Available online at www.sciencedirect.com

SCIENCE DIRECT®

Consciousness and Cognition 15 (2006) 31–45

Consciousness and Cognition

www.elsevier.com/locate/concog

The scrambling theorem: A simple proof of the logical possibility of spectrum inversion

Donald D. Hoffman [*]

Department of Cognitive Science, University of California, Irvine, CA 92697, USA

Received 20 February 2005
Available online 2 August 2005

Abstract

The possibility of spectrum inversion has been debated since it was raised by Locke (1690/1979) and is still discussed because of its implications for functionalist theories of conscious experience (e.g., Palmer, 1999). This paper provides a mathematical formulation of the question of spectrum inversion and proves that such inversions, and indeed bijective scramblings of color in general, are logically possible. Symmetries in the structure of color space are, for purposes of the proof, irrelevant. The proof entails that conscious experiences are not identical with functional relations. It leaves open the empirical possibility that functional relations might, at least in part, be causally responsible for generating conscious experiences. Functionalists can propose causal accounts that meet the normal standards for scientific theories, including numerical precision and novel prediction; they cannot, however, claim that, because functional relationships and conscious experiences are identical, any attempt to construct such causal theories entails a category error.

Keywords: Color; Empirical possibility; Locke; Logical possibility; Nonreductive functionalism; Qualia; Reductive functionalism; Representationalism; Spectrum inversion; Supervenience

[*] Fax: +1 949 824 2307.
E-mail address: ddhoff@uci.edu.

1053-8100/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.concog.2005.06.002

1. Introduction

Is it possible, John Locke pondered in his Essay of 1690/1979, that "the idea that a violet produced in one man's mind by his eyes were the same that a marigold produced in another man's, and vice versa." Could the colors I experience differ from yours, even if experiments reveal no difference between us? Locke's question is raised by inquisitive children but remains hotly debated by scientists and philosophers because its answer is key to current theories of the relationship between brain activity and conscious experience (Bickle, 2003; Braddon-Mitchel & Jackson, 1996; Chalmers, 1995, 1996, 2002; Churchland, 1996, 2002; Clark, 1983; Crick & Koch, 1998; Gregory, 1998; Metzinger, 2000; Tye, 2000). If color scrambling of the type Locke envisioned is logically possible, this would entail that conscious experiences could change without concomitant functional changes in brain states. Locke's question has stirred prolific debate through the ensuing centuries (see, e.g., Palmer, 1999, and its commentaries) but no mathematical articulation or proof. Here, we prove that the answer is "Yes": color scrambling is logically possible.

2. Reductive functionalism

Reductive functionalism asserts that the type identity conditions for mental states refer only to relations between inputs, outputs, and each other (e.g., Block & Fodor, 1972). It does not assert that such relations *cause* mental states but rather that they are *numerically identical to* (i.e., one and the same as) mental states. Thus, each conscious experience, being a species of mental state, is numerically identical to some such relations. Although in humans such relations are presumably instantiated in the nervous system, they could in principle be equally well instantiated in other physical systems, such as computers.

Reductive functionalists typically deny that the experiences of one person could be scrambled from those of another without experimental consequences (Churchland, 2002; Dennett, 1998). For if conscious experiences are identical to functional relations, then to scramble experiences *is* to scramble the functional relations, and this would create measurable differences between the two persons in controlled experiments. Nonreductive functionalists, who hold that conscious experience is determined by functional organization but not reducible to functional organization, can allow the logical possibility of scrambled experiences but deny the empirical possibility (Chalmers, 1995). Such nonreductive functionalism is not the primary target of the Scrambling Theorem developed in this article.

A critic of reductive functionalism might ask how it could be that functional relations could cause conscious experiences. The reply would be swift and short: to ask that question is a category mistake, betraying a fundamental misunderstanding. Reductive functionalism does not claim that functional relations *cause* conscious experiences, or that functional relations are *correlated* with conscious experiences, but that they *are* conscious experiences. Hence, to ask how functional relations could cause conscious experiences is the same mistake as to ask how 12 could cause a dozen. One cannot impugn reductive functionalism by complaining that reductive functionalists do not say how functional relations cause, or why they are correlated with, conscious experiences. One can, however, reasonably ask a nonreductive functionalist to give a scientific theory describing how functional states cause or determine conscious experiences.

Then, how might one refute reductive functionalism? It is, after all, an empirical claim and must therefore be open to potential refutation. The answer is simple. To refute reductive functionalism, one needs to show that it is *logically* possible that conscious experiences are not numerically identical to functional relationships. This does not mean that we must show that, as a matter of empirical fact, conscious experiences are not the same as functional relations. This would be a much harder program and much stronger than is necessary to refute reductive functionalism. All we need to do is to show that it is logically possible that conscious experiences are not the same as functional relations.

How could we do that? One way is to show that it is imaginable that conscious experiences are not the same as functional relations. If something is imaginable, then it is logically possible. This appeal to imagination is often made in philosophical attempts to refute claims of numerical identity. Of course, this approach has a severe weakness. Claims of imagination can be disputed, leading not to clean refutation but merely to partisan bickering. The opponent of reductive functionalism can claim to imagine conscious experiences apart from functional relations, and the reductive functionalist can simply deny that the opponent has succeeded in the claimed act of imagination. The result is an unsatisfying stalemate.

There is, however, another approach that is more compelling than bald claims of imagination: give a mathematical proof that one member of a proposed identity could have some property that the other does not, thus entailing, by Leibniz's Law, that the two are not identical. This eliminates argument over imagination and replaces it with sober discussion of the assumptions required for the proof.

3. The Scrambling Theorem

This section presents a simple proof that color experiences and functional relations are not identical. The proof requires one assumption.

