Synthese (2011) 180:443–463
DOI 10.1007/s11229-010-9712-8

Supervenience and neuroscience

Pete Mandik

Received: 5 June 2008 / Accepted: 7 January 2010 / Published online: 27 January 2010

Abstract The philosophical technical term "supervenience" is frequently used in the philosophy of mind as a concise way of characterizing the core idea of physicalism in a manner that is neutral with respect to debates between reductive physicalists and nonreductive physicalists. I argue against this alleged neutrality and side with reductive physicalists. I am especially interested here in debates between psychoneural reductionists and nonreductive functionalist physicalists. Central to my arguments will be considerations concerning how best to articulate the spirit of the idea of supervenience. I argue for a version of supervenience, "fine-grained supervenience," which is the claim that if, at a given time, a single entity instantiates two distinct mental properties, it must do so in virtue of instantiating two distinct physical properties. I argue further that despite initial appearances to the contrary, such a construal of supervenience can be embraced only by reductive physicalists.

Keywords Supervenience · Physicalism · Neuroscience · Reductionism

1 Introducing fine-grained supervenience

The philosophical technical term "supervenience" is frequently used in the philosophy of mind as a concise way of capturing a, if not *the*, core idea of physicalism, which, if stated as a slogan, is the idea that there are no mental differences without physical differences.[1] Since most physicalists aren't physicalists just about the mind, the slogan really is that there are *no* differences without physical differences. One

[1] For an excellent review of key notions and applications of supervenience, see McLaughlin and Bennett (2006). See Wilson (2005) for criticisms of the adequacy of supervenience for formulating physicalism.

P. Mandik (✉)
William Paterson University, Wayne, NJ, USA
e-mail: mandikp@wpunj.edu

feature of supervenience that explains much of its appeal is that it seems to offer a way of spelling out physicalism that is neutral with respect to reductive and nonreductive varieties of physicalism.

It is a major goal of this paper to offer novel arguments against this alleged neutrality. I will be arguing for reductive physicalism. Further, I'll be taking sides in the version of the debate between reductive and nonreductive materialists as that debate is played out between psychoneural reductionists or type-identity theorists on the one hand and nonreductively inclined functionalists motivated by multiple-realization considerations on the other. Given this, then, I will have little, if anything, to say that explicitly addresses versions of the debate that hinge on other motivations for nonreductionism—motivations such as those concerning the alleged normativity of mental-state ascriptions. It is my intention, then, to argue for the reduction of the mental to the neural and against the functionalist multiple realizability of the mental by the physical. Central to my arguments will be considerations concerning how best to articulate the spirit of the idea of supervenience.

The most explicit early statement directly relevant to the philosophy of mind of what supervenience consists in is due to Davidson (1970, p. 88). Davidson relayed the idea of saying that the mental supervenes on the physical in a way that may be paraphrased as the following pair of propositions:

(1) No two entities can differ at a time with respect to their mental properties without differing at that time with respect to their physical properties.
(2) No single entity can change with respect to its mental properties without changing with respect to its physical properties.

As formulated, (1) and (2) do not make sufficiently explicit the idea that, more than being correlated with certain physical properties, mental properties are had *in virtue of* having certain physical properties. That Davidson had such a determination thesis in mind is made clear in a later work wherein he explicates the mental being supervenient on the physical by writing that "the physical characteristics of an event (or object or state *determine* the psychological characteristics" (Davidson 1973, pp. 716–717, emphasis in original).[2] It is better, then, to formulate Davidsonian supervenience as *this* pair of propositions:

(1) If, at a given time, two entities instantiate two distinct mental properties, they must do so in virtue of instantiating two distinct physical properties.
(2) If, at two distinct times, a given entity instantiates two distinct mental properties, it must do so in virtue of instantiating two distinct physical properties at those times.[3]

[2] I will have relatively little to say about how "in virtue of" should be spelled out, but it will suffice for the present discussion to think of "in virtue of" as signaling a kind of noncausal determination of ψ by ϕ that is consistent with, but does not entail, that ψ and ϕ are one and the same.

