

# An argument for the impossibility of spectrum inversion without difference in representation

PHIL 93507

Jeff Speaks

September 22, 2009

1	Radical anti-intentionalism . . . . .	1
2	Moderate anti-intentionalism . . . . .	3
2.1	The idea that color phenomenologies represent color properties . . . . .	3
2.2	Indication and teleology . . . . .	3
2.3	Covariation . . . . .	4
2.4	Vertical vs. horizontal relations: conceptual role . . . . .	5
2.5	Mixed theories . . . . .	5
3	Quietist moderate anti-intentionalism . . . . .	6
3.1	Time-restricted intentionalism collapses into radical anti-intentionalism . . . . .	6
3.2	Time-unrestricted intrapersonal intentionalism collapses into interpersonal intentionalism . . . . .	7
4	Intentionalism and functionalism . . . . .	7

## 1 Radical anti-intentionalism

Suppose that spectrum inversion without misrepresentation is possible, and two experiences can be alike at the level of content while differing in phenomenal character. But then what is the connection between phenomenology and content? In particular, we can ask: is it a necessary truth that there be *some* systematic correlation, for any given subject, between phenomenology and content?

Suppose first that the proponent of spectrum inversion without misrepresentation says ‘No’. This yields the following *radical anti-intentionalist* view:

Content and phenomenal character are independent aspects of perceptual experience. It so happens that in our case there is a systematic (though contingent) connection between experiences with a certain phenomenal character (‘red-feeling’ experiences, for example) and experiences which represent objects as having certain color properties (redness, for example). But that is just an accidental feature of our constitution and relation to our environment; there could have been creatures for whom there was no systematic connection between content and phenomenology.

Radical anti-intentionalism is very implausible. For, if radical anti-intentionalism were true, then the following two scenarios would be possible:

Psychedelic phenomenology + constant content

A subject is looking intently at a well-lit surface which occupies the whole of the subject's visual field. Over the course of a few seconds, his experience goes from being (as we would put it) bright-red-feeling to being bright-green-feeling to being bright-blue feeling, and constantly repeats this pattern. But, the whole time, he is visually representing the wall as yellow; it visually seems to him throughout that the wall is yellow; according to his experience, the wall is yellow throughout.

Constant phenomenology + psychedelic content

A subject is looking intently at a well-lit surface which occupies the whole of the subject's visual field. The only thing notable about the phenomenology of his experience of the surface is its monotony. The experience is charcoal-gray-feeling, and remains so for the duration of the experience. Nonetheless, the subject is visually representing the color of the wall as constantly changing from bright red, to bright green, to bright blue; it visually seems to him that the wall is changing from bright red, to bright green, to bright blue; according to his experience, the wall is changing from bright red, to bright green, to bright blue.

Possible objections to the view that cases of these sorts are impossible: (i) color constancy ; (ii) these cases are possible if we understand 'seems' in the epistemic sense.

Response to (i): in most cases of color constancy, the color of *something* — like the light shown on a surface — is represented as changing. But we could just run the argument there. In other cases of color constancy, what changes is the relative lighting of a surface. But it is stipulated in the above case that there are no changes in lighting.

Radical anti-intentionalism is even a stronger thesis than the denial of intrapersonal time-restricted intentionalism local intramodal intentionalism (our weakest intentionalist thesis). Radical anti-intentionalism as applied to the case of color can be stated as follows:

Radical anti-intentionalism

For any color-phenomenologies  $P1, P2$  and any colors  $C1, C2$ , a single subject can have consecutive experiences which, respectively, have phenomenologies  $P1$  and  $P2$  and represent colors  $C1$  and  $C2$ .

But radical anti-intentionalism is pretty close to the denial of time-restricted intrapersonal intentionalism; once we see why radical intentionalism must be false, I think that we can see why any denial of time-restricted intrapersonal intentionalism would have to be false as well. The examples above would not get less impossible if, for example, the changes in color phenomenology or color content were not as dramatic. One exception: changes in color phenomenology which are not even in principle noticeable, if such there be.

Let 'moderate anti-intentionalism' be a name for the denial of interpersonal intentionalism conjoined with the denial of radical anti-intentionalism.

