “Always just add 2 to the preceding number.” No one is, as a matter of logical, conceptual, or metaphysical necessity, designated as the one whose judgments, no matter what they are, are definitive of what accords with that rule.

Further, no one is, by definition or stipulation, exempt from going so wrong in trying to follow a particular rule that they are correctly following no rule whatever. If S tries but fails to follow a particular rule R, her failure may be explained by her following a different rule, which she confuses with R. But this isn’t the only possible explanation. S’s judgments (actual and counterfactual) about the dictates of R may be such that there is no rule that she is in those judgments applying correctly. Let us call such a person a blank misunderstander of R.

Actually, the definition of a blank misunderstander needs to be slightly more complicated in order to deal with the fact that, for any set of particular possible cases—and thus for any set that a blank misunderstander would judge to accord with a particular rule R—there is some property (however logically gerrymandered it might be) that is in common to all and only those possible cases, and hence there is a rule (different from R) to the effect that all and only cases with that property are in accord with it. It may nevertheless be the case that our blank misunderstander of R is quite incapable of detecting any such uniquely common feature in the particular cases and so she could not be going by that other rule and so her doing so could not explain her set of dispositions.

So we define a blank misunderstander as follows:

S is a blank misunderstander of R =def S is disposed to make judgments about the dictates of R in various possible cases, but those judgments and S’s detection capabilities are such that there is no rule that those judgments align correctly.

Now, a natural question to raise about the majority-agreement view of what determines accordance with a rule is this: if it is possible with respect to any given individual that she should be a blank misunderstander of a given rule R, why is it not possible that every individual, every would-be user of R, should be a blank misunderstander of R?

Imagine a world of just 1,000 people learning arithmetic as we do. If it is possible that any one of them should be a blank misunderstander of, say, the rule for constructing an endless sequence of numbers that says, “The first number is 0 and for any number n in the sequence, the next number in the sequence is n + 2,” then it is possible that any 10 or 100 or 400 of them should be blank misunderstanders of that rule. Why not all 1,000 of them? Suppose that there is one, S, whose judgments regarding the application of this rule (all of the judgments S is disposed to make) constitute blank misunderstanding of it. Why couldn’t it have been that all 1,000 of them were just like S, having those same judgments, and thus having been blank misunderstanders of that rule?
They certainly could all be disposed to those same judgments, but it
does not follow that they would then all be blank misunderstanders of that
rule. For what would then make it that rule that their judgments were about?
When only part of them are blank misunderstanders and there are others
whose judgments do pick out that rule, we can say that the judgments of the
blank misunderstanders are judgments about that rule because, but only
because, they are so related to the correct judges as to make it right to say
that it is the rule picked out by the correct judges’ judgments that the blank
misunderstanders intend to be applying. For example, the correct judges are
the teachers of the blank misunderstanders and the blank misunderstanders
are trying to learn the rule their teachers are trying to teach them. But, if
everyone’s judgments were like those of the blank misunderstander, then
there would be no such basis for saying that it is that particular rule that they
were all attempting to apply. Nor would there be any basis for selecting any
other particular rule that their judgments are trying to apply.

In particular, there would be no reason to say that all of the blank
misunderstanders’ agreeing judgments simply define a rule (by defining what
accords with it) and are therefore correct applications of a rule that their
agreeing judgments define. It would, rather, be a case where they are all
under the impression of following a rule but there is no rule of which it is
ture that they think they are following it. From the fact that there is no par-
ticular rule that they all mistakenly apply in the same way, it clearly does
not follow that there is a rule that they are all correctly applying. But this
would follow if the majority-agreement view about what determines accor-
dance with a rule were correct.

I now take a more “naive” view about what determines accord with a
rule that avoids this error. What we should say, I believe, is that what makes
cases accord with a rule is an objective property belonging to the cases, that
is, a property they have independently of any individual’s or community’s
judgment about whether or not they accord with the rule. Every objective
property determines a rule for a possible term meaning that property, a rule
such that the cases that accord with the rule are all and only those with
the property. Such a rule is understood by someone only if they are able in suit-
able circumstances to detect the property that determines what accords with
the rule. Many such rules exist only in the abstract and are never grasped or
referred to by anyone. And a particular such rule can be referred to only if
there is someone who grasps it.