[3] It is perhaps worth stressing here that what is relevant about Davidson to the current project is not the particular considerations that motivated the nonreductive portion of his nonreductive physicalism (considerations having to do with the alleged normativity of the mental, etc.). What is relevant is his influential supervenience-based explication of the physicalist portion of his nonreductive physicalism. Such an explication of physicalism has been embraced as consistent with a variety of motivations for nonreductivism, varieties such as functionalist multiple realizability considerations.

Subsequent writers offered formulations of supervenience that varied either in terms of their modal force (e.g., logical, nomological, etc.) or in terms of what entities were under consideration (e.g., persons, space-time regions, entire worlds, etc.).[4] Such variations are of little interest here. What *is* of interest is a formulation of the core idea of physicalism—the "no differences…" slogan—overlooked by Davidson and subsequent authors:

(3) If, at a given time, a single entity instantiates two distinct mental properties, it must do so in virtue of instantiating two distinct physical properties.[5]

Since this formulation will be the focus of the current paper, I prefer a more informative label and will use "fine-grained supervenience" or FGS for short. In case it has gone unnoticed, the key difference between FGS on the one hand and (1) and (2) on the other is that where (1) concerns multiple entities at a single time and (2) concerns a single entity at multiple times, FGS concerns a single entity at a single time. I think it should be fairly obvious that FGS is a distinct yet reasonable way of cashing out the "no differences …" slogan, but I will spell this out further anyway.[6]

One thing that isn't obvious, but will be a major aim of this paper to argue for, is that FGS leads to reductive physicalism. Sections 2–4 are dedicated to spelling out arguments from FGS to reductive physicalism. Sections 2 and 3 articulate thought experiments concerning how certain functionalist nonreductive considerations give rise to highly counterintuitive scenarios of what I shall call "mental–mental supervenience"—scenarios that physicalists will want to rule out and will need to embrace reductive physicalism to do so. Where the line of argument developed in Sects. 2 and 3 specifically targets functionalist varieties of nonreductivism, the argument in Sect. 4

[4] For a recent and brief review of these variations, see Lynch and Glasgow (2003, p. 202). Variations concerning modal force are perhaps more directly relevant to clarifying the "in virtue of" locution than variations concerning the relata of the supervenience relation. While I'm largely bypassing such issues, I will note that my interest in insisting on focusing on Davidson's 1973 "determination" formulation instead of the 1970 formulation, is that a formulation of fine-grained supervenience that follows the pattern of the 1970 wording seems like a relatively uninteresting thesis. It would be the thesis that a thing that has two mental properties at a time must have two physical properties at that time. Few things have fewer than two physical properties. Especially any things complicated enough to have any mental properties. It seems on the face of it a more interesting claim that a thing that has two mental properties at a time must have two different physical properties in virtue of which the two mental properties are had.

[5] Hereafter, reference to properties in this paper is limited to determinate properties as opposed to merely determinable properties. Arguably, without such a restriction, a determinable property and one of its determinates can share a supervenience base and thus constitute a counterexample to this formulation.

[6] Note how different this is from older formulations such as those discussed by Hofweber (2005, pp. 6–7) or this formulation of "strong" supervenience discussed by Wilson (2005, p. 433), wherein it is

> formulated as holding between families of properties A and B, elements of which are co-instantiated in individuals in a domain D:
> A strongly supervenes on B iff $\Box(\forall x \in D)(\forall a \in A)(x \text{ has } a \rightarrow (\exists b \in B)(x \text{ has } b \wedge \Box(\forall y \in D)(y \text{ has } b \rightarrow y \text{ has } a)))$.

FGS is not entailed by formulations such as strong supervenience. Instead, strong supervenience is compatible with the falsity of FGS. We might state this compatibility in the following way: Whereas strong supervenience is compatible with multiple realizability insofar as there might be a physical property, b^*, other than b that suffices for a, strong supervenience is compatible with the falsity of FGS insofar as there might be some mental property, a^*, other than a that b suffices for.

develops a regress argument that targets all versions of nonreductive physicalism. After Sect. 4, I turn from spelling out the entailments of FGS to consider, in Sects. 5 and 6, reasons for embracing FGS. Finally, in Sect. 7, I will address why, if one is convinced of reductive physicalism, one should believe in the reduction of the mental to the neural.

