

Intentionalism and the inverted spectrum

Jeff Speaks

March 6, 2006

1	The challenge to minimal intentionalism	1
2	Intentionalist reply I: Appearance properties	3
3	Intentionalist reply II: Invert misperceives the color of the violet	4
3.1	The argument from ‘getting around in the world’	5
3.2	The argument from interpretability	5
3.3	The argument from the covariational theory of content	6
3.4	The argument from the supervenience of content on wide functional role	6
3.5	The argument from public language meaning + introspection	7

1 The challenge to minimal intentionalism

Remember that minimal intentionalism is just a supervenience claim: it is the claim that any two experiences of a given sense modality which have the same content will also have the same phenomenal character: i.e., if two such experiences have the same content, then ‘what it’s like’ to have the one experience will be the same as ‘what it’s like’ to have the other.

The first thing to see is how certain kinds of inverted spectrum examples can be used to challenge minimal intentionalism. Any counterexample to minimal intentionalism will be a pair of experiences which satisfy the following description:

e_1 and e_2 are experiences of the same sense modality, and have the same content (represent the world as being the same way). But e_1 and e_2 have different phenomenal character: what it’s like to have e_1 is different from what it’s like to have e_2 .

To see how the inverted spectrum can seem to provide an example of this sort, consider the passage by which Locke introduced the inverted spectrum scenario into the philosophical literature:

“Neither would it carry any Imputation of Falsehood to our simple Ideas, if by the different Structure of our Organs, it were so ordered, That the same Object should produce in several Men’s Minds different Ideas at the same

time; v.g. if the Idea, that a Violet produced in one Man's Mind by his Eyes, were the same that a Marigold produces in another Man's, and vice versa. For since this could never be known: because one Man's Mind could not pass into another Man's Body, to perceive, what Appearances were produced by those Organs; neither the Ideas hereby, nor the Names, would be at all confounded, or any Falsehood be in either. For all Things, that had the Texture of a Violet, producing constantly the Idea, which he called Blue, and those which had the Texture of a Marigold, producing constantly the Idea, which he as constantly called Yellow, whatever those Appearances were in his Mind; he would be able as regularly to distinguish Things for his Use by those Appearances, and understand, and signify those distinctions, marked by the Names Blue and Yellow, as if the Appearances, or Ideas in his Mind, received from those two Flowers, were exactly the same, with the Ideas in other Men's Minds." (*Essay on Human Understanding*, §II.xxxii.15)

Here we are asked to imagine two different subjects whose 'blue' and 'yellow' experiences are inverted. That is, when *A* looks at a violet, his experiences seems to him exactly how *B*'s experiences when looking at marigolds seem to him. But, Locke says, this description does not "carry any imputation of Falsehood" about the experiences of either. He argues for this claim when he says each "would be able as regularly to distinguish Things for his Use by those Appearances" as the other.

Let's suppose that Locke is right about this, and that two perceivers could have their experiences as of blue and yellow things inverted with respect to each other without either misperceiving the world at all. This thought can be turned into an argument against minimal intentionalism as follows (using, as has become standard, 'Invert' for the name of the perceiver whose experiences are inverted with respect to ours, and 'Nonvert' for the name of the perceiver whose experiences are not inverted with respect to ours):

1. When Invert and Nonvert look at a violet, their experiences have different phenomenal characters. (Premise)
 2. When Invert and Nonvert look at a violet, neither misperceives the color of the violet. (Premise)
 3. The violet is only one color. (Premise)
 4. Invert and Nonvert represent the violet as having the same color. (2,3)
 5. There is no other difference in the contents of the experiences of Invert and Nonvert. (Premise)
 6. The experiences of Invert and Nonvert have the same content. (4,5)
-
- C. The experiences of Invert and Nonvert have the same content, but different phenomenal characters. (1,6)

(1) and (2) are the assumptions built into the scenario that Locke imagines; but once we have these assumptions in hand, we need only the plausible further assumptions (3) and (5) to yield (C), which is inconsistent with minimal intentionalism. This is why the possibility of 'spectrum inversion without misrepresentation' is often thought to show that minimal intentionalism must be false.

We will now consider some ways that an intentionalist might reply to this argument.

2 Intentionalist reply I: Appearance properties

One line of response is to deny premise (5). If you say this, then you think that the experiences of Invert and Nonvert have different contents because there are some properties *other than color properties* with respect to which the contents of their experiences differ; one is then free to endorse the claim that they have different phenomenal characters without giving up on intentionalism. The natural next question is: what are these properties?