If we take this view then we must hold, I think, that for no rule R is it
possible that everyone be a blank misunderstander of R. That is:

(A) For any rule R: it is impossible that every subject S is a blank
misunderstander of R.

This follows from what is required for it to be the case that someone’s inten-
tion refers to a particular rule. For any particular rule that we all correctly
understand, it is possible that there should appear among us an S who is a
blank misunderstander of that rule, someone who seeks to understand this
rule we try to teach her and thinks she does understand it but in fact blankly
misunderstands it. Given this possibility, it is clear that there is another pos-
sible world in which everyone’s judgments and detection capabilities are the
same as those of this possible S who is a blank misunderstander of our rule.
But in that possible world there would be no basis for saying of any of its
inhabitants that it is our rule, or any other particular rule, that their judg-
ments incorrectly apply.

There would be no basis for saying this, that is, unless you suppose,
with the majority-agreement view, that the agreed judgment of everyone
defines what accords with a rule, and therefore it is sufficient for there being
a rule they all correctly understand that they agree about what it dictates,
unless, that is, you suppose the following:

(B) It is necessary that if, for any S, S judges that there is a rule R
that dictates such-and-such applications, then there is a rule R
that dictates such-and-such applications.

But one need not suppose this. Compatibly with accepting A, one can say
instead, about the possible world in question, that there is no particular rule
that its inhabitants, in their judgments of the form “This case is in accord
with the rule,” refer to and think they follow. In being forced to accept A, by
considerations having to do with what is necessary for reference to a partic-
ular rule, the objective-property view is not forced to give way to the major-
ity-agreement view. This is clear once we distinguish A from B.

(3) However, what is required for reference to a rule provides another
way of supporting premise P-1 of the private-rules argument (a way other
than that provided by the majority-agreement view). So that argument does
not need the majority-agreement view and can be embraced by the naive
objective-property view.

The important point about what is required for a person’s thought to
refer to a particular rule can be put as follows:

α: It is necessary that, for any R and S: if S thinks she understands
R, then there is someone who understands R and is suitably
related to S.

From α we can infer:

β: It is necessary that for any R and S: if it is possible that (a) S
thinks she understands R but does not understand R, then it is
possible that (b) there is someone distinct from S who under-
stands R.
To see this, note that, given α, the first conjunct of (a) in β entails

There is someone who understands R,

and the second conjunct of (a) entails

Anyone who understands R is distinct from S.

But these two together obviously entail (b). So altogether (a) entails (b). From this it follows that the possibility of (a) entails the possibility of (b), which is what β claims.

From β we can deduce P-1 (viz., that it is necessary that for any S and R: if S understands R and it is impossible that there be a subject distinct from S who understands R, then it is impossible that S thinks she understands R when she does not). It will be easier to see this if we adopt some abbreviations for the components of P-1, as follows:

- U(S,R) = S understands R.
- I(S,R) = It is impossible that there be a subject distinct from S who understands R.
- T(S,R) = S thinks she understands R but does not.

Using these abbreviations, P-1 can be expressed as follows:

\[ P-1 \text{ (abb)} \quad \text{It is necessary that, for any S and R: if U}(S,R) \text{ and I}(S,R) \text{ then it is impossible that T}(S,R). \]

And β can be expressed as follows:

\[ \beta \text{ (abb)} \quad \text{It is necessary that for any S and R: if it is possible that T}(S,R) \text{ then it is not the case that I}(S,R). \]

We quickly see that P-1 follows from this if we look at the contrapositive equivalent of P-1:

\[ P-1 \text{ (equiv)} \quad \text{It is necessary that, for any S and R: if it is possible that T}(S,R) \text{ then it is not the case that both U}(S,R) \text{ and I}(S,R). \]

I have already argued for premise P-2 of the private-rules argument, which we can now express as

\[ P-2 \text{ (abb)} \quad \text{It is necessary that, for any S and R: it is possible that T}(S,R). \]

So my defense of the private-rules argument is complete, and we can take it to have established its conclusion, which can now be expressed as follows:

\[ P-3 \text{ (abb)} \quad \text{It is necessary that for any S and R: it is not the case that both U}(S,R) \text{ and I}(S,R). \]

This conclusion can be used to argue that qualia-instances cannot be private in the sense that only their subjects can know what qualia they instantiate. The argument is suggested (though not constituted) by the inference that Wittgenstein makes in the passage from PI § 243 that I quoted earlier. There he speaks of a language that consists of words that "refer to what can only be known to the person speaking; to his immediate sensations," and then says, "So another person cannot understand the language." Applying this to qualia-instances in visual experience, we can spell out the inference that seems to be made here as follows:

(C) If the qualia instantiated in an experience can be known only to its subject, then any rules a subject follows in applying qualia-terms to her qualia-instances can be understood only by that subject.