Assumption. Conscious experiences can be represented mathematically.

One might object to this assumption on the grounds that there are no conscious experiences or that, even if there are, they cannot be represented mathematically. The first objection is rare, and compelling to few, but has been made (e.g., Dennett, 1998). It not only obviates the Scrambling Theorem proved here but renders pointless any discussion of whether conscious experiences are identical to functional relations, the key claim considered here. Therefore, we have no further concern with it.

The second objection precludes the possibility of a science of consciousness. For if conscious experiences cannot be represented mathematically, then one can make no quantitative predictions. This objection might turn out to be true but we should not embrace it unless forced by repeated failure to construct a scientific theory of consciousness.

We therefore assume that conscious experiences can be represented mathematically. This allows us to assert that the color experiences of a person can be represented by a set, X. From this assertion, the Scrambling Theorem follows. Of course, representing color experiences by a set does not entail any assumption that different color experiences are unrelated to each other. The set X can have, as we shall see, further mathematical structures imposed on it that model the relations among the color experiences.

$$\begin{tikzcd}
X \arrow[r, "f"] \arrow[d, "b"'] & W & & X' \arrow[r, "f'"] \arrow[d, "b"'] & Z \\
Y \arrow[ur, "g"'] & & & Y' \arrow[ur, "g'"'] &
\end{tikzcd}$$

Fig. 1. Commuting diagrams of the Scrambling Theorem.

Scrambling Theorem. *Denote by X the color experiences of one person and by Y the color experiences of a second person. Let the following be any mathematical structures and functions that describe the functional relations among color experiences of the first person*: (1) *f: $X_1 \times \cdots \times X_n \to W$, where $X_1, \ldots, X_n$ are copies of X, n is a positive integer, and W is any set*; (2) *$X' \subseteq 2^X$; and* (3) *$f' : X'_1 \times \cdots \times X'_m \to Z$, where $X'_1, \ldots, X'_m$ are copies of X', m is a positive integer, and Z is any set. Let b: $X \to Y$ be any bijection, i.e., any scrambling, between the two sets of color experiences. Then, there exist*: (1) *g: $Y_1 \times \cdots \times Y_n \to W$, where $Y_1, \ldots, Y_n$ are copies of Y*; (2) *$Y' \subseteq 2^Y$; and* (3) *$g' : Y'_1 \times \cdots \times Y'_m \to Z$, where $Y'_1, \ldots, Y'_m$ are copies of Y', such that* (1) *$f = gb$*, (2) *$X' = b^{-1}(Y')$, and* (3) *$f' = g'b$.*

Proof. (1) Define $g = f\,b^{-1}$. (2) Define $Y' = b(X')$. (3) Define $g' = f'b^{-1}$. □

The Scrambling Theorem is illustrated, for the case $n = m = 1$, by the two commuting diagrams shown in Fig. 1. The diagram on the left illustrates that, given sets X, Y, and W, and given any function f from X to W, then for every bijective scrambling b from X to Y, there exists a function g from Y to W that makes the diagram commute, i.e., such that $g = fb^{-1}$. Similar comments apply, mutatis mutandis, for the diagram on the right. (Recall that a bijection is a one-to-one and onto mapping, i.e., a mapping that is both injective and surjective.)

4. The meaning of the Scrambling Theorem

Suppose we show Jack a display, and he has color experiences x_i. If we show Jill the same display, then she will have color experiences $b(x_i)$, since b describes how Jill's color experiences are scrambled relative to Jack's. We ask Jack to perform a color task; his responses are determined by the functional relations f, X', and f'. We ask Jill to perform the same task; her responses are determined by the functional relations g, Y', and g'. The Scrambling Theorem says that it is always possible to choose g, Y', and g' so that Jill responds identically to Jack, even though her conscious experiences are scrambled relative to his. In this case, their functional relations are identical but their conscious experiences differ; therefore, their conscious experiences cannot be numerically identical to the functional relations. To put this in another way, conscious experiences do not supervene logically on functional relations.

Notice that the Scrambling Theorem does not require a specific mathematical structure for the functional relations f, X', and f'. The function f could be, for instance, a distance metric, a quasi-pseudometric, a partial order, or any one of countless other relations. The structure X' could be the open sets of a topology, the measurable sets of a σ-algebra, or any one of countless other structures. The function f' could be, for instance, a metric or probability measure. It does not matter, since the Scrambling Theorem makes no assumption about the specific nature of the functional relations and holds for whatever particular functional relations happen to be appropriate. The Theorem applies, therefore, not just to the scrambling of color experiences but also to the scrambling of any sensory experiences. The spaces of the sensory experiences need have no special symmetries for the Theorem to apply, since any bijection will do.

5. Objections to the Scrambling Theorem

Several objections have been raised to the Scrambling Theorem and its interpretation. One objection is that the points of X and Y could not represent conscious experiences because such points could, instead, be taken to represent internal codes or states of, say, unconscious robots. Thus, according to this objection, although the Scrambling Theorem is mathematically correct, it does not apply to conscious experiences. This objection makes an elementary mistake of logic. Although the Scrambling Theorem could indeed be applied to the unconscious states of a robot, this possibility does not preclude applying the Scrambling Theorem to conscious experiences any more than using the integers to count apples precludes using them to count oranges.

A second objection is that the Scrambling Theorem is not adequately tied to the plethora of empirical facts about color perception, facts about cone sensitivities, opponent color channels, the role of surface reflectances, and the known metrical structure of color space. The Scrambling Theorem, according to this objection, is interesting only insofar as it connects to the "real world," and otherwise does not tell us about genuinely possible inverted spectra in actual human beings.

