Philosophica 68 (2001, 2) pp. 31-40

CONSTRAINED INVERSIONS OF SENSATIONS[1]

Erik Myin

ABSTRACT

Inverted sensation arguments such as the inverted spectrum thought experiment are often criticized for relying on an unconstrained notion of 'qualia'. In reply to this criticism, 'qualia-free' arguments for inversion have been proposed, in which only physical changes happen: inversions in the world, such as the replacement of surface colors by their complements, and a rewiring of peripheral input cables to more central areas in the nervous system. I show why such constrained inversion arguments won't work. The first problem is that the world lacks the symmetry required to invert physical properties in the way required. The second problem concerns 'rewiring'. Empirical evidence indicates that the rewirings are either impossible, or would not result in an inversion of sensation. I propose the deeper reason for the failure of constrained inversion arguments lies in the fact that sensations are not properties of brain states, but spread into the world and the body.

1. Inversions without qualia

Inverted sensation arguments remain popular in the philosophy of mind. The most famous one, no doubt, is the inverted spectrum thought experiment, in which it is claimed that two persons' color sensations could be inverted with respect to each other without any behavioral sign of it (for the original version, see Locke (1690/1987)).

---

[1] I would like to thank Robert Kirk, who quite strongly disagrees, for comments upon this manuscript, as well as the audience at the Fourth Annual Conference of the Association for the Scientific Study of Consciousness, Université Libre de Bruxelles, 2000. For a treatment of related issues in a different context, see Myin (2001). For financial support, I wish to acknowledge the FWO-Vlaanderen (project G.0175.01) and the Vrije Universiteit Brussel (project OZR 396).

Inverted sensation arguments mostly have the notion of a *quale* as an essential ingredient, for it is claimed that it are *qualia*, the phenomenal or qualitative aspects of sensations that are inverted. However, these arguments have been criticized precisely because the notion of quale appealed to is unconstrained (in particular, see Dennett (1990)). The criticism, roughly sketched here, is that the notion of a quale is such that one can imagine anything about it. For example, Dennett has argued that if qualia taken to be 'atomistic', 'intrinsic', 'nonrelational' and 'ineffable', then, they are something with which you can *of course* do anything you want in any thought experiment (see also Myin (1999), for some related points).

As a reaction to such criticism, some people have come up with what will be called *Constrained Inversion Arguments*: arguments where you get inverted sensations -but without invoking qualia. As far as this author knows, the most detailed version of such a Constrained Inversion Argument is due to Robert Kirk, who presents it in his *Raw Feeling* (Oxford University Press (1994)).

Kirk's argument involves two stages. First it is shown that inverted sensations are possible as a consequence of an uncontroversial operation in the world. This shows that having inverted sensations is unproblematic. What it is to have inverted color sensations is simply like what it is to be in a world in which surface colors are inverted: a world in which red surfaces are replaced by green surfaces, yellow surfaces by blues, etc... This operation could be done -just imagine a huge team of painters and lots of paint. This first step, then, is aimed at proving that the idea of inverted sensations can be made crystal clear just by invoking an unproblematic, if unpractical, physical operation. The second ingredient, necessary to show that sensations can be inverted in two persons who are in the *same* physical situation, is the idea of a *rewiring*. Just imagine two parallel worlds, a repainted one besides an original, and you and your molecule for molecule duplicate. Now further imagine your peripheral nerve cables would be rewired so as to arrive at the central neural stages which are activated in your twin when she's looking at the parallel scene in her world. Clearly, so the defender of the Constrained Inversion Argument presses, this is another, unproblematic, physically possible, operation, that would certainly lead to inverted sensations.

In the remaining of this paper, an attempt will be made to cast doubt on the two steps of this constrained inversion argument. Nothing that will

be said is meant as a direct criticism of qualia or inverted sensations arguments that involve qualia. So, importantly, this paper has no pretension of offering an empirical refutation of spectrum inversion. Yet it is meant to offer an attack on empirical construals of spectrum inversion -construals which go beyond the claim that spectrum inversion is only conceptually possible.