2 Some things qualia cannot do

Regardless of whether one is a reductive or a nonreductive physicalist, there are certain things about qualia that one must regard as impossible. If you have qualia, and someone else is your physical doppelgänger, then it is impossible for them to have any of the following: (a) inverted qualia, (b) absent qualia, (c) fading qualia, and (d) dancing qualia.[7] While readers will likely already be familiar with these impossibilities, I'll spell them out a bit more and also spell out what is supposed to make them impossible. This will help convey the gist of how supervenience is useful in formulating physicalist attitudes about what can and cannot happen with qualia. This will also be useful for setting the stage for the arguments for FGS to come.

2.1 Inverted qualia

Seeing red things involves having a certain kind of qualia. Let's call them "red qualia" for short and remain neutral on whether red qualia are so called due to their literal redness or their relations to literal redness. Literal redness is literally the opposite of literal greenness. If we arranged hues in such a way that adjacency reflects similarity, we'd wind up with a circle and a 180° rotation—an inversion—of that circle would put red where green was and vice versa. We could presumably do the same thing with your physical doppelgänger's qualia and, so to speak, put her red qualia where your green qualia are and vice versa. However, we could do this only if physicalism was false. If physicalism is true, then we can't. And here supervenience articulates the reason why: (1) says that no two entities can differ mentally without differing physically. This rules out (a) because it involves two entities differing mentally without differing physically.

2.2 Absent qualia

If physicalism were false, then your physical doppelgänger could be utterly devoid of qualia—it could be a philosopher's zombie—in spite of being just like you with respect to how many and what kind of particles it is made of and how those particles are arranged and move through space and time. But physicalists do not tolerate absent qualia for the same reason they do not tolerate inverted qualia: version (1) of

[7] See Chalmers (1996, pp. 247–275) for an extended discussion of the sorts of things, such as dancing, that physicalists forbid qualia to do.

supervenience prohibits entities differing in a mental respect without differing in a physical respect.

2.3 Fading qualia

Fading qualia are like absent qualia, but it takes a different time, not a different person, to get them. If you were to remain the same physically but you gradually changed from having qualia to not having qualia, then you would have fading qualia. Such a possibility is excluded by physicalism, and here the exclusion is due to clause (2) of supervenience. Since an entity cannot change mentally without changing physically, and "fading qualia" is shorthand for situations where qualia change without physical changes, fading qualia are impossible.

2.4 Dancing qualia

Dancing qualia are rapid intrasubjective qualia inversions and reversions. Like inverted qualia, there is a 180° re-mapping. Like fading qualia, the difference is across times, not persons. Like fading qualia, the impossibility of dancing qualia is entailed by clause (2) of supervenience.

3 More things qualia cannot do and trouble for functionalism

We have not yet begun to consider arguments against nonreductive physicalism. Nor have we put FGS to work. Let us change all that. To get this change rolling, let us consider a new thing that may be impossible for qualia: doubled qualia.

Doubled qualia occur when two minds, one whose qualia are inverted with respect to the other—a "green mind" and a "red mind," respectively—share a supervenience base. What it means to share a supervenience base is that there are no physical differences in virtue of which the red mind and the green mind differ.

Davidsonian supervenience, the conjunction of (1) and (2), does not rule out doubled qualia. To appreciate this failure of Davidsonian supervenience, it will help to attempt to imagine doubled qualia. To imagine doubled qualia, begin by imagining someone other than you, call him or her "Person X", who is *not* your physical doppelgänger. Let us stipulate that there is *some* physical difference between you and Person X. That is, there are two minds inside of Person X—a red mind and a green mind—whereas you only have the standard-issue red mind. Suppose further that all and only the physical properties that give rise to X's red mind are the *same* physical properties that give rise to X's green mind. Suppose, as well, that the physical properties that give rise to X's red mind are different from the physical properties that give rise to your red mind. If you are having a hard time imagining this, it is likely due to your tacit or explicit acceptance of FGS. Your acceptance of (1) and (2) can't explain this. In particular, the difference between you and X—the fact that you have only a red mind whereas X has a green one as well as a red one—is fully consistent with (1), since we have stipulated that there is a physical difference between you and Person X.