## 2 Moderate anti-intentionalism

### 2.1 *The idea that color phenomenologies represent color properties*

On the basis of the impossibility of the scenarios described above, the proponent of the possibility of spectrum inversion without misrepresentation should recoil from radical anti-intentionalism to *moderate anti-intentionalism*: she should say that, even though inversion examples show that there is no necessary connection between the phenomenal characters and contents of two arbitrarily chosen experiences, it is yet a necessary truth that there be, for any subject, *some* systematic connection between phenomenology and content.

There is nothing initially incoherent in the idea that phenomenology and content might be related in this way. The example of expressions and what they signify.

A problem with this analogy: the possibility of coreferential expressions. Suppose that in perception ‘coreference’ of this sort were possible; suppose, to fix ideas, that for a given agent green-feeling and purple-feeling experiences represent the same color property. Then the agent could have an experience which flipped rapidly back and forth between (as we would put it) a green-feeling phenomenology and a purple-feeling phenomenology, while nonetheless visually representing the color of the relevant surface as constant. But this is incoherent, in the same way that the above example of psychedelic phenomenology + constant content is incoherent.

We could run a similar argument using the possibility of ‘ambiguity’; in this case, the counterexample would be a version of constant phenomenology + psychedelic content.

The connection between phenomenology and content, the moderate anti-intentionalist should say, is like the relationship between expressions and their referents in some ways, but unlike it in others; and one of the ways in which it differs is that, as a matter of necessity, it is never the case that, for any one agent (or for any one agent in a suitably short interval), experiences of distinct phenomenal types represent a single color, or that experiences of a single phenomenal type can on different occasions represent distinct color properties.

One way to put this is by understanding the position of the moderate anti-intentionalist as an instance of the following claim:

For any subject, there is some relation  $R$  which is such that experiences with color-phenomenology  $F$  represent color-property  $G$  if the  $F$ -feeling experiences bear  $R$  to  $G$ .

The impossibility of psychedelic phenomenology + constant content and constant phenomenology + psychedelic content can then be thought of as putting a constraint on the relation  $R$ :  $R$  must be a one-to-one relation between experiences with certain types of phenomenal character and the color properties represented by those experiences.

### 2.2 *Indication and teleology*

This constraint rules out several otherwise plausible candidates for  $R$ :

- *Indication*. Suppose that a state of a certain kind indicates  $x$  for a subject iff were optimal conditions to obtain, the subject in question would come to be in that state only because of  $x$ . Clearly, there’s nothing to rule out two distinct states indicating the same thing.

- *Teleology.* The same goes for teleological theories, which might try to explain the connection between content and phenomenology for a creature in terms of facts about what evolutionary role was played by perceptual states with a certain phenomenology in the history of the creature’s species. There’s clearly no impossibility in states with two distinct color phenomenologies both playing the same evolutionary role in the history of a species. Since neither of these make  $R$  a one-to-one relation, neither are plausible candidates for explaining the representation relation between phenomenal character and color.

### 2.3 Covariation

But other candidates for  $R$  might seem more promising, such as the view that experiences with a given phenomenology represent a certain color property just in case they (under certain conditions) *co-vary* with that property. (See, for example, Tye (1995, 2000).) Since a subject is able to have at most one visual experience with color phenomenology at a time, it is plausible to think that it is impossible for experiences of distinct phenomenal types to each, for a single subject, co-vary with a single color property. For this reason, it seems that identifying  $R$  with some sort of covariation correctly rules out examples of psychedelic phenomenology and constant content, as well as the converse examples of psychedelic content and constant phenomenology.

However, covariational accounts are open to the same form of objection as the accounts considered above: they entail that a clearly impossible situation is possible. Consider the following scenario:

#### Full color phenomenology + no color content

A subject is (still) looking intently at a well-lit surface. The subject’s experience has the same phenomenology as your experience of something bright red. However, the subject’s perceptual experience does not represent the color of the surface at all; in fact, it does not represent anything as having any color. So, when the subject’s experience changes its phenomenal character to one similar to the phenomenal characters of the experiences you have when looking at bright green things, the subject’s experience does not represent the color of the wall as having changed.