2. Problems for lightness inversions without qualia

To address constrained inversion of sensation arguments, I will turn to the usual example of spectrum inversion. Instead of starting with hue, however, I will start with lightness inversion: the inversion of surface blacks and whites and all the gray values in between (The idea is considered, for example in Wittgenstein(1977), Shoemaker(1975) and Kirk(1994)). Only later on, I will turn to fully chromatic, or hue inversion. The rationale behind this somewhat indirect approach is to indicate first how deeply problematic constrained inversion is in the case of lightness, to make a case then that the difference between surface lightness and hue is not as great as it seems, in other words, to argue that the treatment for lightness can perhaps be applied to hue also.

2.1. Physical problems

The question addressed in this section is whether you can really imagine an inverted environment, in which surface lightnesses would be inverted and in which my twin would have inverted sensations.

If only the environment were two-dimensional there would be little problems for that. Think only of a checkerboard pattern where the blacks and whites can be switched, or think of a photographic negative. However, in a three-dimensional environment, things become different.

The problem is that the being black or white of a surface has effects that 'spread into three dimensional space' Being black or white is not only a local property of surfaces, it is a property that determines how light is distributed over the whole scene. A black surface diminishes the amount of light available, a white surface keeps constant the light available in a scene; it just spatially redistributes it. This means that the presence of black and white surfaces has consequences for the overall

distribution of light in a scene, and thus for overall visibility.

The difference shows up, for example, in what is visible in a white walled versus a black walled room, illuminated by the same light source. An object lying in a distant corner will be visible in the white walled room, but not in the black walled room.

What this shows is that black and white, as surface colors, lack a certain kind of symmetry that would be necessary to allow the first phase in a constrained inversion argument to go through. To be in a black walled room is *not* having sensations that are the symmetrical 'mirror image' of the sensations one has in a white walled room. One will not see the same contrasts with a different direction of contrast, one will simply see less.

Now I want to move on to some problems for the second phase of the constrained inversion argument, again in the context of lightness inversion.

2.2. Behavioral problems

Let's now focus on the second stage in the constrained inversion argument: the idea of 'rewiring'.

With respect to rewiring, things may look fine for a stationary observer, having just one look at an environment. Things become seriously more complicated, however, when one draws behavior into the picture. Behavior can be looked at various levels, at the subpersonal level of brain processes and at the personal level of the behaving person or organism. I'll touch upon some aspects of both in turn.

Problems at the subpersonal level

A subpersonal problem arises from the asymmetry between black and white noted above, white being related to distribution of light and leading to an increase in visibility, black being connected with the decrease of the light, and thus decreasing visibility. The pupil acts to regulate inflow of light in response with overall lightning conditions. With much light coming in, for example, in a well lit white walled room, it will contract, in situations in which less light is available, it will try to maximize resources, by dilating, so as to allow as much light as possible to come in.

This all makes perfect sense in the normal situation. But with rewired

connections it does no longer do so. Consider a person who has undergone 'rewiring' in a white walled room: she will have the sensations a normal person would have in a black walled room. Because of the asymmetry noted, there will be less visible in this situation -that's a point already made. But the situation gets worse because for the rewired person, the pupil will constrict -so as to even further decrease visibility (while it was already decreased). So we see that, when the behavior of the pupil is taken into account, the problems created by the basic asymmetry become aggravated.

Personal level problems

Now consider behavior at the personal level: overt behavior. Let's return again to our white walled and our black walled rooms. It's obvious that the different experiences undergone (by a normal subject) in the two situations will be paired up with different behaviors. In the black walled room, the tendency to open a curtain or put on the light on will be much greater. This has consequences for the constrained lightness inversion thought experiment. It means that the problems encountered at the subpersonal level will simply re-occur at the personal level. The rewired person in the black walled room will have a tendency to close the curtain, diminishing available light just as pupil constrictions does.

This implies that, besides the rewiring of 'input-to-central' connections, also 'central-to-output' connections should be tampered with, so that behavior would be similar, despite inverted sensations (I take it behavioral equivalence is necessary in inversion arguments).

Suppose such behavioral inversions could be established, so that 'black' behavior could go with 'white' sensations, and vice versa. The question is: why would one have any confidence that such a central-to-output rewiring would not lead to a *reinversion of sensations* (see Dennett (1990), Cole (1990)?