This objection is also an elementary mistake. The target of the Scrambling Theorem is not an empirical claim about actual spectrum inversions in human beings. The target is the *reductive* functionalist claim that color experiences and functional relations are identical. To refute this claim of identity, the Scrambling Theorem establishes the *logical* possibility of inverted spectra, not their *empirical* existence or nature. The Scrambling Theorem shows that no appeal to empirical details about color is needed to establish the logical possibility of inverted spectra. Therefore, discussion of the empirical data would, in this context, obscure the fundamental point of logic made by the Scrambling Theorem. If our goal was to disprove *nonreductive* functionalism, then empirical data on color perception would, in all likelihood, be relevant in formulating such a proof.

A third objection is longer. According to this objection, the possibilities for set-theoretic transformations used in the Scrambling Theorem are well known and not in dispute in the inverted-qualia literature. The real issue is whether any such formal transformations could in fact correspond to alternative phenomenal realities. Many claim that they could not, arguing that specific qualia, i.e., conscious experiences, have intrinsic natures that inherently constrain their relations, including their similarity and resemblance relations. If that were so, it might well not be possible to preserve those phenomenal resemblance relations under an inversion or scramble. The mere formal possibility of defining some distance metric or other mathematical structure that could be preserved does not in itself show that those structural relations could correspond with the reality of phenomenally perceived resemblance. The key point here is whether or not specific qualia are mere items in a set which can be structured arbitrarily or whether they have intrinsic natures that determine, or at least constrain, their possible relations. The Scrambling Theorem seems to assume the former and does not really give any arguments to refute the latter alternative view.

One interpretation of this objection is that color experiences cannot be represented by a set. This interpretation precludes any mathematical analysis and so is probably not intended. A second interpretation, then, is that a person's color experiences can be represented by a set but, because of the intrinsic qualities of such experiences, only certain mathematical structures on this set are possible. But we must ask, Possible in what sense? Empirically or logically? If empirical possibility is meant, then this objection makes the same elementary mistake as the second objection. If logical possibility is meant, then the Scrambling Theorem proves the logical possibility by producing the required mathematical structures. This result might strain the imagination: How could a red be more similar to a green than to another red? But imagination here appeals to empirical intuitions when what is at question is *logical* possibility. When imagination conflicts with mathematical proof, imagination must yield, for it is proved wrong. If the argument of the Scrambling Theorem is well known, it is apparently not well understood.

A fourth objection is that the Scrambling Theorem is mathematically trivial. Indeed it is. Once one has formulated the question of logical possibility at the appropriate level of mathematical abstraction, the proof is a single line. The real work, then, is finding the proper mathematical formulation. It is a bonus, not a defect, that once the proper formulation is found the answer becomes disarmingly clear.

A fifth objection concerns representationalist accounts of qualia (e.g., Harman, 1996; Shoemaker, 2000; Tye, 1996, 2000). For representationalists, qualia are not properties of brain states but properties that those states represent objects as having, and those representational relations are determined by the functional or causal or information–theoretic relations between the brain states and the objects. Given such a view, the idea of switching red qualia states with green ones in terms of their total functional roles is incoherent. Being a red qualia state just is a matter of representing objects as being red, and that representational content is determined by the state's role and connections to the relevant sorts of objects. Lots of people reject this sort of strong representationalism, and indeed, they sometimes appeal to the intuitive appeal of inverted-qualia thought experiments to challenge it, but nonetheless it is a serious view in the current debate, and if it was true it would surely undercut the Scrambling Theorem's claim about the possibility of qualia scrambling.

This objection simply gets representationalism wrong. Representationalism does not entail, as this objection asserts, that qualia inversion is incoherent. Tye, for instance, notes that "the pure representationalist can happily allow the possibility of phenomenally inverted color experiences in normal observers" (p. 107, 2000) and "the pure representationalist can admit with impunity the conceptual possibility of phenomenal inversions without representational difference in both swamp duplicates and human beings" (p. 112, 2000). Moreover, the objection that representationalism, if true, "would surely undercut the Scrambling Theorem's claim about the possibility of qualia scrambling" misses the point. Representationalism, at least according to Tye, claims each conscious sensory experience is *identical* to some tracking relationship, i.e., to causal covariation under optimal conditions. If 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 must therefore be impossible.

A sixth objection is that the Scrambling Theorem uses discrete mathematics and therefore fails to apply to some sensory experiences, such as color experiences, that are plausibly assumed to be continuous. This objection misunderstands the formalism of the Scrambling Theorem. The sets X and Y can be finite, countable, or uncountable, and the Theorem applies in each case. By describing functional relations on X by the mapping f: $X_1, \ldots, X_n \rightarrow W$, one does not thereby require X to be discrete or to have only n elements.

A seventh objection is that the mapping f: $X_1, \ldots, X_n \rightarrow W$ requires the listing of (possibly infinite) sets of n tuples. According to this objection, one can get away with that for small sets of discrete elements, where listing such sets is a reasonable way to represent them and searching through them is a reasonable way to verify them, but in very large and/or continuous domains, however, the relations are going to have to be computable by some productive algorithm other than retrieval from lists of n tuples, and it does not seem reasonable that such systems can pre-store the set of n tuples for a true continuum.

This objection misses the point. The Scrambling Theorem answers Locke by establishing a point of logic. It is not intended to provide a computational model of human sensory processing. The Scrambling Theorem uses mathematical structures that capture all relevant functional relations. However, to answer Locke's question, the Theorem does not need to specify how these functional relations might in fact be represented or computed in human sensory systems. This empirical issue, although interesting, is irrelevant here. Indeed, to focus on such details would be to obscure the simple answer to Locke's question.

