Lessons for Mary

Pär Sundström

New York University and Umeå University

par.sundstrom@philos.umu.se

Abstract. What could you learn if you saw a colour after being confined from birth to a black-and-white room? It turns out that this is surprisingly hard to say. I suggest that reflection on this question teaches us that colour perception has a richer content than we might have thought.

1. Introduction

Frank Jackson (1982) once told us the story of Mary, who is confined from birth to a black-and-white room but one day gets to see red. Jackson's interest in this story concerned the truth or falsity of physicalism. But as Michael Thau (2002) and Alex Byrne (2002) have recently pointed out, the story raises a hard and interesting question that is independent of the issue of physicalism, namely: what could Mary *learn* when she sees red? This is the question I shall discuss.

Consider the following version of Mary's release. While she has still experienced no chromatic colour, Mary is told that her thumbnail, which is behind her back, is painted red. Although Mary is told that the nail is red, it seems that she will learn something when she sees it.

But what could she learn? One might think that,

        (BELIEVE-RED) when Mary is told that her nail is red, she comes to *believe* that it is red,

and

        (SEE-RED) when she gets to see it, she *sees* that it is red.

But then Mary *sees* what she already *believes* to be the case. And if so, what does she *learn*?

There are three main strategies for answering this question. The first is to deny (BELIEVE-RED). Perhaps Mary doesn't really believe that her nail is red before she sees it. If so, it is obvious what she learns: she learns that her nail *is red*. The second strategy is to deny (SEE-RED). Perhaps what Mary sees is not that her nail is red, but that it is some way other than red. If so, what she learns is that her nail is *that way*. The third strategy is to accept both (BELIEVE-RED) and (SEE-RED) and identify something that Mary can learn even though they are both true.

In sections 2 and 3, I briefly discuss the first and the second strategy. In section 4, I present my own proposal, which is a version of the third strategy.

2. First strategy: Deny (BELIEVE-RED)

Perhaps one needs to have had experiences of red in order to believe of something that it is red. If so, Mary won't believe that her nail is red even if she is told so. She may believe other things, for example, that her nail has some property called 'red'. But she won't believe that it is red.

Some philosophers think this is correct (e.g. Harman 1990), but in my view there are considerable reasons to doubt that.

As philosophers like Burge (1979) and Kripke (1979) have emphasised, we quite often have beliefs about objects and properties that we have had no acquaintance with and know quite little about. In fact, we can have beliefs about objects and properties even if we are significantly *mistaken* about them. For example, I could have beliefs about arthritis even if I had no acquintance with this disease, knew very little about it, and were importantly mistaken about what it is.

Further, it seems -- to me -- that beliefs about *colour* are just like beliefs about arthritis in this regard. To illustrate this, it may be helpful to first consider some "propositional attitudes" other than belief. Thus, it seems plausible that Mary could *hope* that she will one day see red, and *doubt* that this will ever happen. And if Mary can *hope* and *doubt* that she will see red, surely she can *believe* it too.

For these briefly indicated reasons, I think the first strategy is not promising.

Before I turn to the second strategy, this is the place to make one important observation about our current puzzle. Because of Jackson's interest in physicalism, he supposed that Mary knew "all" the physical facts about colour and colour vision. I have not assumed this, and I can now make clear why The present problem is hard insofar as it is plausible that Mary can *believe* that someting is red before she has seen red. And if this section is right, Mary can believe this even if she is not an expert on colour. She may have to know a few things about red to have beliefs about it. Perhaps she needs to know that red is a visually detectable surface property. But she needn't an expert on red to have beliefs about it any more than I need to be an expert on arthritis to have beliefs about it.

I shall eventually consider the case where Mary in some sense knows "all" the facts about colour and colour perception. But for the time being, I will take Mary to be only as educated about these things as the average person.