We know that in cases of spatial input inversion, by wearing input inverting goggles, behavioral adaptation leads, after some time, to seeing the world (roughly) as normal again -despite keeping the goggles on (for discussion and further references, see Hurley (1998), O'Regan & Noë (in press)). Dennett, and more explicitly David Cole have rightly asked: why wouldn't the same happen in the case of color (Dennett (1990), Cole (1990)?.

Robert Kirk has proposed an answer to this question (Kirk (1994),

chapter 2). Basically, his position is to deny that in the case of lightness and hue there are any natural behavioral consequences of perceiving a surface to be this or that color (including black and white). Or better: if there are behavioral consequences, they are contingent: they could as well have been otherwise, without change in the experience. Kirk admits spatial perception has noncontingent behavioral consequences: it's no accident that I reach for objects on my left hand side (normally) with my left hand. But in the case of color, all that counts is contrast and direction of contrast (that the contrast proceeds between black and white or between white and black doesn't really seem to matter).

But the examples given above show that what Kirk says is not true once one considers three-dimensional scenes and behavior in those that's extended in time.

In fact, from what has been said above, one should rather conclude that reinversions of sensation after behavioral reinversion would be *more* probable in the lightness case than they are in the spatial case. This is precisely because of the asymmetry between black and white. Some spatial inversions, such as mirror inversions, are *symmetrical.* (one finds a very nice discussion of these in Hurley (1998), chapter 7). This means that it would be possible to invert both peripheral input-to-central connections and central-to-output connections, and in this way get a functioning individual with inverted sensations. Or at least, even if this sounds implausible (see the discussion in Hurley(1998), chapter 7 for difficulties), the point is that lightness inversion is even less plausible!

3. From lightness to hue?

But can what is said about lightness be applied to hue?

Two prima facie reasons to answer this question positively are the phenomenological and the functional similarity between hue and lightness. The similarity between hue and lightness is in the first place phenomenological, in the sense indicated by Wittgenstein (1977, III 85: "In a brightly colored pattern black and white can be next to red and green, etc. without standing out as different"). Besides, there's also a functional similarity: both hue and lightness allow surfaces to be distinguished from each other and from a background.

But one might still hold there's a crucial difference because the basic

asymmetry that was present in the white versus black case doesn't exist in the case of hue. Black is connected to the absence of light energy, white with its presence. However, one might point out, in the case of hue we don't have the same contrast. Hue is less 'brute', in not having to do with absence or presence of light, but with a sensitivity to *which kind* of light is present. Hue essentially involves the selective reflectance of wavelength. Hue *always* presupposes light, so the contrast drawn between black and white as absorbing (and darkening) versus reflecting (and illuminating), which lay at the heart of the asymmetries in lightness inversion, cannot be drawn.

But there *is* a conception of hue in which the analogy seems to hold: Jonathan Westphal's conception of hues as selective darkeners (Westphal(1991)). Instead of thinking of the colors of surfaces as selective reflectances, he invites us to think of them as selective absorbers a blue surface is one that selectively refuses to reflect yellow light, and vice versa, a red surface is one that selectively darkens red light, or the red parts of a composed light.

Once one sees color from this perspective, one sees one can apply the arguments from asymmetry in a similar way. For consider a scene illuminated by a reddish (evening) light: as far as red and green surfaces are concerned, the situation is completely analogous to the situation of black and white surfaces in neutral (white) light. Reddish surfaces will increase visibility, while greenish surfaces will not -in an asymmetrical and irreversible fashion.

So we can find a precise analogy to the physical asymmetry between blacks and whites in the domain of hue? Can we also find the motor asymmetries? Can we find something similar to pupil dilatation/constriction in the case of hue? Again we can. I'll just give one example: the differential distribution of color sensitivity over the retina. It is an established fact that, with eyes fixating a central point, red and green can only be seen in the central part of the visual field, while the yellow and especially the blue zone extend significantly further (see Hurvich (1981), p. 21). The upshot of this is that a different pattern of movement will be necessary to scrutinize the color of a red versus a blue color outside the central fovea. Because blue can be spotted in a wider area of the visual field, it will be easier to detect and will need less exploratory movement.

