Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Consciousness and Cognition 15 (2006) 48–50

Consciousness and Cognition

www.elsevier.com/locate/concog

Commentary

Hoffman's "proof" of the possibility of spectrum inversion ☆

Alex Byrne [a], David Hilbert [b,*]

[a] *Department of Linguistics and Philosophy, Massachusetts Institute of Technology, 77 Massachusetts Avenue,
Cambridge, MA 02139-4307, USA*

[b] *Department of Philosophy, Laboratory of Integrative Neuroscience, University of Illinois at Chicago,
601 S. Morgan Street, Chicago, IL 60607-7114, USA*

Philosophers have devoted a great deal of discussion to the question of whether an inverted spectrum thought experiment refutes functionalism. (For a review of the inverted spectrum and its many philosophical applications, see Byrne, 2004.) If Hoffman is correct the matter can be swiftly and conclusively settled, without appeal to any empirical data about color vision (or anything else). Assuming only that color experiences and functional relations can be mathematically represented, a simple mathematical result—the Scrambling Theorem—shows that color experiences can be permuted while keeping functional relations constant, thus contradicting functionalism.

It will be helpful to illustrate Hoffman's argument with a very simple example. Suppose that there are two color experiences—experiences of red (state R) and experiences of green (state G), and that the functionalist identifies R with functional state F_{R}, and G with functional state F_{G}. A person can only be in one of these functional states at any particular time. According to the functionalist, then, it is impossible for R to obtain without F_{R}, and vice versa.

First, let us put the functionalist theory into the format of the Scrambling Theorem, using Hoffman's notation, and 'R,' 'F_{R}', etc., to stand for mathematical surrogates for the experiences and functional states. Let X (the set of "color experiences of one person"), and Y (the set of "color experiences of a second person") both be $\{\mathrm{R}, \mathrm{G}\}$. As Hoffman emphasizes, the argument does not depend on restricting the choice of the other functions and sets mentioned in the Scrambling Theorem, so we may take them to be as simple as possible. Let W be $\{F_{\mathrm{R}}, F_{\mathrm{G}}\}$, and define f: $X \rightarrow W$ as $f(\mathrm{R}) = F_{\mathrm{R}}$ and $f(\mathrm{G}) = F_{\mathrm{G}}$. (Hence $n = 1$.) Define the bijection b: $X \rightarrow Y$ as: $b(\mathrm{R}) = \mathrm{G}$ and $b(\mathrm{G}) = \mathrm{R}$.

Then, in this simple case, the Scrambling Theorem tells us that there is a function g: $Y \rightarrow W$ such that $f = gb$. And obviously there is: $g(\mathrm{G}) = F_{\mathrm{R}}$, and $g(\mathrm{R}) = F_{\mathrm{G}}$.[1]

So far, so good. Why does Hoffman think this mathematical result shows that, *contra* the functionalist, it *is* possible for R to obtain without F_{R}? (*Note*. In this last sentence we are using 'R' and 'F_{R}' to stand for an experience and a functional state, not the mathematical surrogates.)

☆ Commentary on Hoffman, D.D. (2006). The scrambling theorem: A simple proof of the logical possibility of spectrum inversion. *Consciousness and Cognition, 15*, 31–45.

* Corresponding author.
E-mail address: hilbert@uic.edu (D. Hilbert).

[1] We are ignoring X', f', etc., since for maximal simplicity we may suppose they do no work in representing the functionalist theory. Incidentally, Hoffman defines Y' as $b(X')$, but X' is a subset of 2^X, and b is not defined on subsets of 2^X; presumably he meant that $Y' = \{b(x) | x \in X'\}$.

1053-8100/$ - see front matter © 2006 Published by Elsevier Inc.
doi:10.1016/j.concog.2006.02.003

Consider Jack, as described by the functionalist. He has color experiences R and G, and the functionalist theory identifies R and G with F_R and F_G, respectively, as specified by the function *f*, $f(R) = F_R$ and $f(G) = F_G$. The Scrambling Theorem tells us that there is another function *g*, $g(G) = F_R$, and $g(R) = F_G$. So consider Jill, whose experiences and functional states are described by *g*, not *f*. Since the range of Jill's functional states is the same as Jack's, she will "[respond] identically to Jack." But, because she is described by *g*, when she is in functional state F_G (and hence not in F_R) she will be in R, not G. Hence the functionalist theory is refuted.

Hoffman points out that "[t]he proof applies not just to color experiences, but to all experiences in all sensory modalities," but here he has been overcome by modesty. The argument makes no use of the special features of *functional* states: by applying the Scrambling Theorem we can just as easily conclude that color experiences can be inverted with respect to neural states, ectoplasmic states, or whatever. Moreover, the argument makes no use of the special features of *experiences*, as opposed to other mental states and events. If the argument works at all, it works against the widely held theory that *belief* is a functional state, and indeed against *any* reductive theory of mind.