An eighth objection claims that representations of experiences could differ from one another in essentially nonformal ways, depending on whether the logical properties of the represented relations (such as symmetry and transitivity) were necessarily present in the representing relations (as in the case of representing color experiences by points in a three-dimensional geometric space) or present only by virtue of the representational mapping into more general (i.e., less constrained) structures (as in the case of representing color experiences by elements in sets of n tuples). The reason why set-theoretic formulations are so tremendously useful in mathematics is probably because they are so tremendously unconstrained: anything seems to be representable within them. It is very much in this spirit that set-theoretic concepts are used to formulate the Scrambling Theorem and why it seems so completely general. The issue is whether the very generality of the formalization methods used might throw the proverbial baby out with the bathwater. It is certainly worth entertaining the possibility that more intrinsic constraints than are available in set theory are required to adequately capture the nature of the functional relations relevant to functionalist accounts of experience.

This objection makes a fundamental mathematical mistake. It contrasts the unconstrained nature of representations based on the sets to the necessarily constrained nature of representations based on, e.g., three-dimensional geometric spaces. But three-dimensional geometric spaces are themselves nothing but sets with certain added structures, e.g., metrics or group actions. The Scrambling Theorem uses sets but allows any formal structure to be imposed on those sets that is required to model functional relations. If one needs to transform the set of color experiences into a three-dimensional metric space, then the functions f and f' of the Scrambling Theorem are there to oblige. If one needs to transform the set of color experiences into a topological or measurable space, the structure X' is available for this purpose. The way one obtains constrained or structured representations in mathematics is to start with sets and add structure. Even moves from set theory to the theory of classes or to category theory do not obviate this method of creating structured representations.

A ninth objection claims that the Scrambling Theorem assumes that an observer's experiences can be represented as a simple set of simple elements and that a scrambling of experiences can be represented as a bijection between the simple elements of two such simple sets. One might allow for the possibility of a mathematical representation of experience as some sort of set-theoretic structure but deny that a simple set of independent elements is adequate and in turn insist that such a representation is properly a more complex set-theoretic structure. This would then allow that a scrambling of experience might be represented by a map to a different set-theoretic structure or a set-theoretic structure not necessarily isomorphic to the first. The implication of this possibility for the main argument here is that the Scrambling Theorem has shown that scrambling of experience entails no functional differences in any world where experiential states and functional properties of such states can be mathematically represented by simple sets and bijections between such sets. The salient question now, in consequence, is whether a simple unstructured set is an adequate mathematical representation of our experiential states. The challenge then is to argue against such a representation. This is progress in the sense that it helps to clarify the proper nature of the disagreement.

This objection misunderstands the Scrambling Theorem. The Theorem does not require representing experiential states as simple unstructured sets. It provides the functions f and f' precisely to allow the sets of experiential states to be appropriately structured. Once such structures are properly in place, the bijections induce morphisms of those structures, i.e., the bijections preserve the structures. The next section provides a concrete example of the Scrambling Theorem in which these structures and morphisms can be clearly seen.

A tenth objection states that a radical functionalist—one who thinks that all there is to mental states is their relations to all other mental states—would simply deny that the Scrambling Theorem actually scrambles anything of consequence at all. Rather the scrambling would simply "rename" the experiences by assigning them to different formal symbols. The reason is, of course, that for the radical functionalist, there is absolutely nothing to any mental state beyond these relations: the mental states are devoid of content except for the relations, and if *all* of the relations are preserved, then *everything* about the mental states is preserved. And since the Scrambling Theorem proves that all the same relations are intact after scrambling, nothing of any functional consequence has been changed in the slightest by the scrambling. In this sense, the Scrambling Theorem is essentially a proof that, within radical functionalism, any assignment of mental states (or experiences) to symbols is as good as any other because it proves that all of the same relations are there despite the scrambling. To say this in another way, for a functionalist, all scramblings of a given mental domain are members of the same equivalence class: there is simply no basis on which to distinguish among them.

This objection misunderstands the purpose of the Scrambling Theorem. The radical functionalist this objection considers could be of two types: (1) one who denies that there are any conscious experiences and (2) one who asserts that there are conscious experiences and that these experiences are identical to certain functional relations. The Scrambling Theorem is not addressed to the first type of radical functionalist. If such a functionalist refuses to grant the existence of conscious experiences, that is fine. The Scrambling Theorem does not try to dissuade one from this position. It is not surprising, however, that few claim to be of this first type; if we know anything, we know we have conscious experiences. But if, as is more common, the radical functionalist claims to be of the second type, claims that conscious experiences are *identical* to functional relations, then the Scrambling Theorem refutes that claim.

An eleventh objection claims that the Scrambling Theorem shows only something about relations between formal representations, but not about relations between conscious experiences. Strictly speaking, this objection is true. But then it is also true, in the same strict sense, that Peano's axioms apply only in number theory, but not, e.g., to apples. But we do not conclude, therefore, that Peano's axioms are irrelevant to counting apples. Nor should we conclude that the Scrambling Theorem is irrelevant to identity claims made about conscious experiences. Once one grants that there are conscious experiences and that these experiences can be represented by sets, one has granted the relevance of the Scrambling Theorem to conscious experiences. All empirical applications of mathematics require a similar granting of relevance.

6. The Probabilistic Scrambling Theorem

The Scrambling Theorem is intentionally abstract to make clear the logical possibility of spectrum inversion. For some readers, however, the implications of this theorem might more readily be grasped if cast more concretely and discussed in more detail. This is the goal of this section.

To do this, let us again denote possible color experiences of Jack by X and those of Jill by Y. Such experiences can be more or less certain: thus we can describe *probabilities* of sensory events, e.g., the probability that Jack sees orange. The probabilistic nature of sensory experiences is a basic assumption universally employed by psychophysicists who use Signal Detection Theory (Green & Swets, 1988; Wickens, 2002) and by researchers in computational vision who model perception as a process of Bayesian inference (Bennett, Hoffman, & Prakash, 1989; Knill & Richards, 1996). Therefore, following the standard theory of probability, we assume that X and Y are probability spaces, in which events of X are certain subsets of X (Fine, 1973). In this setting, we can consider the probabilities, p, of various color events of X.