The conclusion to draw is that different colors come with their specific

'package' of motor patterns, each uniquely adapted to the physical characteristics of their physical stimulus. This creates problems for constrained inversions which seem similar to the motor (pupil) problems for lightness inversion pointed at above.

4. The root of the problem

After having raised some problems for constrained inversion arguments, it is time to look for the reasons why constrained lightness and hue inversion seem prima facie so compelling.

Maybe its plausibility lies in its reliance on the widely accepted traditional 'computational' view of color perception, and visual perception in general. In this view, color consciousness arises at a certain high level stage of processing. Initial peripheral stages are unconscious, till an output is produced, a representation at a certain stage of processing that somehow 'glows with the light of consciousness'.

It is the picture Dennett has called 'Cartesian materialism' (Dennett(1991)). In it, color consciousness is completely an internal affair: a strictly internal labeling of form -a way of spreading mental paint on pregiven forms. Just as a child can fill in a line drawing with color in any way it wants, the mind in principle can color the forms it gets in any way it wants.

But there is an alternative: it arises from the tradition of vision research in which vision is seen as an interaction rather than as an internal event (see Myin (2000)) for a sketch of this tradition and its alternative). A recent theory of vision within this paradigm has been developed by Kevin O'Regan and Alva Noë (O'Regan and Noë (2001), Myin and O'Regan (2002)). As said, in this picture, vision, including color vision, is not an internal affair - it is an interaction between organism and environment that is extended in time and space. The differences between colors are not arbitrary, but are rooted in different ways organisms interact with different surfaces. On this view, but not on the traditional one, the way a surface differently reacts to light (think about black and white) and the differences in the perceptual apparatus of the organism all play a role in color experience. Also, the natural bodily reactions to a stimulus determine which qualitative 'feel' will be present when the stimulus becomes consciously perceived. On this view,

experiencing a color is no longer the activation of an internal representation that necessarily stands in an arbitrary relation to the external stimulus, but it is the very specific pattern of interaction of the organism and its embodied perceptual apparatus with a surface in the world. Colors then are 'spread into the world', and thereby lose even the possibility of becoming arbitrary with respect to it. Against such a background, arguments concerning constrained inversions of sensations might make much less sense then against the traditional one.

Vrije Universiteit Brussel

REFERENCES

Cole, David (1990), 'Functionalism and Inverted Spectra', *Synthese* **82**, 207-222.

Dennett, Daniel C. (1990), 'Quining Qualia', in William G. Lycan (ed.), *Mind and Cognition: A Reader*, Blackwell, Oxford, 519-547.

Dennett, Daniel C. (1991), *Consciousness Explained,* Little, Brown, Boston.

Hurley, Susan (1998), *Consciousness in Action*, Harvard University Press, Cambridge Mass.

Hurvich, Leo M. (1981), *Color Vision*, Sinnauer, Sunderland Mass.

Locke, J. (1690/1987). *An Essay concerning Human Understanding*. Oxford: Clarendon Press

Kirk, Robert (1994), *Raw Feeling. A Philosophical Account of the Essence of Consciousnes*, Oxford University Press, Oxford.

Myin, Erik (1999), 'Beyond Intrinsicness and Dazzling Blacks', *Behavioral and Brain Sciences* **22**, 964-965.

Myin, Erik, (2000), 'Two sciences of Perception and Visual Art', *Journal of Consciousness Studies* **7**, No. 8-9, 43-55

Myin, Erik (2001), 'Color and the Duplication Assumption', *Synthese* **129**, 62-77.

Myin, & O'Regan, J.K. (2002), 'Perceptual Consciousness, Acces to Modality and Skill Theories of Perception', *Journal of Consciousness Studies* **9**, No. 1, 27-45.

O'Regan, J.K. & Noë, A. (2001) 'A sensorimotor accoount of vision and visual consciousness', *Behavioral and Brain Sciences* **24** (5), 883-917.

Shoemaker, Sydney (1975), 'Functionalism and Qualia', *Philosophical Studies* **27**, 292-315.

Westphal, Jonathan (1991), *Color: A Philosophical Introduction*, Basil

Blackwell, Oxford.
Wittgenstein, Ludwig (1977), *Remarks on Color,* ed. G.E.M. Anscombe, trans. Linda L. McAlister and Margarete Schättle, University of California Press, Berkeley.