Shoemaker, in ‘Phenomenal Character’, gives one answer to this question. Shoemaker endorses the Lockean assumptions we sketched above:

“For suppose Jack and Jill are spectrum inverted relative to each other. When both are looking at a ripe tomato, their experiences will be markedly different in phenomenal character . . . Yet I would want to say . . . that the experiences of both represent the tomato, and represent it correctly, as being red.” (26)

Yet Shoemaker endorses minimal intentionalism (which is entailed by what he calls ‘representationalism’). So he must find *some* difference in the contents of Jack and Jill; that is, he must find some property which is such that the experiences of one but not the other of Jack and Jill represent the tomato as having that property. And we know from the above that this property will not be a color property.

Shoemaker presents his view as follows:

“Once these desiderata for a solution to the problem have been made clear, it begins to be clear what sort of solution it must have. How can the experiences of Jack and Jill represent the tomato differently and yet neither of them misrepresent it, given that the same information about its intrinsic nature is getting to both? They can only because the different properties their experiences attribute to the tomato are *relational* properties. So the bare bones of the solution is this. Let Q1 be the quale associated with redness in Jack, and let Q2 be the quale associated with redness in Jill. There is a relational property consisting in producing or being disposed to produce experiences with Q1. And there is one that consists in producing or being disposed to produce experiences with Q2. Jack’s experience represents the tomato as having the first of these relational properties, and Jill’s experience represents it as having the second of them. And in fact it has both. Neither property is the property of being red, which is also attributed to the tomato by the experiences of Jack and Jill. . . .” (27)

Objections to this view:

1. The property which Jack's experience but not Jill's ascribes to the tomato appears to be an intrinsic property of the tomato, not a relational one. The example of 'to the right of'.
2. There is a sense in which this view seems to make colors inaccessible to visual experience. This seems counterintuitive.

There is logical space for a number of different kinds of views of what's going on in these sorts of cases which are similar to Shoemaker's in that they find a difference in representational content in the perceptions of Invert and Nonvert which is not a difference in the colors their experiences represent objects as having. Thau, e.g., defends a view like this in *Consciousness and Cognition*. One could also hold a view which is like Shoemaker's in refusing to say that either Invert or Nonvert misrepresents the color of the tomato, but to say that one of them does misrepresent the appearance properties of the tomato, where as above appearance properties \neq colors.

3 Intentionalist reply II: Invert misperceives the color of the violet

One might think that there's an easier reply open to the intentionalist here: deny one of the Lockean assumptions. The core Lockean assumptions are (1) that two subjects could be spectrum inverted relative to each other, and (2) that neither of two such subjects would misrepresent the colors of things.

As Byrne notes, there are good reasons for doubting that there could be an agent spectrum inverted relative to you or me whose color-behavior was indistinguishable from our own. But the kinds of cases we need for the present argument needn't involve behaviorally undetectable spectrum inversion — it seems that cases of what Byrne calls 'shifted qualia' will do as well. So in the case of the use of spectrum inversion examples to argue against minimal intentionalism, giving up Lockean assumption (1) does not look like a plausible way out. (However, as we'll see, the possibility of certain kinds of inverted spectrum scenarios will become important again below, when we consider the plausibility of rejecting Lockean assumption (2).)

It might seem that denying (2) is a better response: maybe we should grant that the relevant kind of spectrum inversion is possible, but deny that in such cases both agents represent the colors of things correctly. Why this can seem like a plausible thing to say. An intuition pump: suppose that you put on spectrum-inverting lenses. Wouldn't that change the way that your experience represents the world as being? Why this kind of example would not convince someone with certain views about mental representation.

If one says this, the way out of the argument against intentionalism is clear: once we say that Invert and Nonvert do not both represent the colors of things correctly, then we will have to say that their experiences have different contents. And then we will not have an example of a pair of experiences which have the same content, but different phenomenal character. So we will not have a counterexample to minimal intentionalism.

The anti-intentionalist should then give the following kind of reply: Invert and Nonvert are alike with respect to some property type F ; there can be no representational differences between two subjects alike in their F -properties; therefore the kind of representational

difference posited by the intentionalist can't exist. There are several different plausible candidates for '*F*'.

3.1 *The argument from 'getting around in the world'*

One idea is that Invert and Nonvert are equally good in using their color vision to get around in the world; they are equally good at making distinctions between colors, and grouping things together on the basis of their colors. Given this, it might seem odd to say that one of them, but not the other, is in error about the colors of things.

Why it is not implausible to deny that the truth of one's beliefs supervenes on their practical usefulness.