Those physicalists who find doubled qualia to be weird are likely to find the following intolerably bizarre: intermittently doubled qualia. To get warmed up for intermittently doubled qualia, stop to appreciate the following stipulation about Person X: the red mind and the green mind need not have any awareness of each other whatsoever. What this means, then, is that for all you know, your qualia are doubled right now. For all you know, there's someone else "in there" with you right now and what it's like to be them is just like what it is like to be you except for the inversion-of-the-color-qualia bit. Now, imagine further that you undergo the following recurring physical change: The set of physical properties *P*1, which suffice to instantiate just the red mind are replaced by a set of distinct physical properties *P*2, which suffice for both a red mind and a green mind. So, according to the intermittently doubled qualia thought experiment, you change from *P*1 to *P*2 and back again, which changes you from undoubled qualia to doubled qualia and back again. And all of this happens without you—the red mind—even noticing.

Perhaps you think that intermittently doubled qualia are impossible. You may be right. But here's one thing that will not rule them out: clause (2) of supervenience. Clause (2) prohibits you from changing mentally without changing physically, but the intermittently doubled qualia thought experiment stipulated that there were physical changes accompanying the change to a doubled state.

Physicalists should agree that doubled qualia are absurd. They should likewise agree on the absurdity of intermittently doubled qualia. (Hereafter I will use "doubled qualia" as shorthand for both the intermittent and nonintermittent varieties.) If you call yourself a "physicalist" but you believe in a theory that entails the possibility of doubled qualia, then you have some serious problems.

To sum up the section so far: Davidsonian supervenience rules out only some of the obvious impossibilities concerning qualia. Clause (1) rules out (a) inverted and (b) absent qualia. Clause (2) rules out (c) fading qualia and (d) dancing qualia. Neither (1) nor (2) rules out doubled qualia. And I am betting physicalists would very much like to rule out doubled qualia. If so, then physicalists should explicitly embrace FGS. They probably implicitly already have, but more on this in Sects. 5 and 6.

One thing that one might say about doubled qualia is that no one really needs to worry about them since, even though no theory rules them out, no theory entails them. However, I think maybe there are theories that entail the possibility of doubled qualia. I think perhaps functionalism does. This is interesting because many, if not most, nonreductive physicalists are functionalists.

Suppose physicalism, P, just is (1)&(2)&FGS. Suppose also that functionalism, F, entails doubled qualia. Doubled qualia contradict P, especially the FGS part of P. Physicalist functionalists adhere to the conjunction F&P. If F entails doubled qualia, then that would be a reductio of F&P. In the face of such a reductio, one must consider either giving up F or giving up P.

Before getting into the account of how functionalism leads to doubled qualia, some brief comments are in order about what functionalism is supposed to be.[8] Functionalism has a positive part and a negative part. The positive part is what makes a mental

[8] Classic formulations are to be found, of course, in Putnam (1967) and Fodor (1974).

state the mental state that it is. The negative part is what *doesn't* matter to making a mental state the mental state that it is. The positive part says that essential to the instantiation of some mental state are the causal relations that state bears to other states. (Different varieties differ on further requirements for those causal interactions, e.g., whether they must include interactions to states outside of the body or whether they must be describable as implementing computations, but little hinges on these variations here.) The negative part is an insistence upon the psychological irrelevance of certain lower-levels of physical organization. (Different varieties differ on how low one must go to find the irrelevant level—e.g., cells or chemicals—but little hinges on these variations for now.)

One distinctive and widely known feature of functionalism is its commitment to multiple realizability, the idea that two entities can differ in their low-level physical properties while those distinct low-level physical properties suffice for—that is, realize—instances of the same mental property. It's the negative part of functionalism that leads to multiple realizability, and it is multiple realizability that makes functionalism fit so well with nonreductivism. If Jones and Smith can have a mental property in common while having no physical properties in common, then that mental property cannot reduce to—that is, cannot be identical to—either of those physical properties. It is also the negative part of functionalism that leads to doubled qualia. Multiple realizability may be construed as allowing for mental *realizers* as well as mental *realizeds*. I turn now to spell out further how functionalism leads to either doubled qualia or something very close to doubled qualia that I will call "mental–mental supervenience."