Like the examples of psychedelic phenomenology + constant content and constant phenomenology + psychedelic content, the case of full color phenomenology + no color content is, it seems to me, clearly impossible.

But, if  $R$  were some sort of covariation, it would be possible. Suppose that when a creature’s visual system comes across a green object (under the right conditions), it sometimes triggers experiences with a green phenomenology, and sometimes experiences with a red phenomenology. Then neither type of experience — neither those with red phenomenology, nor those with green phenomenology — will covary with either color property. Since green surfaces will sometimes trigger experiences with red phenomenology, experiences with green phenomenology can’t covary with greenness; since green surfaces will sometimes trigger experiences with green phenomenology, experiences with red phenomenology won’t covary with greenness. So neither type of experience — neither those with red phenomenology nor those with green phenomenology — will covary with green, and neither will covary with instantiations of any other color. So, in this sort of case, the covariational theory yields the result that when the agent in question has experiences of this sort, she simply fails to represent the objects in question as having a color; it does not visually seem to the agent that the objects in question have a color. But this is extremely implausible.

If it does not seem immediately obvious that this is an absurd result, it may help to consider ‘mixed’ cases. We can imagine that experiences of the agent which have a yellow phenomenology

or an orange phenomenology, do satisfy the requirements of the covariational theory even though experiences with a red phenomenology, as above, fail to covary with any color property. But in this kind of case, if the covariational theory is to be believed, if the subject is looking at a screen with colors being projected upon it, and her experience switches from yellow phenomenology to orange phenomenology to red phenomenology, what has happened is that the screen first visually seemed yellow to the subject, then visually seemed orange to the subject, and then ceased to seem to have any color at all. This is hard to believe. Surely the switch from orange phenomenology to red phenomenology can't be a switch from representing the relevant surface as having a color to simply failing to do so.

#### 2.4 *Vertical vs. horizontal relations: conceptual role*

So far we've focused on attempts to explain  $R$  in terms of 'vertical' relations between color experiences and the external features which, under various conditions, cause the subject in question to have color experiences of the relevant type. The analogy here is with the distinction between broadly causal and conceptual role theories of the contents of mental states, where the former explain the contents of symbols in terms of symbol-world relations, and the latter explain the contents of symbols in terms of the relations between internal states. A 'conceptual role' approach to the relationship between phenomenology and content seems well-suited to avoid the problems we've been discussing, since a theory which explains content in terms of relations between phenomenal states needn't allow distinct phenomenal states to represent the same color property.

However, it's clear that the moderate anti-intentionalist can't make do with these sorts of relations alone. After all, if color content is explained wholly in terms of relations between certain phenomenal states, then color content should supervene on phenomenology. But this is just what the proponent of the possibility of spectrum inversion without misrepresentation must deny.

#### 2.5 *Mixed theories*

The natural move is then for the moderate anti-intentionalist to shift to a theory which explains  $R$  in terms of some combination of vertical and horizontal relations. (This is the analogue of the move from solipsistic to non-solipsistic conceptual role semantics, in the terminology of Harman (1987).) The idea would then be that the horizontal relations could secure distinctness of content between distinct phenomenologies for an individual at a time, while the vertical relations could secure variance of content between individuals (or between distinct times in the life of a single individual).

But the problem is that it's hard to see how these two sorts of relations could be combined without yielding the problems faced by covariational accounts. Horizontal relations between experiences of certain phenomenal types are introduced to avoid the problem of experiences of distinct phenomenal types representing the same color property; we can therefore think of the use of these horizontal relations as imposing the requirement that distinct phenomenal types represent distinct colors. But now suppose that, for a subject, experiences of distinct phenomenal types  $P_1$  and  $P_2$  bear the relevant vertical relations to a single color property  $C$ . By parity, experiences of type  $P_1$  represent  $C$  if and only if experiences of type  $P_2$  do. But if both do, then scenarios of psychedelic phenomenology + constant content are possible; the subject's experience could switch rapidly between distinct phenomenal characters  $P_1$  and  $P_2$  while representing the color of the relevant surface as constant. But cases of this sort are impossible, so both can't represent  $C$ ; so neither can. But presumably neither represents any color other than  $C$ . So experiences with phenomenal characters  $P_1$  and  $P_2$  do not, for this subject, represent objects as having any colors

at all. But this is just a case of full color phenomenology + no color content, and, as argued above, cases of this sort are not possible. So combinations of vertical and horizontal relations fare no better than either taken singly.