We also wish to assume as little as possible about the scrambling function, b, that maps Jack's color experiences X to Jill's color experiences Y. At a minimum, we need this scrambling function to respect the structure of color events for X and Y, so that statements about probabilities for Jill's color events can be translated into corresponding statements for Jack's color events. A scrambling function that does this is called a measurable function or, if it is real valued, a random variable (Ash & Doleans-Dade, 2000).

Note the purely probabilistic nature of our setting. The sensory spaces X and Y need not have a dimension. The scrambling function b need not be linear or continuous.

A shade of red is more readily discriminated from green than from another shade of red. We wish to model all such empirical similarities and differences between color experiences. All of them must be preserved by our function that scrambles color experiences between Jack and Jill, otherwise controlled experiments could detect the scrambling.

One way to describe these similarities and differences is to define a distance metric, d, between color events such that, the more similar two events are, the less distance lies between them. Again, we want to assume as little as possible. We have a probability, p, for Jack's sensory events, and we wish to assume nothing more about them. Therefore, we derive our notion of distance solely from this probability.

This can be done simply. As illustrated in Fig. 2A, let event A represent one color experience and B a second color experience. The symmetric difference of A and B, denoted $A\Delta B$, is the

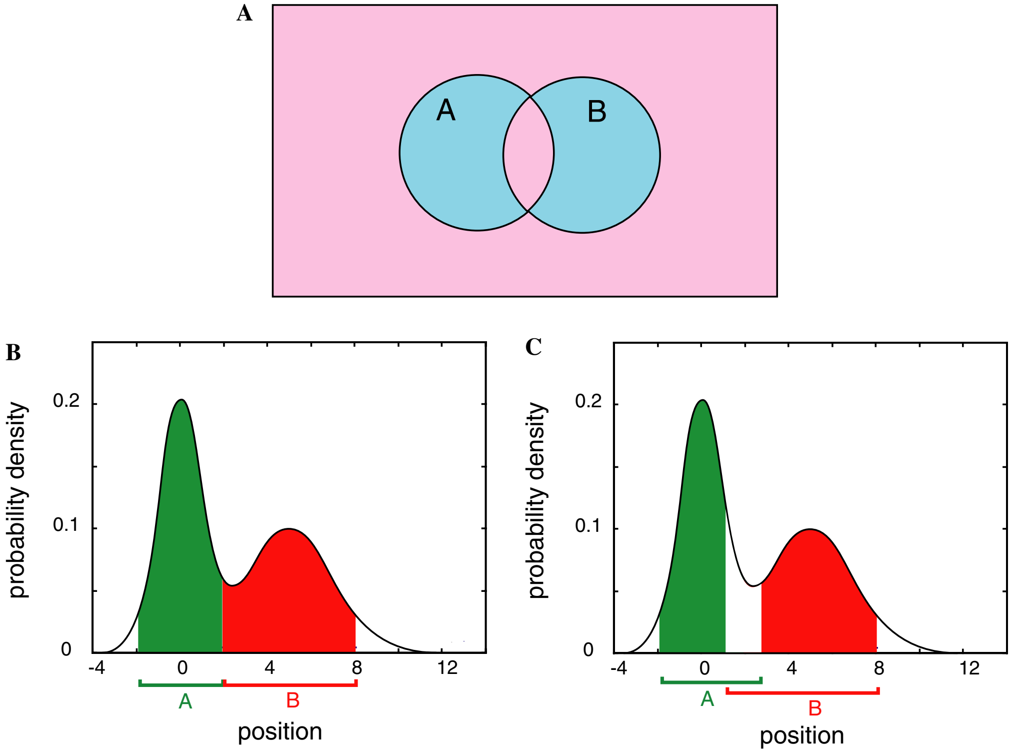

Fig. 2. The probability of symmetric difference (PSD) metric. (A) The symmetric difference, shaded blue, of events *A* and *B*. The symmetric difference of *A* and *B* is their union minus their intersection. (B) The PSD distance between events *A* and *B*. This distance is the green area above *A* plus the red area above *B*. The probability measure in this example is $(G_1 + G_2)/2$, where G_1 is a Gaussian with mean 0 and standard deviation 1, and G_2 is a Gaussian with mean 4 and standard deviation 2. (C) The PSD distance between events *A* and *B* that overlap. The region of overlap is not included in the symmetric difference, and so the PSD distance is less than in (B). (For interpretation of the references to colors in this figure legend, the reader is referred to the web version of this paper.)

blue-shaded region in Fig. 2A. Intuitively, it is the part of *A* that is outside of *B* plus the part of *B* that is outside of *A*. Then the distance, $d(A, B)$, between color experience *A* and color experience *B* is just the probability *p* of this symmetric difference (Blumenthal, 1970; Ferrer, 2003). That is, $d(A, B) = p(A \Delta B)$. This probability-of-symmetric difference (PSD) metric specifies the distances between all pairs of color experiences, not just a particular pair *A* and *B*, and therefore captures all empirically measurable similarities and differences among these experiences. According to functionalism the distances, *d*, between color experiences, but not the color experiences themselves, enter into functional architecture.

Given the probability *p* of Jack's color experiences *X* and given our scrambling function *b*, we can canonically transport *p* to Jill's color experiences *Y*. This transport, which is a probability measure *q* on *Y*, is simply the distribution of *b*. Recall that, if *D* is any event for Jill's color experiences *Y*, and if *C* is its corresponding unscrambled event for Jack's color experiences *X*, then the probability $q(D)$ is just $p(C)$.