3. Second strategy: Deny (SEE-RED)

Thau (2002, chapter 5) and Byrne (2002) both argue that, since Mary comes believe that her nail is red when she is told so, and learns something when she sees it, we must conclude that she doesn't see that her nail is red. Thau embraces this conclusion "at the level of reference". According to him, the *property* that Mary sees and comes to believe that her nail has is, contrary to what we naively think, not the property red. Byrne takes the conclusion to hold only "at the level of sense". On his view, Mary sees and comes to believe a content which *refers* to the same property that her antecedent belief was about, namely redness, but which represents this property under a new *concept*.

If Mary doesn't see that her nail is red, what does she see? Thau and Byrne both argue that what Mary sees and learns cannot, in principle, be expressed by any public language predicate. The argument is this. Suppose we could introduce a public language predicate -- say, 'R' -- with which to express what Mary learns when she comes to see her nail. Then we could tell Mary, and thus make her believe, that her nail is R before she sees it. But at this point, our problem would repeat itself: if Mary already believes that her nail is R, how can she learn something -- as she surely will -- when she sees it to be R? The only way to prevent the problem from arising again is to assume that we could not introduce a predicate to express what Mary learns.

I don't think we are forced to accept this conclusion. I think it is possible that Mary learns something even if she sees what she already believes to be true. Moreover, I think that what she learns can be expressed in public -- and, indeed, familiar -- terms. The reason it's hard or impossible to tell her what she will learn is not that she learns something inexpressible, but that she learns *so much*.

4. Third strategy: Accept both (BELIEVE-RED) and (SEE-RED)

I shall exploit an analogy: Consider a complex, irregular, bounded, two-dimensional shape, composed of large number of straight lines. Call this shape-type 'VERYCOMPLEX'. Now, suppose I'm told that I will see a VERYCOMPLEX figure. For the reasons mentioned in section 2 above, it seems possible that, even if I have never seen a VERYCOMPLEX and know quite little about this shape, I can believe that the figure I will see is VERYCOMPLEX. But even if I do, it is clear that I can learn any number of things when I see the figure (as VERYCOMPLEX). For example, I can learn that the longest line segment of a VERYCOMPLEX is at least three times as long as the second longest one, and that there is at least one angle between line segments that is greater than 90°.

There are two features of VERYCOMPLEXNESS that explains why I can learn something from seeing this shape even if I already believe that I will see it. The first is that VERYCOMPLEXNESS can, in a sense, be analysed in terms of its "parts" or "aspects". Thus, having a longest line segment that is more than three times as long as the second longest segment is *part* of what it is to be VERYCOMPLEX, as is having at least one angle between line segments that is greater than 90°. Second, these "parts" are *visually accessible* in this sense: one can come to know on the basis of seeing a VERYCOMPLEX that it has a longest line segment that is more than three times as long as the second longest segment. Since the shape has different "parts", I can know that a figure is VERYCOMPLEX and yet not know everything about it -- not even everything that is entailed by its being a VERYCOMPLEX. And since those parts are visually accessible, I can learn that a VERYCOMPLEX has them when I see one.

I suggest that this provides us with a useful model for what Mary learns when she sees her nail. This suggestion may appear outrageous, for the following reason. Unlike VERYCOMPLEX, red does not have visually accessible parts. True, philosophers of colour disagree about whether red is a simple, unanalysable quality or not. But whatever it *is*, red is *given to perception* as a simple and unanalysable quality. Or so we tend to think.

I suggest, however, that red -- that apparently simple and unanalysable quality -- has, in my sense, visually accessible parts or aspects: it is possible to learn any number of things about how red is -- or about how red things are -- on the basis of seeing red.[1]

Here is a sample of the truths that, if I'm right, Mary can learn when she sees her nail (as red), even if she already believes that the nail is red:

> (a) In normal daylight, it is easy to distinguish red from black, white and grey.
> (b) As daylight fades, red becomes indistinguishable from black and grey before it becomes indistinguishable from white.
> (c) Red can (unlike black) be the colour of a light.
> (d) If it's dark and clear, it is easy to detect both a white light and a red light,
> (e) but if it's foggy, a red light is, all else equal, easier to detect than a white light.