There is a familiar class of alleged counterexamples to functionalism, instances of which are Searle (1980) Chinese Room and Block (1978) Chinese Nation. The way such counterexamples are supposed to work is by coming up with hypothetical instances in which the functionalist causal-relational criteria for mentality are satisfied by a system that nonetheless does not instantiate mentality. Functionalists correctly challenged whether these were obvious counterexamples instead of just obvious instances of begging the question against functionalism. I don't think extant fans of the counterexamples gave compelling defenses of them. I think, however, that their status as counterexamples is defensible.

First, let's spell out in further detail the counterexamples in question, starting with Searle's Chinese Room. Imagine that functionalists have codified the causal interactions between system states alleged to suffice for that system to count as understanding Chinese. Such a codification can be expressed as a program or set of instructions that, if running on a computer, would enable it to thereby implement an artificial intelligence capable of understanding Chinese. A person can follow these instructions, or instructions isomorphic to them. A person, then, can execute the program. In Searle's Chinese Room thought experiment, we are to imagine this program being written in English and executed by a non-Chinese-speaking English speaker. Searle offers that we imagine *him* playing this role, sitting in a room in which Chinese queries written on cards come in through the "in" slot and appropriate Chinese responses go out the "out" slot. Crucially, Searle claims that *he* can follow the English instructions that would result in appropriate Chinese responses to Chinese queries without understanding Chinese himself. Searle's argument, in brief, against functionalism is that running the program—implementing the causal interactions specified by the program—cannot

suffice for understanding Chinese, since Searle can run the program without understanding Chinese.

One functionalist response—the "systems response"—is that while Searle may not understand Chinese, Searle alone is not instantiating the requisite causal interactions. Instead, the causal interactions that suffice for understanding Chinese include more than those inside of Searle. They also include the causal interactions with the cards coming in and out the slots as well as with the instruction manual wherein the program is written. In short, the objection is that Searle comprises only a part of the total system responsible for implementing understanding of Chinese. Searle counters that the cards and manual are inessential and that the crucial features of the thought experiment can be preserved in a version wherein Searle memorizes the instructions. Chinese queries are put to him verbally and he utters appropriate responses all without actually understanding Chinese.

At this point in the dialectic we have something very close to doubled qualia—an instance of what I call "mental–mental supervenience." Consider what bullet-biting functionalists must say to Searle at this point: They must say that Searle's mental activities, which do not themselves *constitute* understanding of Chinese, nonetheless *give rise to* a second mind that does understand Chinese. In other words, the Chinese-understanding mind supervenes on Searle's monolingual English-understanding mind, which in turn supervenes on Searle's brain.

I want to emphasize here that the scenario that I am describing as an instance of mental–mental supervenience and thus a violation of FGS is *not* the scenario as initially described in the systems reply—call this "the whole-room scenario." It is instead the scenario described by Searle in response to the systems reply—call this "the internalized-room scenario." The whole-room scenario looks to be easily described as consistent with FGS and thus not an instance of mental–mental supervenience. The whole-room scenario may plausibly be described as providing two different physical supervenience bases for the Chinese-understanding mind and Searle's monolingual English mind. The Chinese-understanding mind supervenes on a larger physical system than does Searle's—a physical system that includes various contents of the room, the cards, and Searle himself. In contrast, Searle's own mind plausibly supervenes on only a proper part of that larger physical system, a proper part likely restricted to just Searle's own brain. (For simplicity's sake I here assume a kind of internalism about Searle's mind.) Turning from the whole-room scenario to the internalized-room scenario, it looks like there is nothing external to Searle's brain to figure in a distinct physical supervenience base for the Chinese-understanding mind. Searle has here memorized the rules and manipulates the symbols involved for following the program in his mind. The Chinese-understanding mental states supervene on Searle's rule-following mental states, which in turn supervene on Searle's brain states. Thus does the internalized-room scenario count as an example of mental–mental supervenience and a violation of FGS. Bullet-biting functionalists simply accept the possibility of the internalized-room scenario.