We have canonically transported Jack's probabilities p of color experiences X to Jill's probabilities q of color experiences Y. Now, we can derive from q a new distance metric, d', on Jill's color experiences Y, just as we derived from p the distance metric, d, on Jack's color experiences X. And here is the striking result:

> For *every* choice of probability, p, and *every* choice of scrambling, b, Jack's distance metric, d, and Jill's distance metric, d', *always* agree.

A proof of this "Probabilistic Scrambling Theorem" is given in Appendix A.

The Probabilistic Scrambling Theorem entails that, if A' and B' are any two sensory experiences for Jill and if A and B are the corresponding unscrambled sensory experiences for Jack, then the distance between A' and B' for Jill is identical to the distance between A and B for Jack. Thus, all perceptual experiments involving Jill evince the same results as for Jack. A simple illustration of this is shown in Fig. 3A, where the color experiences of Jill are scrambled relative to those of Jack, but all distances between corresponding experiences remain unchanged. This example uses no inputs from the world and no outputs from Jack or Jill; scrambling is possible nonetheless.

Fig. 3B extends this example by adding inputs and outputs, and illustrates that Jack and Jill will apply the same color names to the same objects, so that one cannot tell by talking with them that

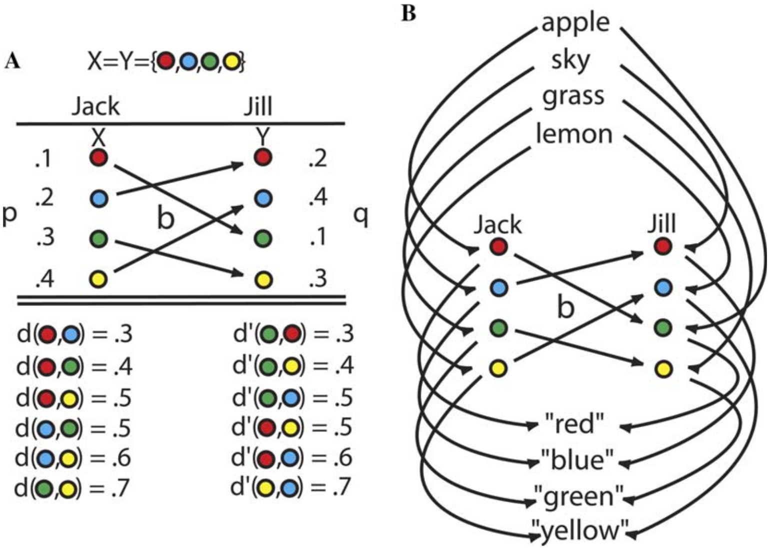

Fig. 3. Simple examples of the Probabilistic Scrambling Theorem. (A) A scrambling of experiences preserves all distances. The arrows define the scrambling and transport the probability p to its distribution q. Since the four basic color events are disjoint, the measure of the symmetric difference, and therefore of the distance, between any two of these color events is just the sum of their measures. One can easily verify that the distances d and d' between corresponding color events remain unchanged. (B) Objects, subjective experience, and language. Although Jack and Jill differ in their color experiences, they still use the same color names to describe objects. An apple leads to one color experience for Jack, as indicated by its arrow on the left, and to another color experience for Jill, as indicated by its arrow on the right. But both Jack and Jill call their color experience "red," as indicated by their arrows to the color words. Mathematically, one says that the diagram commutes. (For interpretation of the references to colors in this figure legend, the reader is referred to the web version of this paper.)

their experiences differ. This example uses language, but the same result applies to nonlinguistic experiments. Thus, the color experiences of Jack and Jill could be differently connected to the external world without any experimental divergences to betray that phenomenal difference. If Jack makes finer discriminations among stimuli in the region of color space called "green" than in the region called "blue," so also will Jill, even if Jill's color experience upon viewing grass is the same as Jack's upon viewing the sky. The nonuniformity of color space (Mausfeld & Heyer, 2003) or of any other phenomenal space is no obstacle to application of the Probabilistic Scrambling Theorem. But if we suppose instead that Jack is red-green color blind and Jill is not, then of course experiments would reveal differences between them, with no attendant violation of the Probabilistic Scrambling Theorem.

The Probabilistic Scrambling Theorem models subjective similarities between conscious experiences with the PSD metric, a metric that is derived solely from probabilities governing those experiences and, as described in Appendix A, that generalizes standard measures of perceptual discrimination from Signal Detection Theory (Green & Swets, 1988; Wickens, 2002). This metric, however, is not required for the Probabilistic Scrambling Theorem. The Theorem holds if, instead of probabilities of symmetric differences, one uses probabilities of *any* measurable function of the relevant events, or of their unions, intersections, and differences.

7. Conclusion

We have proved the logical possibility of Locke's suggestion that color experiences might vary from person to person, with no measurable differences. The proof applies not just to color experiences but to all experiences in all sensory modalities. This result proves that reductive functionalism is false: conscious experiences are not identical to functional relations. Conscious experiences do not supervene logically on functional relations. This still leaves open the logical possibility that nonreductive functionalism might be true; however, since there is, as yet, no scientific theory of consciousness based on nonreductive functionalism, we do not yet know *precisely* what might be true.

