Functionalist theories of content

PHIL 93507
April 22, 2012

Let's assume that there is a certain stable dependence relation between the physical internal states of subjects and the phenomenal characters of their experiences. I'll refer to the internal state which is sufficient for a subject S to have an experience with a given phenomenal character — say, RED — as RED$_S$.

This will be, for all I have said, a non-rigid designator of internal states. It is also worth emphasizing that the assumption that there are states of this sort does not involve the assumption that phenomenal properties supervene on the intrinsic properties of subjects, even if a reasonable case can be made for this claim. All that is required is the very plausible assumption that there are causally sufficient conditions for a subject to be in a certain phenomenal state which are specifiable in terms

We can think of a functionalist theory of perceptual content as specifying a relation R between internal states-types of this sort and the properties those internal states represent (for a given subject) as instantiated. Then the impossibility of Scenario A places a constraint on functionalist theories of content, which can be stated as follows:

> [C] Necessarily, there is no subject S such that RED$_S$ and GREEN$_S$ both bear R to redness at the same time.

Suppose for *reductio* that [C] is false. Then take the subject whose possibility is guaranteed by [C]'s falsity, and imagine that subject being first in state RED$_S$, and then, immediately after, being in GREEN$_S$. This would be a case of psychedelic phenomenology + constant content; but such cases are impossible.

Intuitively, to show that a functionalist theory fails to meet [C], one has to do two things: (1) show that distinct mental representations can bear R to just the same property, and (2) show that such co-reference is possible when tokenings of those mental representations are associated with distinct phenomenal states.

A possible candidate for a theory which fails to meet [C]: indication theories, which say that the content of a perceptual state is determined by what would cause the subject to be in that state, were the subject in certain ideal conditions.

On (1): there seems to be no impossibility in the claim that given state of affairs could, in conditions optimal for the subject, sometimes cause one internal state and sometimes another.

(2) is harder. If we avoid relying on controversial internalist theses about the supervenience of intrinsic qualities on phenomenal character, we have to rely on a modal claims like the following:

> [M] Possibly, there is a subject S for whom RED$_S$ and GREEN$_S$ both indicate the color red.

If [M] is true, an indication theory entails that the subject whose possibility it guarantees would, when alternating between the two states RED$_S$ and GREEN$_S$, be an instance of psychedelic phenomenology + constant representation of color properties. [M] is not obvious. But a case can be made for it.

> For a pair of states x and y to indicate the same property F (or state of affairs) for a subject is for it to be the case that, in optimal conditions, the subject would be caused to be in x only by instances of F, and the same for y. Optimal conditions are, intuitively, conditions under which the cognitive system of the subject is functioning perfectly. So the possibility of (a) turns on this question: is it possible that a subject could be such that conditions optimal for it are conditions in which instances of F sometimes cause x and sometimes y, and nothing else ever causes a state of either type?
>
> But this does seem to be possible. Imagine first that our subject is an evolved creature, and that we think of optimal conditions for such creatures as "involving the various components of the visual system operating as they were designed to do in the sort of external environment in which they were designed to operate." (Tye) It is well-known that the process of evolution often issues in creatures who are not "put together" in the way one might expect a creature designed *ex nihilo* for the relevant environment to be put together. This sort of observation makes it very plausible that, even if it seems odd a priori, it is possible for a creature to have evolved to operate in an environment in which both RED$_S$ and GREEN$_S$ were caused by red surfaces. Perhaps there was a surprising absence of green things in the environment, and red things were very important to the creature's survival — so important that it was evolutionarily useful to have redundancy in one's red-indicating capacities.
>
> One might get around this problem by giving another, non-evolutionary view of optimal conditions. But it is hard for views of this sort to steer a middle course between circularity and an implausible indeterminacy of content. On the one hand, we can't define optimal conditions as, e.g., "conditions under which all of the creatures beliefs are true," since that builds facts about the truth conditions of beliefs into the story about what fixes the contents of the subject's mental representations. One the other hand, one can't fall back on a view of optimal conditions according to which they are something like "conditions under which the creature would flourish", since it seems plausible that for at least some

> creatures, they would flourish as well in a scenario in which RED$_S$ was caused by greenness as (with appropriate changes to the causes of other phenomenal states) in a scenario in which it is caused by redness. But it is implausible to think that, for this reason, that creature's perceptual states with phenomenal character RED represent surfaces as indeterminately red-or-green.
>
> Here's a somewhat fanciful way to get a case which illustrates the problems posed by non-evolved creatures. Familiar spectrum inversion scenarios feature inverting lenses which would give our ordinary experiences of yellow things a BLUE phenomenal character and our ordinary experiences of blue things a YELLOW phenomenal character. Now imagine a newer model of type-2 lenses, which gives our ordinary experiences of blue things a GREY phenomenal character and our ordinary experiences of yellow things a BLUE phenomenal character (and leaves the phenomenal character normal for experiences of grey things unchanged). While wearing the type-2 lenses, therefore, one never has an experience with YELLOW phenomenal character. Type-3 lenses fix this problem. Type-3 lenses are just like the type-2 lenses, but for the fact that when confronted with the light reflected from a normal yellow surface, they sometimes do what the type-2 lenses do, and cause a BLUE experience, but other times just "let the light through", causing a YELLOW experience. Let's suppose that it is random when type-3 lenses do the one, and when they do the other.
>
> Now imagine a Swamp-duplicate of me who materializes on earth with type-3 lenses as part of its visual system. It's hard to see how conditions optimal for Swamp-me could be anything other than the conditions in which it materialized; so it looks like (presuming that the phenomenal character of Swamp-me's states is the same as the phenomenal character of my states when I'm wearing the lenses), both BLUE$_{swamp\text{-}me}$ and YELLOW$_{swamp\text{-}me}$ will indicate the color yellow. Hence, if the indication theory sketched above is true, both of these states will represent the color yellow; which will make cases of psychedelic phenomenology + constant content possible.