...

The cases discussed above can be thought of as constituting a kind of dilemma for the moderate anti-intentionalist: (i) if a candidate for  $R$  is tolerant enough to rule out cases of full color phenomenology + no color content, then  $R$  will fail to be one-to-one, and so will make possible either or both of the scenarios of psychedelic phenomenology + constant content or psychedelic content + constant phenomenology; but (ii) if a candidate for  $R$  is made demanding enough to block the possibility of these cases, it will end up making color representation implausibly difficult, making cases of full color phenomenology + no color content possible.

### 3 Quietist moderate anti-intentionalism

Here, though, the anti-intentionalist may wish to raise a question about the dialectical situation. Let us suppose that we have failed to specify a plausible candidate for the relation  $R$ . Why should the moderate anti-intentionalist feel pressed to specify this relation? The moderate anti-intentionalist will want to distinguish between the following three theses:

#### Intrapersonal Time-Restricted Intentionalism

Necessarily, if two experiences of a single subject within some minimal time interval  $t$  differ in phenomenal character, then they differ in content.

#### Intrapersonal Time-Unrestricted Intentionalism

Necessarily, if two experiences of a single subject (whatever the interval between them) differ in phenomenal character, then they differ in content.

#### Interpersonal Intentionalism

Necessarily, if two experiences (whether of a single subject or two subjects) differ in phenomenal character, then they differ in content.

The radical anti-intentionalist denies each of these; but the moderate anti-intentionalist is as such committed only to denying the third. The foregoing argument against moderate anti-intentionalism amounts, basically, to a demand that the moderate anti-intentionalist specify that relation between phenomenology and content which will make Interpersonal Intentionalism false, while letting at least one of the two versions of Intrapersonal Intentionalism be true. But perhaps the moderate anti-intentionalist can resist this demand for explanation. The two versions of Intrapersonal Intentionalism are just supervenience theses, as is Interpersonal Intentionalism. Why can't the moderate anti-intentionalist simply accept the first or first and second of these theses, deny the third, and be done with it?

#### *3.1 Time-restricted intentionalism collapses into radical anti-intentionalism*

First, suppose that the quietist moderate anti-intentionalist endorses only Intrapersonal Time-Restricted Intentionalism, and rejects Intrapersonal Time-Unrestricted Intentionalism. In this case, the if two consecutive experiences of a subject have the same color content, they must

have the same color phenomenology; but if the two experiences are separated by some minimal interval of time — call this  $t$  — they might have the same content, and different phenomenology. But presumably it is possible for the subject to be having perceptual experiences during  $t$ , which have some color phenomenology. Since, by hypothesis,  $t$  is the minimal interval of time by which two experiences alike in color content but distinct in color phenomenology must be separated, they cannot have the color content appropriate to the perceiver's situation prior to  $t$ ; but by the same reason they can't have the color content appropriate to the perceiver's situation subsequent to  $t$ . And they can't have some third sort of content since, as in the previous cases, this would violate the stipulation that  $t$  is the minimal interval of time by which two experiences alike in color content but distinct in color phenomenology must be separated. So they must have no color content. But this would, contra our conclusions above, make cases of full color phenomenology + no color content possible. To avoid this this interval of contentless perceptual experience, it must be the case that  $t=0$ . But this is to collapse our first version of quietist moderate anti-intentionalism to radical anti-intentionalism; if  $t=0$ , then psychedelic phenomenology + constant content would be possible. But it isn't. So the anti-intentionalist should not reject Intrapersonal Time-Unrestricted Intentionalism unless she wants also to reject Intrapersonal Time-Restricted Intentionalism, and so become a radical anti-intentionalist.