3.2 *The argument from interpretability*

Another idea is that Invert and Nonvert are indistinguishable in terms of their behavioral dispositions, so that an interpreter of their mental states would assign the same contents to all of them. But the contents of mental states supervene on what an ideal interpreter of this sort would assign to an agent.

Note that this assumes a bit more than the previous objection. We can get a case where the phenomenal characters of Invert's and Nonvert's experiences differ while giving them equal ability to navigate the world without assuming a full-scale qualia inversion, but instead just a qualia shift, which is clearly possible. But since qualia shifts are behaviorally detectable, we'll need more than that to get the present objection going.

It might seem that this is easy enough: just imagine a full qualia inversion of the sort that Locke had in mind. Wouldn't this be behaviorally undetectable, as Locke seemed to think?

Suppose that the qualia inversion worked like this: the space of hues is thought of as a circle, with the colors red, green, blue, and yellow equally spaced. Then a qualia inversion would map reds to greens (and vice versa) and yellows to blues (and vice versa).

This kind of qualia inversion would be behaviorally detectable, because human color space is not symmetrical in the way that this inversion requires (see the Byrne piece for a nice discussion of this). Three reasons for the asymmetry (given the kind of spectrum inversion under consideration):

1. There are more distinguishable hues between blue and red than between yellow and green.
2. Saturation and brightness interact differently in the cases of yellow and blue. The most saturated shades of blue are darker than the most saturated shades of yellow.
3. Dark yellow — i.e., brown — is experienced as a qualitatively distinct hue from light yellow, whereas dark blue is not.

But if the differences between Invert and Nonvert are behaviorally detectable, then it is not clear that an ideal interpreter would assign the same contents to their mental states. So the objection from the supervenience of content on what an ideal interpreter would say does not go through.

A response: even if our color space is not symmetrical in a way which would permit undetectable spectrum inversion, the color space of some other sort of creature could be. Since minimal intentionalism is necessary if true, we can run the argument using those (perhaps merely imaginary) creatures. E.g.: creatures with only black/white vision.

Why one might doubt that such examples are genuinely possible. Counterexamples to the link between conceivability and possibility.

A second reply: what reason do we have to believe that mental representation could not outstrip what an ideal interpreter would say?

3.3 The argument from the covariational theory of content

Another way of filling out the inverted spectrum argument against intentionalism is via what seems like a plausible theory of mental content, at least for observable contents: the idea that the content of an inner state for an agent is whatever property usually causes (or in ideal conditions would cause) the agent to be in that state. How this can be used to defend the idea that the experiences of Invert and Nonvert have the same content.

A strength of this way of spinning the argument is that it can be run with just qualia shifting examples, without any stronger assumptions about the possibility of undetectable spectrum inversion.

Some problems for the covariational theory of content. Why what is really needed to get the problem going is just that (normal, ideal, whatever) covariation with a property is a necessary condition for representation of that property.

3.4 The argument from the supervenience of content on wide functional role

If one doubts the plausibility of the covariational theory of mental content, then one can re-run the objection using the more plausible idea that the content of a representation supervenes on its 'wide functional role' (roughly, its pattern of causal connections to property exemplifications and other internal states).

Why this version of the anti-intentionalist argument requires substantive assumptions about the possibility of inverted spectrum cases; if behaviorally undetectable inverted spectrum examples are impossible, so is spectrum inversion between functional duplicates.

A return to the case of black/white vision. A reason to doubt that inversion among functional duplicates is possible even in this case.

Is it plausible to deny the supervenience of mental content on wide functional role?

3.5 *The argument from public language meaning + introspection*

A different version of the argument is based on less substantive assumptions about spectrum inversion. Suppose that Invert and Nonvert are spectrum inverted relative to each other; forget for now about whether this is behaviorally detectable or not. The following seem like plausible claims:

- Invert and Nonvert each belong to the same linguistic community, and use that language to communicate seamlessly. Given this, it is reasonable to assume that, at least in cases where they habitually apply expressions to the same things, that they use their expressions to mean the same thing. In particular, it is reasonable to assume that they use their color words to denote the same properties.
- If there is a sentence S which a subject understands, then our reliability about the nature of our own experiences guarantees at least this: usually, if a (normal, attentive) subject says that the content of his experience is expressed by S , then that subject will be correct.

But the conjunction of these is inconsistent with the view that the contents of the experiences of Invert and Nonvert systematically differ. So, their conjunction is inconsistent with the defense of minimal intentionalism sketched in this section.