To make the example even closer to doubled qualia, we can note that there is nothing essential to the "understanding Chinese" stuff—we could have had the program be a simulation of seeing red and Searle be color-blind.

Let's turn to look at Block's Chinese Nation counterexample to functionalism. As in Searle's thought experiment, imagine that functionalists have codified a specification of the causal roles alleged to suffice for instantiating a mental state. Instead of imagining Searle himself following the rules, imagine the populous nation of China put to the task. Perhaps we imagine each individual Chinese citizen playing the role of a neuron (which would require it being *very* populous) and interneural communication replaced with intercitizen communication via walkie-talkie. Like Searle, this collection of individuals is imagined to implement a program. Like Searle, Block claims that the functionalists are mistaken in supposing that running the program suffices for mentality. Block alleges that no mental states are implemented aside from the mental states of the individual citizens. Supposing that the program was an English-understanding program, the claim is that non-English-speaking Chinese-speaking citizens could run the program without the individual citizens or any emergent "group-mind" thereby understanding English. Changing the example to be about qualia, we can stipulate that the program in question is a "seeing red" program and that the Chinese are all color-blind.

The example—seeing red as opposed to understanding language—matters little for this paper. What does matter is what certain functionalists say to such examples. They bite the bullet and affirm that the activity of the citizens suffices for the instantiation of a separate "group" mind—that a distinct solitary übermind arises out of the collective action of these individual unterminds.[9]

Does such a bullet-biting-functionalist response to the Chinese Nation constitute an affirmation of mental–mental-supervenience? Not yet, but with minor modifications it will. Before saying what the modifications are, let's first say why they are needed. Why they are needed is because the situation as described so far does not have the übermind and the collection of unterminds sharing a supervenience base. The unterminds, let us suppose, supervene on the brains of the citizens. The übermind, in contrast, supervenes on the brains plus the walkie-talkies. The addition of the walkie-talkies suffices to distinguish the übermind's supervenience base from the untermind-collection's supervenience base.

There are two ways of changing the example, though, to make it a genuine case of mental–mental supervenience. The first way is that we can, à la the Clark and Chalmers (1998) extended-mind thesis, stipulate that conditions are satisfied allowing that each citizen's mind "extends" beyond their skull to supervene on the walkie-talkies and much else besides. The second way is that we can get rid of the walkie-talkies and implement direct mind-to-mind (or brain-to-brain) communication between the citizens. On either option the color-blind untermind collective, which contains no red qualia, supervenes on exactly the same set of physical properties that the red-qualia-containing übermind supervenes on. And thus we have another case of mental–mental supervenience provided by the bullet-biting functionalist.

At this point a functionalist who sees where this is all going may attempt the following objection:

[9] See, for example, Lycan (1987, pp. 33–34) and Braddon-Mitchell and Jackson (1996, p. 106).

> The kinds of mental–mental supervenience that arise in the bullet-biting responses to Searle and Block arise *only* in those responses and, instead of biting bullets, we wish simply to dismiss such "counterexamples" as silly.

In response to the imagined objection I say the following: I don't think that functionalists can easily dismiss mental–mental supervenience as arising only in these outlandish scenarios. At least some functionalists are explicitly aware of how functionalist considerations lead to the view that group minds are a common occurrence. And one version of functionalist embrace of the widespread occurrence of group minds is the view—homuncular functionalism—that holds that each individual human mind is itself a group mind.

Homuncular functionalism (aka homunctionalism) as advocated by Lycan (1987) and Dennett (1978), also involves mental–mental supervenience. According to the homunctionalists, a human mind taken at the personal level is decomposable into a handful of subpersonal homunculi, each of which is decomposable into further homunculi. At each level of decomposition the units at that level have genuinely mental properties, but of a stupider, simpler sort than those found at the levels above. The recursive decomposition bottoms out with units so simple and stupid as to succumb to wholly mechanistic and nonpsychological explanations. But any two adjacent levels above the "bottom-out" level offer examples of mental–mental supervenience. Consider the personal level and the first subpersonal level of homuncular decomposition right below it. The personal level, which contains one mind, supervenes on the next level down, which consists of many interacting homuncular minds. The homuncular functionalist view of the mind is analogous to the bullet-biting functionalist construal of Block's Chinese Nation insofar as both hold that a high-level "group" mind can supervene on a multitude of lower-level minds. Thus, mental–mental supervenience is not restricted to bullet-biting functionalism. If homunctionalism is true, then it happens all of the time.