From this and the lemma given us by the conclusion of the private rules argument, namely,

(D) There cannot be a rule for applying a term which is such that it is impossible for more than one person to understand it,

we can infer the conclusion,

(E) Therefore, it is not the case that the qualia instantiated in an experience can be known only to its subject.

A shorter way of putting this argument:

(C) If qualia-instances are private objects, then a subject's application of qualia terms to her own qualia-instances is guided by private rules.
(D) Private rules are impossible.
(E) Therefore, qualia-instances cannot be private objects.

Is the inference expressed in (C) to be granted? The antecedent of (C) entails:

(1) Only I can know what qualia my experience instantiates.

If we suppose, as is quite plausible, that

(2) You cannot know what rule I follow in applying a quale-term to my experience unless you know what quale is instantiated by the qualia-instances of mine that I take as guiding paradigms.

then it does seem to follow that only I can know what rule I follow in applying a quale-term to my sensations.

But this is not quite the consequent of (C), which is
Only I can understand (know the content of) the rule I follow in applying a quale term to my experience.

Does (3) follow from (1) and (2)? It may seem that it does not. Suppose you and I each use a rule for applying a certain term to one of our own color qualia, say, the term “Q1.” Now could it not happen, by coincidence, that the particular quale-instances of yours that you use as your guiding paradigms are qualitatively the same as those of mine that I use as my guiding paradigms? It just happens that you select the same quale instantiated in your experience to mean by “Q1” as I select in my experience to mean by “Q1.” In that case, even though, given (1), neither of us could know that the other follows the same rule, wouldn’t it still be true that we both do follow and understand the same rule?

At first blush, the answer may seem to be yes. But I think that the truth is no. In the case imagined, although the particular paradigm instances you and I use are alike, it does not follow from this that we follow the same rule, if reference to a particular paradigm instance is essential to the definition of the rule, to its content. And it is plausible to think that it is, to take it that the particular paradigm instance that guides my application of the term is, as Wittgenstein put it in PI § 16, an instrument in my language. Your rule and my rule are no more the same rule than would be the rule for the application of “glick” followed in tribe A, that an object is one glik long if it is the same length as that object (pointing to their paradigm glik stick), and the rule followed for the application of “glo” in tribe B that an object is one glo long if it is the same length as that object (pointing to their paradigm glo rod), where the two paradigms happened by chance to be the same length.

So I conclude that (3) does follow from (1) and (2) and, hence, that, given (2), (C) holds. If (C) holds and (D), the conclusion of the private rules argument, holds, then we have the conclusion (E) that quale-instances cannot be private objects in the sense that only their subjects can know what they are like (what qualia they instantiate).

IV

Given this conclusion and the argument given earlier that, owing to the functional indefinability of qualia, they are not intersubjectively knowable on the basis of ordinary stimulus-response evidence, what are we friends of qualia to do? If we wish to remain friendly to qualia, we must find some other basis on which one subject might know what another’s quale-instances are like.

In “The Inverted Spectrum,” Shoemaker mentions two things that might be thought to be such other possible ways for me to know what your quale-instances are like.

One is the possibility of our fusing into one person in such a way that that person can remember both of our experiences and thus directly compare their quale-instances. Sydney mentions this only in a footnote, and he notes that it is possible “only if human beings are fusible—and that seems rather questionable.” He does not explain why he thinks it is questionable, and it’s not clear to me that this possibility ought to be dismissed out of hand, but I won’t pursue it further here.