That might sound impressive enough, but it does not even begin to approach the limits of Hoffman's argument. When it is detonated, no putative necessary connection is left standing. Suppose you think that gold is the element Au, lead is the element Pb, and (hence) that it's not possible that gold is not Au, and that lead is not Pb. A simple variant of the above argument, deploying the bijection that maps gold to Pb and lead to Au, shows that you are wrong. Suppose you think that I. Lewis Libby is identical to Scooter Libby, that Hesperus is identical to Phosphorus, and (hence) that it's not possible that I. Lewis Libby is not Scooter Libby and that the planet Hesperus is not the planet Phosphorus. ('Hesperus' and 'Phosphorus' are Ancient Greek names for Venus.) Wrong again. The bijection mapping I. Lewis Libby to Phosphorus, and Hesperus to Scooter Libby, shows that I. Lewis Libby might not have been Scooter Libby, and might instead have been Phosphorus. Since nothing in Hoffman's argument depends at all on the nature of the individuals and properties the Scrambling Theorem is applied to, the argument applies to everything. In Hoffman's hands, the Scrambling Theorem is the *alkahest* of metaphysics, dissolving necessity into possibility.

At this point, we can be quite sure that something has gone wrong. Universal solvents are no more to be found in philosophy than they are in chemistry. A clue to Hoffman's error can be seen in the discussion of his fifth objection. There he considers the claim that "each conscious sensory experience is *identical* to some tracking relationship." Hoffman replies that "[i]f we wish to evaluate that claim, then we must ask if it is logically possible that the two are not identical, and hence we look for proofs like the Scrambling Theorem; we cannot evaluate the identity claim by assuming it is true and then concluding that any logical proof to the contrary is impossible" (p. 14). But we equally well can't proceed by assuming non-identity, which is essentially Hoffman's procedure. If we go back to our toy illustration of Hoffman's argument, the error is in the step from the existence of the function *g*, to the conclusion that there could be someone whose functional states and experiences are correctly described or modeled by *g*. There is a function that maps gold to Pb, and one might use this function to represent the eccentric theory that gold really is Pb. But it would be a blunder to conclude from this that the eccentric theory is *possible*—that gold *might have been* Pb, and so is not (necessarily) identical to Au; Hoffman's error is of exactly the same kind.

What could have led Hoffman into making such an elementary mistake? Part of the explanation might be that he has conflated a narrow sense of logical possibility with what philosophers usually call *metaphysical* possibility. The latter is the sense of possibility at issue in discussions of the inverted spectrum; rather confusingly it is sometimes also called 'broadly logical possibility' (Plantinga, 1999). To add to the confusion, 'logical possibility' has another, more standard use, on which it means something like *ideally conceivable*; whether logical possibility in this sense is the same as metaphysical possibility is controversial (see Chalmers, 1999).

To say that a sentence S is logically possible (in the narrow sense) is simply to say that S is not a logical contradiction (or, more precisely, that its canonical representation in some logical language, say the first-order predicate calculus with identity, is satisfiable). And at least once we are given the canonical representation of a sentence S, logical possibility (in the narrow sense) *is* the sort of thing that admits of mathematical proof, in particular proofs that invoke various set-theoretic structures. For instance, assuming that '$\sim a = b$' is the canonical representation of 'I. Lewis Libby is not Scooter Libby,' one may prove that 'I. Lewis Libby is not Scooter Libby' is logically possible by defining an interpretation function *I* on a domain (set) with at least two elements x and y, such that $I(`a') = x$ and $I(`b') = y$. However, this does not show that I. Lewis Libby *might not have been* Scooter Libby—in other words, that it is metaphysically possible that I. Lewis Libby is not Scooter Libby. There is no doubt that spectrum inversion scenarios are logically possible in the narrow sense, but that is not what the philosophical debate is about.

There is important lesson to be learned from Hoffman's "proof," though, albeit not the one he had hoped. Excess abstractness is a constitutional disorder of the philosophical temperament, and everybody would prefer proof to inconclusive argument. This sometimes leads to extravagant claims to settle substantive philosophical questions solely by using logic and mathematics, as beautifully exemplified by Hoffman's article. As Saul Kripke famously put it: "It should not be supposed that the formalism can grind out philosophical results in a manner beyond the capacity of ordinary philosophical reasoning. There is no mathematical substitute for philosophy" (1976, p. 416).[2]

References

Byrne, A. (2004). Inverted qualia. In E. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2005 ed.). Available from: <http://plato.stanford.edu/archives/sum2005/entries/qualia-inverted/>.

Chalmers, D. (1999). Materialism and the metaphysics of modality. *Philosophy and Phenomenological Research, 59*, 473–493.

Kripke, S. (1976). Is there a problem about substitutional quantification? In G. Evans & J. McDowell (Eds.), *Truth and Meaning*. Oxford: Oxford University Press.

Plantinga, A. (1999). Modalities: basic concepts and distinctions. In J. Kim & E. Sosa (Eds.), *Metaphysics: An Anthology*. Oxford: Blackwell.

[2] Thanks to Agustín Rayo for discussion.