It is worth spelling out exactly how these functionalists got into this position and what a problematic place this is to be. The possibility of mental–mental supervenience follows from functionalism. According to functionalism, all that matters to the instantiation of a mental event is that certain causal relations obtain between parts of the realizing system. Functionalism allows, then, that (i) the parts of the realizing system can have their own mental properties and (ii) the causal relations between parts of that system can be instances of mental causation.

The possibility of mental–mental supervenience, however, poses a serious threat to theorists who subscribe to the conjunction of physicalism and functionalism, because the possibility of mental–mental supervenience leads to a *reductio ad absurdum* of that conjunction. The key to the *reductio* is the fact that the possibility of mental–mental supervenience contradicts physicalism, since the possibility of mental–mental supervenience is the possibility of mental differences obtaining without physical differences obtaining.

As mentioned previously, functionalism is just one kind of nonreductive physicalism. Perhaps nonreductive physicalists who are functionalists will react to the mental–mental supervenience arguments by abandoning functionalism in favor of some other nonreductive physicalism. The following argument seals off all escape routes. It is not

simply a *reductio* of physicalism and functionalism. It is a *reductio* of physicalism and nonreductionism.

4 The regress argument

Nonreductive physicalism about mental properties is nonreductive because it holds that no mental property is identical to any physical property. Call this the nonidentity thesis. It is physicalist because it holds that there can be no mental differences that obtain without obtaining in virtue of physical differences. Call this the supervenience thesis. In the present section I will work with a generalized supervenience thesis, a thesis that consists in a conjunction of the generalizations of (1), (2), and FGS. The generalized versions—I'll call them "(1*)," "(2*)," and FGS*—are generalized because they concern properties in general instead of a restricted focus on mental properties. They are as follows:

(1*) If, at a given time, two entities instantiate two distinct properties, they must do so in virtue of instantiating two distinct physical properties.

(2*) If, at two distinct times, a given entity instantiates two distinct properties, it must do so in virtue of instantiating two distinct physical properties at those times.

(FGS*) If, at a given time, a single entity instantiates two distinct properties, it must do so in virtue of instantiating two distinct physical properties.

In case anyone is wondering why physicalists should accept generalized versions of (1), (2), and FGS—versions that aren't just about mental properties, but about *all* properties—we can perhaps gesture toward considerations along the lines considered in connection with qualia in Sects. 2 and 3. If doubled qualia and intermittently doubled qualia strike physicalists as sufficiently repugnant to embrace FGS, then we might invite them to consider analogous situations involving, instead of the doubling of qualia, the doubling of moral, political, economic, or aesthetic properties. For example, contemplate how obviously inconsistent with physicalism would be a scenario of intermittently doubled *morality* whereby a single situation or agent switched from having one set of moral properties to two sets of disparate moral properties in concert with a minute physical change, such as the firing of a single neuron. I will not spell this out further here, but I hope it suffices to suggest that what makes doubled qualia repugnant to physicalists is the doubling and not anything peculiar to qualia.[10]

I turn now to the construction of the regress. Given (1*), (2*), and FGS*, if a and b differ with respect to the mental properties M_1 and M_2 that they instantiate, they must do so in virtue of instantiating some distinct physical properties P_1 and P_2. Here's where problems arise. The demands of supervenience don't require only that *mental* differences give rise to physical differences. *All* differences must give rise to physical differences. (That's what "no difference without a physical difference" means.) It

[10] An anonymous referee has made a suggestion about these doubling arguments, a suggestion that I want to here note as a welcome one. The suggestion is that these considerations depend for their intuitive force on realism about properties and qualia. A physicalist who also embraces such realisms is highly unlikely to want to embrace doubling.

follows, then, that if the thesis of nonidentity holds, there is sense to be made of the question of in what consists the difference between M_1 and P_1. If they are distinct, there must be some physical difference in virtue of which they differ. Call that physical difference P_3. P_3 also must differ from M_1, otherwise nonidentity is violated. P_3 must further differ from P_1, otherwise there would be no physical difference in virtue of which M_1 and P_1 differ. So the difference between M_1 and P_1 demands the positing of a distinct property, P_3.