The other way, and the one in which Sydney puts stock, is that I might discover what type of neurophysiological state of mine realizes, for example, the color quale of my visual experience—say, of seeing blue—and then discover that you had the same type of neurophysiological state, from which, he suggests, I could infer that your physiological state realizes an instance of the same quale as does mine. I can, he says, in principle discover what neurophysiological property realizes the quale instantiated in my experience of seeing blue, despite the fact that that quale is not itself functionally definable. It is enough for the purpose that functional definitions are possible for what it is for an experience to instantiate a quale (what it is for it to have qualitative character) and for what it is for two qualia instantiations to be qualitatively the same. He says:

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. . . . 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.4

He goes on to say:

There is no reason in principle why we should not be able to discover whether the color experiences of different human beings 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.5

Sydney’s thinking, I take it, goes as follows: We can use the functional definition of being a quale-instance to identify the neurophysiological realizations of quale-instances. Then we use the functional definition of qualia identity across quale-instances to identify the neurophysiological property that realizes a specific quale. That is, I can discover in this way what neuro-
physiological property realizes in me the quale-instance I have when I see a blue surface in normal conditions, and I can in the same way discover the neurophysiological property that realizes in you the quale-instance you experience when you see a blue surface. If it is the same neurophysiological property in you and in me, then I can infer that our experiences instantiate the same quale. I need not, in making this inference, be troubled by the argument from the possibility that you or I may undergo a radical qualia change. For a radical qualia change in one of us would entail a radical change in the neurophysiological realizations of the qualia experienced in the same external circumstances. And this would allow us to determine whether it was before or after the radical change that you and I experienced the same qualia in the same circumstances: it was whichever time the realizations of our qualia-instances were the same in the same circumstances.

It seems to me that this reasoning makes a questionable assumption. It seems to assume that the same neurophysiological realization must mean the same quale, that each quale supervenes on a specific neural property, so that the current total neural state of a subject currently experiencing qualia metaphysically necessitates that it is those specific qualia she experiences. But is this assumption really any more secure than was the assumption that same stimulus-response means same quale? Doubt was cast on that assumption by the possibility of intrasubjective qualia inversion, or radical qualia change, wherein sameness of stimulus and response are preserved. But isn’t it also possible that there should be intrasubjective radical qualia change wherein sameness of neurophysiological realization is preserved? And if so doesn’t this in a parallel way cast doubt on this new assumption?

Does Sydney suppose that it follows from the functional definability of qualitative similarity that a radical qualia change entails a radical change in the realizations of the qualia? As far as I can see, this seems to follow only if we assume that the only way that the functional definition of qualitative similarity could be realized is for the realizations of the qualitatively similar instances to be themselves similar, to have a neurophysiological property in common. Now, that is, to be sure, a natural way for it to be realized and the first thing one is likely to think of, but I can’t see that it is entailed if qualitative identity is to be functionally definable.

There must be, of course, some neurophysiologically specifiable relation between two particular realizations of the same quale that realizes the functional definition of their qualitative identity. But, as far as I can see, this relation need not be, or need not always be, similarity between their neurophysiological realizations. It might sometimes be, say, a specific sort of intervening event between those neural states that systematically changes what quale is subsequently realized by each of the subject’s quale-realizing neural states (an event, that is, that induces a radical qualia change without inducing a change in the neurophysiological realizations that are caused by the same external circumstances). We discover, let us suppose, that a subject with an accurate memory of an earlier quale experience judges that it and a current quale experience are qualitatively the same iff either no such event has intervened and their neurophysiological realizations are the same, or such an event has intervened and their neurophysiological realizations are different in a certain specific way (dictated by the system of the radical change). She judges that the qualia are different in a certain specific way (dictated by the system of the radical change) iff either no such event has intervened and their realizations are different, or such an event has intervened and the realizations are the same.

Some may ask whether we could distinguish such an event’s causing a radical qualia change from its causing a radical memory deception. But I see no reason to suppose that the neurophysiological facts couldn’t possibly distinguish these, couldn’t assure us, that it is not damage to the subject’s memory that is responsible for her reporting a radical qualia change. It might be that the neurophysiological event causing the radical change leaves a trace that explains the subject’s regarding qualia before and after the change as different despite their having the same realizations, and it explains this without implicating any damage to the neurophysiological mechanisms underlying the subject’s memory of the qualitative character of her earlier experience.

If a radical change in your qualia were caused in this way, then, it seems to me, there would be nothing in the way that change is physically realized or in the comparisons between how qualia are realized in our two neural systems to justify either believing that before the change your qualia were in similar circumstances the same as mine or believing that after the change your qualia are in similar circumstances the same as mine. And thus knowledge of how qualia and relations of qualitative similarity and difference are physically realized would put me in no better position to know what your qualia are like than does stimulus-response evidence. It would, to be sure, remove any doubt I might have had, when my belief that your qualia are the same as mine was based only on stimulus-response evidence as to whether neurophysiological evidence would confirm that belief. But it would leave my belief vulnerable to essentially the same sort of skeptical argument that it was vulnerable to when it was based just on stimulus-response evidence.