Acknowledgments

I thank Jeff Barrett, Mike Braunstein, Ann Carruthers, Dave Chalmers, Charlie Chubb, Paul Churchland, Stuart Cracraft, Dan Dennett, Jason Ford, Ron Frostig, Richard Gregory, Duncan Luce, Rainer Mausfeld, Louis Narens, Alan Nelson, Steve Palmer, Barron Ramos, Whitman Richards, Bob Schwartz, Terry Sejnowski, and Manish Singh, for helpful comments on earlier drafts. I thank Johannes Andres, Bill Batchelder, Ann Carruthers, Juergen Golz, Geoff Iverson, Louis Narens, and Reinhard Niederee, for helpful discussions. In particular, Johannes Andres and Reinhard Niederee pointed out to me, during a talk with Rainer Mausfeld's color group in Kiel, that the Probabilistic Scrambling Theorem could be generalized to the Scrambling Theorem. This work was supported by a grant from US National Science Foundation.

Appendix A

Probabilistic Scrambling Theorem. *Let (X, ξ) and (Y, ζ) be measurable spaces, b: $X \rightarrow Y$ a measurable function, p a probability measure on (X, ξ), and d the metric on ξ induced by p defined, for all $A, B \in \xi$, by $d(A,B) = p(A\Delta B)$, where Δ denotes symmetric difference. Then, the probability measure q on (Y, ζ) defined, for all $E \in \zeta$, by $q(E) = p(b^{-1}(E))$, induces a metric d' on ζ satisfying, for all $E, F \in \zeta$, $d'(E, F) = d(b^{-1}(E), b^{-1}(F))$.*

Proof. $d'(E, F) = q(E\Delta F) = p(b^{-1}(E\Delta F)) = p(b^{-1}(E)\,\Delta b^{-1}(F)) = d(b^{-1}(E), b^{-1}(F))$. □

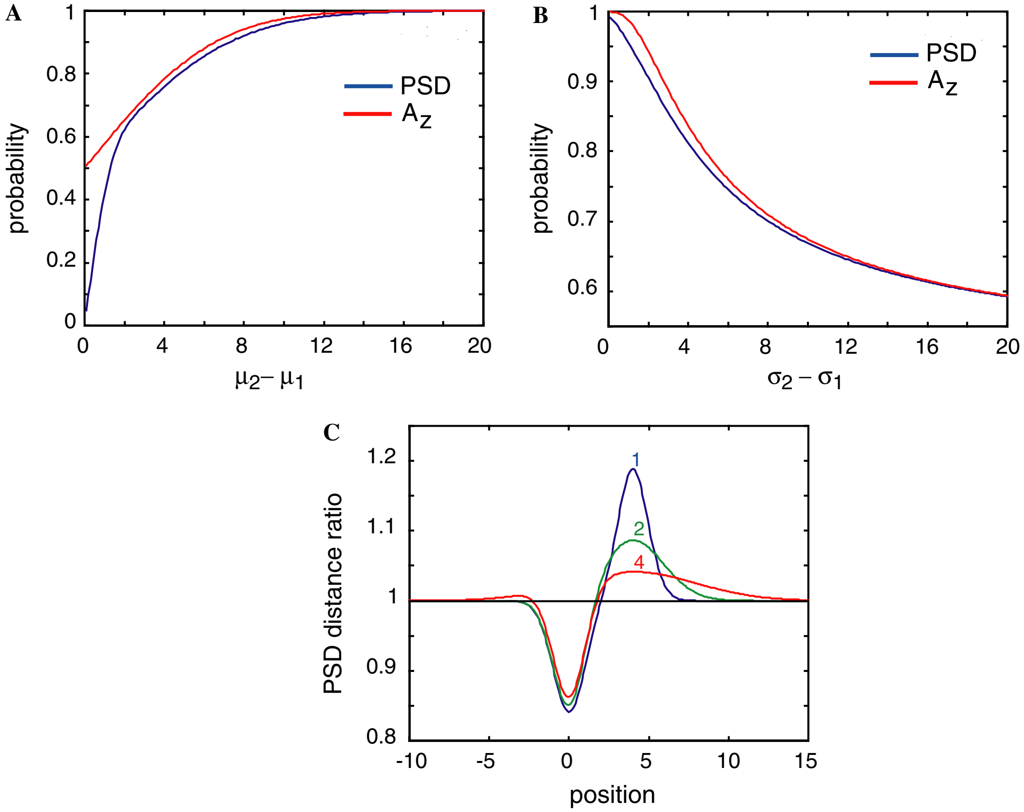

Fig. 4. Plots of the PSD metric and the function A_z of Signal Detection Theory. (A) PSD and A_z for two Gaussians with standard deviations 1 and 5, as the difference in their means varies from 0 to 20. PSD and A_z are related monotonically, but PSD has greater sensitivity than A_z, where sensitivity is needed most, viz, for small differences between the mean values. (B) PSD and A_z for two Gaussians with mean values 0 and 5 as the difference in their standard deviations varies from 0 to 20. Again, PSD and A_z are related monotonically but PSD has greater sensitivity than A_z for small differences between the standard deviations. (C) Ratio of PSD distances. For each position of an event, C, of width 0.2 we plot the ratio $G_1\,(R\Delta C)$/ $G_2(R\Delta C)$, where R denotes the real line. G_1 is always a Gaussian with mean 0 and standard deviation 1. G_2 is a Gaussian with mean 4 and standard deviation 1 for the blue curve, standard deviation 2 for the green curve, and standard deviation 4 for the red curve. On the far left of the plot, the red curve lies above the value 1, indicating that in the PSD metric the event C in this region is closer to the Gaussian with mean 4 than to the Gaussian with mean 0. (For interpretation of the references to colors in this figure legend, the reader is referred to the web version of this paper.)