Unless Mary has studied vision science, she won't know very many such truths prior to seeing her nail. But can she really learn these truths just on the basis of seeing her red nail in normal daylight? Well, to be honest, I'm not quite sure, and I don't know how to find out. But it strikes me as not too implausible that she could. It doesn't seem wildly implausible, for example, that Mary's perception of her nail should give her *some* clue about whether the nail would be easier to distinguish against a black background or a white background as daylight fades. And if it gives her *some* clue, she can come to believe it to *some* degree. And since what she then to some degree believes is *true*, this would constitute a *learning* of sorts.

The proposal is obviously somewhat speculative. All I claim for it, for now, is that it should be recognised as a live option along with other responses to our puzzle.

---

[1] This idea and its elaboration below is remotely inspired by Broackes (1992) and Noë (forthcoming, chapter 4), but I don't know if either of them would accept what I say.

But even if you are sympathetic thus far, you may wonder about the case where Mary is an expert on colour and colour perception? Suppose Mary has been told about (a)-(e) and "all" other truths about red and how it looks. What could she learn in that case?

Consider VERYCOMPLEXNESS again. Suppose I haven't seen this shape, but that you tell me the relative lengths of all its line segments as well as the sizes of all the angles between them. And suppose that, by using some mnemonic technique, I can remember this whole description. Then I know, in a sense, all the facts about VERYCOMPLEXNESS; at any rate, my knowledge of VERYCOMPLEXNESS is as complete as expert Mary's knowledge of red could conceivably be. But suppose, further, that I am unable to "keep in mind" more than a few facts about VERYCOMPLEXNESS at any one time. Let's say that, at any one time, I can keep in mind at most the relative lengths of five contiguous line segments and the size of the angles between them. Then it is again clear that something new and significant happens if I get to see a VERYCOMPLEX.

Part of what is new in this case is that I *do* something that I haven't done before. Whereas before I could, at any given time, access only a few of all the facts I knew about VERYCOMPLEXNESS, I now access them (visually) *all at once*. But by *doing* this, it seems I can also come to *know* things I didn't know before. For example, I may come to know long and complicated conjunctions that were entailed by what I previously knew (e.g., that the figure has a longest line segment at least three times as long as the second longest one, *and* that there are ten contiguous line segments that are approximately the same length, *and*...-- so forth.). While such propositions were entailed by what I previously knew, it was in practise beyond me to entertain them, because of their complexity. But with the aid of my new, representationally rich visual state, I can entertain them and know them to be true.

I suggest that something similar could be true of expert Mary. Even if she in a sense knew all the facts about red and how red looks, she would supposedly be incapable of keeping very many of them in mind at any one time. But when she sees red, she represents many of these things all at once, if I am right. And with the aid of this visual

state, she can learn things that she was unable to figure out before. For example, she can learn long conjunctions of the things she knew, as well as other consequences of her antecedent body of knowledge.

Literature

Broackes, Justin 1992 "The autonomy of colour", in: *Reduction, Explanation, and Realism*, David Charles and Kathleen Lennon (eds.), Oxford: Oxford University Press, 421-65.

Burge, Tyler 1979 "Individualism and the mental", in: Peter French, Theodore Uehling, and Howard Wettstein (eds.), *Midwest Studies in Philosophy* 4, Minneapolis: University of Minnesota Press, 73-121.

Byrne, Alex 2002 "Something about Mary", *Grazer Philosophische Studien* 63, 123-40.

Harman, Gilbert 1990 "The intrinsic quality of experience", in: James Tomberlin (ed.) *Philosophical Perspectives* 4, Atascadero, California: Ridgeview, 31-52.

Kripke, Saul 1979 "A puzzle about belief", in: Avishai Margalit (ed.), *Meaning and Use*, Dordrecht: Reidel, 239-83.

Jackson, Frank 1982 "Epiphenomenal qualia", *Philosophical Quarterly* 32, 127-36.

Noë, Alva, forthcoming, *Perception in Action*. Draft available online at <http://socrates.berkeley.edu/~noe/action.html>.

Thau, Michael 2002 *Consciousness and Cognition*, Oxford: Oxford University Press.