Intuitively, what we want to do is modify our functionalist theory so that it does not permit this sort of co-reference. There are a few different ways in which we might do this.

Response 1: Brute force

Even if the functionalist theories listed above don't meet [C], it might seem pretty easy to modify them so that they do: just add to the theory the stipulation that distinct representations of color properties can't represent the same property. But this does not really help.

Consider again the possible situation (a) in which both RED$_S$ and GREEN$_S$ indicate a single color — say, redness. This would then be a case, if if we add the suggested stipulation to the indication theory, in which neither RED$_S$ nor GREEN$_S$ would covary with a color in optimal conditions, and hence would be a case in which neither represents a color. But now imagine, what is surely possible, that some other states, which we can call ORANGE$_S$ and YELLOW$_S$, do covary with colors in optimal conditions — with orange and yellow, respectively. 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 represent it as colored at all.

This is also a problem for a covariational theory.

Response 2: Internalist functionalism

Let's consider another way in which the functionalist might respond to this problem: she might endorse a purely internalist functionalist theory. If a functionalist theory has the consequence that any intrinsic duplicates are alike with respect to the representational properties of their internal states, then such a theory will not, obviously, permit the permutation of the representational properties of visual states via the permutation of the relationship between those states and instantiations of the various colors in the world, and hence won't be open to an argument like the ones just given.

Problem: hard to see how to construct such a theory.

Response 3: Limited holism

We could instead define R as a relation between the subject's collection of phenomenal states and the color properties; maybe a phenomenal state represents a color iff the "best pairing" of all the phenomenal states with the colors assigns that phenomenal state to that color. (We might define "best pairing" in various ways — maybe in terms of normal causes.)

However, this theory may not be able to avoid entailing the possibility of psychedelic phenomenology + constant representation of color properties after all. However, exactly, we determine which pairing of color properties with phenomenal states does the best job of matching state types with their normal causes, it appears that this pairing is something that can change over time. At least the most obvious definitions of "normal cause" — in terms of frequency, or causation in certain ideal conditions — will have the consequence that the normal causes of a state can change over time, which will be sufficient to change the best pairing of states and represented colors.

If so, a holistic functionalist theory is open to a diachronic version of the argument above:

> Let t be the time in some subject S's life when the "best pairing" changes. Then there will be some pair of internal states with associated phenomenal characters — say RED_S and GREEN_S — which are such that the color property which experiences with RED_S represent before t = the color property which experiences with GREEN_S represent after t. Now suppose that the individual in question is in state RED_S just before t, and then, at t, comes to be in state GREEN_S. This will then be a case of psychedelic phenomenology + constant representation of color properties. Varying the case so that the subject is in RED_S before and immediately after t would give us a case of constant phenomenology + psychedelic representation of color properties.

Objection 1: Perhaps we should say that there is no "one moment" when the best pairing switches; perhaps the best pairing can only shift over a sufficiently long interval of time t.

> *Reply*: This runs into all the problems, discussed above, with the time constraint.

Objection 2: maybe the "normal cause" of a state does vary over time; but nothing says that we have to define the pairing between the phenomenal states and the colors in terms of normal cause. And the argument just given seems to show that we shouldn't, and that we should instead let the pairing be determined by some fact about the subject which, in principle, can't change over the course of the subject's life — like, for example, the subject's evolutionary history.

> *Reply 1*: One consequence of this sort of view is that it seems to make interpersonal spectrum inversion (and spectrum shift) without misrepresentation possible, but intrapersonal spectrum inversion (and spectrum shift) without misrepresentation impossible. As noted above in connection with the interpersonal constraint, this can seem ad hoc, since the intuitions which militate in favor of the possibility of interpersonal inversion are just as strong in the intrapersonal case. A good example here is the sort of spectrum shift discussed in [1]. We're just as disinclined, intuitively, to assert differences in the veridicality of color experiences between the experiences of people of different ages (intrapersonal shift) as between the experiences of people of different sexes and races (interpersonal shifts).
>
> But this might be taken just to be an interesting consequence of, rather than an objection to, the view.
>
> *Reply 2*: A holist teleological theory of this sort can't be the last word, since it is extremely plausible that Swampman could visually represent objects in his environment as having colors. So to make this response general, we also need

some account of what could fix the correlation between Swampman's phenomenal states and the colors which is such that the "best pairing" can't change over the course of Swampman's life. One might appeal to some more general understanding of "optimal conditions" to do this, but this gets into the problems discussed above in connection with modal claim (a) — it is hard to give such a view of optimal conditions without either lapsing into circularity or giving rise to an implausible indeterminacy of content.

If this sort of worry can be overcome, then it seems to me that giving some theory of this sort — a holist functionalist theory which fixes the pairing between phenomenal states and represented properties using some property which is in principle unchangeable over the course of the subject's life — may be the best bet for the externalist functionalist about perceptual content.