### 3.2 *Time-unrestricted intrapersonal intentionalism collapses into interpersonal intentionalism*

So suppose that the quietist moderate anti-intentionalist endorses Intrapersonal Time-Unrestricted Intentionalism but rejects (as he must, to qualify as an anti-intentionalist at all) Interpersonal Intentionalism. Let  $A$  and  $B$  be distinct individuals whose experiences, by the anti-intentionalist's lights, differ in phenomenology but have the same content. Whatever the facts about these individuals are which determine the phenomenologies and contents of their mental states, can't we always imagine a sufficiently long-lived individual who, at one stage in his life, is identical in all relevant respects to  $A$ , and at another stage, is identical in all relevant respects to  $B$ ? If so, then it is hard to endorse Intrapersonal Time-Unrestricted Intentionalism without endorsing Interpersonal Intentionalism.

Caveat: this is blocked by teleological theories. But these theories are implausible for the reasons sketched above, plus swampman cases.

Just as Intrapersonal Time-Restricted Intentionalism (plus the denial of the time-unrestricted version of the thesis) collapses into radical anti-intentionalism, Intrapersonal Time-Unrestricted Intentionalism collapses into Interpersonal Intentionalism. Moderate anti-intentionalism is thus an inherently unstable position; the two stable positions on this topic are radical anti-intentionalism and intentionalism. It is hard to be a moderate anti-intentionalist — even a quietist one.

A restriction on the argument: the difference between inversion without misrepresentation and inversion without difference in representation. Strictly, the above argument aims to rule out only the latter, which is the one relevant to intentionalism.

## 4 **Intentionalism and functionalism**

Intentionalists have often been attracted to the view because it promises a reduction of phenomenal character to mental representation, which can then be explained via some sort of functionalist theory. The argument above may be an aid to intentionalism, but it is no help to this sort of intentionalist.

That is because the discussion above shows that functionalism + intentionalism is open to the

same sorts of arguments as is the anti-intentionalist. In a nutshell, the arguments above seem to show that if one or another functionalist view of perceptual representation were correct, then one of the impossible scenarios described above would be possible. So these functionalist accounts of perceptual representation must be false.

And it's hard to see how this result could be limited to *perceptual* representation. Suppose that you and I are spectrum inverted relative to each other; then we will differ with respect to the color our visual experiences attribute to a marigold. Suppose that we both take our experiences at face value, forming beliefs which attribute to the marigold the same color property our experience represents the marigold as having. Then it seems that our beliefs about the color of the marigold will differ in content as well.

Another way to see the problem for functionalism here is by noting that, if intentionalism is true, then this provides us with a general tool for arguing against various theories of mental representation. Suppose that a theory explains the content of mental representations in terms of the *A*-properties. Then we can ask: is spectrum inversion possible between subjects alike in their *A*-properties? If so, then mental representation fails to supervene on the *A*-properties, and the theory of mental representation in question is false.

The more specific the theory of mental representation, the more powerful this line of argument. It is notoriously difficult to decide whether spectrum inversion between full functional duplicates is genuinely possible. But it might be easier — to use Jerry Fodor's well-known and admirably specific theory of mental representation as an example — to decide whether spectrum inversion is possible between subjects who are alike with respect to, for example, the dependence relations among nomological connections between representations in their language of thought and properties in the world. (See Fodor (1990).) If this is possible, then, if the preceding argument is sound, Fodor's theory of mental representation must be false.

This puts some pressure on the view, which seems fairly popular, that naturalizing phenomenal character is impossible, whereas there must be some naturalistic account of mental representation. Kim (2007) and Chalmers (1996) are two people who seem to come close to endorsing this view.

## References

- David Chalmers, 1996. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.
- Jerry Fodor, 1990. A Theory of Content, I: The Problem. In *A Theory of Content and Other Essays*, 51–88. Cambridge, MA: MIT Press.
- Gilbert Harman, 1987. (Nonsolipsistic) Conceptual Role Semantics. In his *Reasoning, Meaning, and Mind*, 206–232. Oxford: Clarendon Press.
- Jaegwon Kim, 2007. *Physicalism, or Something Near Enough*. Princeton, NJ: Princeton University Press.
- Michael Tye, 1995. *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. Cambridge, MA: MIT Press.
- Michael Tye, 2000. *Consciousness, Color, and Content*. Cambridge, MA: MIT Press.