If b is bimeasurable, then the Probabilistic Scrambling Theorem also holds in the other direction, i.e., for all $G, H \in \xi$, $d(G, H) = d'(b(G), b(H))$. The Theorem also holds if, instead of the PSD metric, one uses the probability of *any* measurable function of the relevant sets. For instance, it holds for the quasi-pseudometric (Ferrer, 2003) $d^*(A, B) = p(B \backslash A)$, where $p(B \backslash A) + p(A \backslash B) = p(A \Delta B)$; d^*, like some psychological judgments of similarity (Tversky, 1977), is not symmetric. The Theorem assumes that ξ and ζ are σ-algebras, but holds more generally. If, for instance, ξ and ζ are closed under *disjoint* union then the metric, where defined, is preserved under scrambling. This more general formulation permits incomparable pairs of subjective experiences, for which there is no distance: What is the distance between the smell of garlic and the sound of a flute?

The PSD metric used in the Probabilistic Scrambling Theorem generalizes the function A_z, the area under a receiver-operating characteristic curve, used in Signal Detection Theory (Green & Swets, 1988; Wickens, 2002). A_z describes an observer's ability to discriminate between two signals with Gaussian distributions G_1 and G_2 on the measurable real line (R, ξ). Using the PSD metric, one obtains this special case by letting $p = (G_1 + G_2)/2$, and choosing two events, A and $B \in \xi$, satisfying: (1) event A is radially symmetric about the mean of G_1 and B about the mean of G_2, and (2) A and B together maximize $p(A \Delta B)$. An example is shown in Fig. 2B. This special case of the PSD metric varies monotonically with A_z as the mean values and variances of the gaussians vary, as shown in Figs. 4A and B. If an observation results in an event C, and one must decide between G_1 and G_2 given C, then an unbiased decision criterion is the value 1 for the ratio of the distances from C to G_1 and G_2, viz, the ratio $G_1(R \Delta C)/G_2(R \Delta C)$, as shown in Fig. 4C.

The PSD metric is preserved, up to a global scale factor, by the entailment relation of the Lebesgue logic on probability measures, in which p entails q if p is a normalized restriction of q (Bennett, Hoffman, & Prakash, 1993). Perceptual inferences of observers are morphisms of the Lebesgue logic (Bennett et al., 1989, 1993) and therefore respect the PSD metric.

References

Ash, R., & Doleans-Dade, C. (2000). *Probability and measure theory*. San Diego: Academic Press.

Bennett, B., Hoffman, D., & Prakash, C. (1989). *Observer mechanics: A formal theory of perception*. San Diego: Academic Press.

Bennett, B., Hoffman, D., & Prakash, C. (1993). Lebesgue logic for probabilistic reasoning and some applications to perception. *Journal of Mathematical Psychology, 37*, 63–103.

Bickle, J. (2003). *Philosophy and neuroscience: A ruthlessly reductive account*. Dordrecht: Kluwer Academic Publishers.

Block, N., & Fodor, J. (1972). What psychological states are not. *Philosophical Review, 81*, 159–181.

Blumenthal, L. (1970). *Studies in geometry*. Freeman: San Francisco.

Braddon-Mitchel, D., & Jackson, F. (1996). *Philosophy of mind and cognition*. Blackwell: Oxford.

Chalmers, D. (1995). Absent qualia, fading qualia, dancing qualia. In T. Metzinger (Ed.), *Conscious experience* (pp. 309–328). Exeter, UK: Imprint Academic.

Chalmers, D. (1996). *The conscious mind*. Oxford University Press: Oxford.

Chalmers, D. (Ed.). (2002). *Philosophy of mind: Classical and contemporary readings*. Oxford University Press: Oxford.

Churchland, P. S. (2002). *Brain-wise: Studies in neurophilosophy*. Cambridge, MA: MIT Press.

Churchland, P. M. (1996). The rediscovery of light. *Journal of Philosophy, 93*, 211–228.

Clark, A. (1983). Spectrum inversion and the color solid. *Southern Journal of Philosophy, 22*, 431–443.

Crick, F., & Koch, C. (1998). Consciousness and neuroscience. *Cerebral Cortex, 8*, 97–107.

Dennett, D. (1998). Instead of qualia. In Brainchildren (pp. 141–152). Cambridge, MA: MIT Press.

Ferrer, J. (2003). Quasi-pseudometric properties of the Nikodym-Saks space. *Applied General Topology, 4*, 243–253.

Fine, T. (1973). *Theories of probability: An examination of foundations*. London: Academic Press.

Green, D., & Swets, J. (1988). *Signal detection theory and psychophysics*. Los Altos, CA: Peninsula Publishing.

Gregory, R. (1998). Brainy mind. *British Medical Journal, 317*, 1693–1695.

Harman, G. (1996). Qualia and color concepts. In E. Villenueva (Ed.). *Philosophical issues* (7). Northridge, CA: Ridgeview.

Knill, D., & Richards, W. (1996). *Perception as Bayesian inference*. Cambridge: Cambridge University Press.

Locke, J. (1690/1979). *An Essay Concerning Human Understanding*. Oxford: Oxford University Press.

Mausfeld, R., & Heyer, D. (Eds.). (2003). *Color perception: Mind and the physical world*. Oxford: Oxford University Press.

Metzinger, T. (Ed.). (2000). *Neural correlates of consciousness: Empirical and conceptual questions*. Cambridge, MA: MIT Press.

Palmer, S. (1999). Color, consciousness, and the isomorphism constraint. *Behavioral and Brain Sciences, 22*, 923–989.

Shoemaker, S. (2000). Phenomenal character revisited. *Philosophy and Phenomenological Research, 60*, 465–467.

Tversky, A (1977). Features of similarity. *Psychological Review, 84*, 327–352.

Tye, M. (1996). *Ten problems of consciousness*. Cambridge, MA: MIT Press.

Tye, M. (2000). *Consciousness, color, and content*. Cambridge, MA: MIT Press.

Wickens, T. (2002). *Elementary signal detection theory*. Oxford: Oxford University Press.