How unwelcome should this conclusion be? Maybe the right attitude here is like the one Moore took toward skepticism about our claims to know the external world. We would in fact be strongly inclined to believe on the basis of the best stimulus-response evidence supplemented by the best neurological evidence—evidence such that no better evidence could possibly be
had—that others' qualia are the same as our own. Why ought not we to do so or why ought not we to think that we are on that basis justified in those beliefs and justified in thinking (as we ordinarily do) that those beliefs are knowledge? Why ought not we to treat it as a fundamental epistemic principle that such evidence does justify such beliefs and such claims to know? It seems that nothing else can justify them any better. The best possible evidence is, to be sure, compatible with the falsity of what we take it as evidence for, and so we are faced with a choice between nevertheless treating the best possible evidence as conclusive or else giving up thinking our beliefs or knowledge-claims are justified. Why should rationality dictate what is clearly against our nature, that we choose the skeptical alternative?

The argument for the skeptical alternative modeled on the one that we considered earlier would be this:

(1) If you came to know (on the basis of stimulus-response evidence) that another person had undergone radical qualia change (and you knew that the neurophysiological story behind this was as I imagined above it might be), then you would not know that her qualia before the change were the same as yours.

(2) Therefore, before the change you would not have known that.

(3) But if before the change, when you had no reason to think any such change was in the offing, you did not know that the other's qualia are the same as yours, then, whether or not any such change is in the offing, you do not know that.

(4) Therefore you do not now know that the other's qualia are the same as yours.

Plausible though it may seem, this argument is not valid. The fallacy lies in the first inference, from (1) to (2). That inference overlooks (or perhaps refuses to accept) that there can be knowledge that is vulnerable to the acquiring of evidence that goes against what one knew, that knowledge can be lost by acquiring new but misleading evidence. Suppose I have just come to know that Cornell defeated Princeton in their recent football game by hearing a newscast that reported that fact, and suppose that I were now to make the following inference:

Premise: If I were later to hear another newscast reporting that Princeton defeated Cornell, I would not then know that Cornell defeated Princeton.

Conclusion: I do not now know that Cornell defeated Princeton.

The inference from (1) to (2) is no better than this one. Which is to say, no good at all. So, having resisted this reason to think otherwise—and seeing no other—I conclude that we can know on the basis of the sort of evidence that is in principle available to us what others' qualia are like, and so there is no problem about our having terms that mean qualia. That is my bolder and qualia-friendly conclusion. Should it turn out (to my surprise) that there is some good reason to give up this conclusion, there would remain the more cautious (but qualia-unfriendly) conclusion that if we can't know on the basis of the sort of evidence available to us what others' qualia are like then we can't know this at all, and the private language argument shows that we cannot have terms that mean qualia and, therefore, that there is something wrong with the whole idea of qualia.

NOTES

5. Ibid., 348.
6. In the discussion following my presentation of an earlier version of this paper (at a conference to honor Sydney held at Cornell in September 1997), someone wondered whether the inference from (1) to (2) would look better if it were spelled out as follows:

(1') If you came to know (on the basis of stimulus-response evidence) that another person had undergone radical qualia change (and you knew that the neurophysiological story behind this was as I imagined above it might be), then you would know that, even if your belief before the change that the other's qualia were the same as yours was true, your evidence for it was not causally connected to the fact in the way required for the belief to be knowledge.

(2') Therefore, your current (stimulus-response) evidence for your belief, even if the belief is true, is not causally connected to the fact believed in the way required for the belief to be knowledge.

I think it does not look better. It looks no better than the following inference (made after I've heard a newscaster report the fact that Cornell defeated Princeton): "If I were later to learn that the newscaster I heard was just guessing when he said that Cornell defeated Princeton, then I would know that, even if my resulting belief that Cornell defeated Princeton was true, my evidence for it was not causally connected to the fact in the way required for the belief to be knowledge. Therefore, my current evidence for this belief, even if the belief is true, is not causally connected to the fact in the way required for the belief to be knowledge."