

formerly worked in fine art and advised on colour, HJA reported that he saw the world only in terms of shades of grey following his lesion. He detected a change in his colour experience. Nevertheless, aspects of colour processing remained. For example, visual evoked responses to isoluminant colours could be measured and his ability to match isoluminant colour patches was above chance. HJA's conscious reports were clearly impaired, however, when he was asked to judge whether he was right or wrong when making colour-based responses (and even when the colour responses were correct). Thus, when asked to point to a colour token matching one pointed to by the examiner, HJA's confidence judgements bore no relation to the accuracy of his performance. He often felt he was guessing when he was right and he felt confident of being right when he was in fact wrong. His conscious experience of colour appeared to be dissociated from the residual colour processing abilities he had. This dissociation of colour experience was also distinct from HJA's conscious experience of other perceptual impairments. Thus his judgements were generally accurate when he was asked to rate his confidence about whether object and face identification responses were correct. He also showed no evidence of residual access to object or face identities, unlike the results with colour. Hence, in this instance, the degree of perceptual deficit – measured in terms of residual perceptual abilities – can be distinguished from the conscious experience of the deficit. For one class of stimulus (colour) there was better residual processing, but less insight into the deficit, than for other classes of stimulus (objects, faces). It is not simply that patients with a more profound perceptual deficit experience a more profound loss of the ability. From this we may conclude several things. For example, it may be possible to distinguish the neural substrates involved in perceptual processing from those involved in conscious awareness of their products. The neural substrate of conscious experience may also take a distributed form, and so can be dissociated for different stimuli. We suggest that detailed analysis of such patients can inform us not only about perceptual processes, but about how such processes are realised in subjective experience.

## Overlooking the resources of functionalism?

Zoltán Jakab

Institute of Interdisciplinary Studies, Carleton University, K1S 5B6 Ottawa, Ontario, Canada. [zjakab@ccs.carleton.ca](mailto:zjakab@ccs.carleton.ca)

**Abstract:** Although the author's critical view of functionalism has a considerable intuitive pull, his argument based on the color room scenario does not work. Functionalism and other relational views of the mind are capable of providing coherent accounts of conscious experience that meet the challenge set up by the "color room argument." A simple example of such an account is presented.

Palmer claims that because functionalism can give only a relational picture of the mind, it will fail to capture the intrinsic qualities of experience. Experiential qualities are below the level of relational isomorphism that can be captured by the methods of behavioral science in general and functionalism in particular. In support of this claim, Palmer sets up an argument in two versions (sect. 2.5 and sect. 4). First let me reconstruct this argument.

**Version one.** The color machine in the color room satisfies all functionalist (computational) requirements associated with color discrimination and color-related behavior. A functionalist should therefore conclude that the color machine in the color room has color experiences. But it is intuitively implausible that the color machine has any color experience. Therefore functionalism is probably wrong.

**Version two.** Put yourself in the color room, thereby bypassing the other mind's problem. Master the computation that the color machine performs; in this case, you become the color machine, you satisfy all functionalist requirements for having color experiences, hence (so the

## References/Palmer: Color and consciousness

functionalist must argue) you will have color experiences simply by means of doing that calculation. But again, very plausibly, you will not have any color experiences by means of doing that calculation. Therefore functionalism is presumably false.

Overall, the structure of the argument is *modus tollens*:

1. If functionalism is right, then the color machine (or you) necessarily have color experiences merely by means of performing the color-related computations.

2. Neither the color machine, nor you (merely by means of doing the relevant calculations) would have any color experience.

Therefore 3. Functionalism is wrong.

Now let me give a reply. Functionalism might be wrong (i.e., incapable of accounting for conscious experience), but Palmer's argument based on the color-room scenario is insufficient to show this. The argument is not sound because the first premise is unsupported. (I will not address the second premise in this commentary, even though doubts might arise about it as well.) Let us see what the problem is with the first premise.

**Version one.** The analogy between the *human brain as a whole* (or some implementation of its functionally/computationally relevant structure) and the color machine does not hold up. The color machine is at most the model of an isolated subsystem of the brain. Should the functionalist conclude that it has color experiences? I think functionalism is not at all committed to drawing this inference.

Compare: Would a visual brain *in itself*, isolated from the rest of the brain, floating in some suitable solution, receiving appropriate optical stimuli, have color experiences? *When embedded in the neural/functional architecture of the rest of the brain*, the well-functioning visual brain does give rise to color experience. But does it do so in isolation? This is questionable, to say the least.

**Version two.** Does functionalism entail that the human subject in the color room must have color experiences simply in virtue of performing the relevant calculations? I think not. Here is why.

There is a possible analysis of experiential qualities, which is (1) relational and (2) not yet ruled out as insufficient: Perhaps the experience of seeing red is a relation between a subject and a certain type of physiological state. The relation is "undergoing" a state: a token of a physiological state type occurs in one's brain *in the appropriate way* – for example, it is a well-characterized activity pattern of area V4. (Additional background conditions like normal awake state or REM sleep, sufficient attention to events of visual perception, etc., can be assumed.) My experience of seeing red *is* the undergoing relation between me and that particular state – the relation set up by that physiological state occurring in my visual brain. Moreover, in describing such a physiological state, we necessarily resort to some kind of abstraction. This already happens when we specify the physiological activity *type* that, for example, is tokened in V4 when the subject sees red. In giving such types we leave out idiosyncratic biochemical and physiological variations as irrelevant and specify a generalizable physiological pattern. Perhaps we can even specify some sort of computational operation that is performed by that physiological event. Furthermore, perhaps seeing a color is a subject's undergoing relation to an inner state *type* – *identified at the computational level*. If this is right, then we have a functionalist account of qualia.

Intuitively, this account may seem too austere. Is this so much the worse for the account, *or* so much the worse for the intuition? A difficult question. It has not yet been convincingly argued by anyone (to my knowledge) that this account cannot be right. An obvious problem with it is that it has yet to be spelled out in reasonable detail; however, even in this extremely sketchy form it absolves the functionalist of the burden of concluding that the subject in the color room has to have color

experiences simply by virtue of performing the color-related calculations. If I realize the color-discrimination process by calculating the appropriate computational algorithm in my head, this process will primarily involve my higher order cognitive machinery. Hence this mental simulation does not at all imply that the physiological states underlying color experiences occur in my visual system; mental calculation need not involve activity in V4 or any other color-processing area. If I am the person in the color room, then it should not be at all surprising that I have no color experiences merely by means of doing that calculation.

Note also that the proposed relational account predicts that zombies physiologically identical to us are logically impossible. Once human subjects entertain some relevant physiological state, then by definition they have the corresponding experience.