Since P_3 is distinct from M_1, the question again arises of in what consists the distinction, and the answer will involve P_4, which by similar chains of reasoning will lead to P_5 and P_6 and so on. This is a bad thing. If we think of the relations between the supervenient and subvenient properties in terms of explanation, then where the regress is infinite the target *never* gets explained. Fans of nonidentity should give up on physicalism altogether. Fans of physicalism should embrace identity theory.[11]

It is worth spelling out that the construction of the regress does not depend on reifying "being different" as a property unto itself. The relevant notion of difference can be unequivocally spelled out in terms of distinctness—that is, failure of identity. Thus, the argument may be stated as follows:

Physicalism requires that *no* properties can be nonidentical without being instantiated in virtue of nonidentical physical properties. In the case of nonidentical mental properties M_1 and M_2, they are instantiated in virtue of P_1 and P_2. The thesis of the nonidentity of the mental to the physical that is essential to nonreductive physicalism requires that M_1 be identical to no physical property and thus be not identical to P_1. An application of physicalism (the conjunction of (1*), (2*), and FGS* that we can summarize with the "*no* properties can be nonidentical without…" slogan) to M_1 and P_1 requires the instantiation of P_3, which is itself nonidentical to P_1 and M_1.

Note here that I am not assuming the existence of P_3 to get the argument rolling. That would indeed be an assumption that reifies difference as a property unto itself. Instead I am showing how the unwelcome P_3 arises as a consequence of the conjunction of physicalism and nonreductivism.

Continuing, then, we see that just as the unwelcome P_3, arises, so does its unwelcome brethren P_4 through P_∞. This is because an application of physicalism to M_1 and P_3 requires, for the sorts of reasons already stated, the instantiation of P_4. It should be clear at this point how M_1 and P_4 require P_5 and so on to P_∞.

At this point the nonreductive physicalist may wish to resist the argument by resisting the reading offered here of all differences requiring physical differences. This would be to protect nonreductive physicalism by rejecting physicalism as I've characterized it here. However, a serious question arises, then, as to what's physicalistic about nonreductive physicalism.

To appreciate the problem, consider the following: Reductive physicalism and nonreductive physicalism are going to have to give different answers to questions such as

[11] Lynch and Glasgow (2003) run a similar regress argument concerning supervenience, but the target of their argument is "superdupervenience," a brand of nonreductive physicalism that affirms the physical explicability of the supervenience relation. The conclusion of their regress argument is that superdupervenience is impossible. What Lynch and Glasgow do not attempt is to argue as I have against generalized nonreductive physicalism.

"in virtue of what are two properties different?" The reductive physicalist will always be able to give the same answer regardless of the properties in question. However, depending on the properties, the nonreductivist will have to give different answers. The different answers will depend on whether the properties in question are (i) two different mental properties, (ii) two different physical properties, or (iii) a mental property and a physical property. Spelling this out further yields the following:

(i) In the mental–mental case, the two mental properties are nonidentical in virtue of being instantiated by physical properties that are nonidentical.
(ii) In the physical–physical case, the difference is due simply to the two physical properties being nonidentical.
(iii) In the mental–physical case, the difference is due simply to the mental property and the physical property being nonidentical.

If we ask the nonreductivist what's physicalistic about (i)–(iii), which is to ask what merits considering the differences in question to be physical differences, we only get satisfying answers for (i) and (ii). For (ii), what makes the difference in question a physical difference is that there are two properties that are nonidentical and both physical. For (i), what makes the difference in question a physical difference is parasitic on (ii). For (iii), the difference is simply that the properties are nonidentical. But since one of the properties in question is nonphysical, there seem to be no grounds for regarding the difference in question as a *physical* difference (since neither (i) nor (ii) can supply the grounds). Further, there seem to be no grounds for considering the position in question a version of physicalism. It is dualism.

It is time now to consider arguments for FGS. Perhaps we have pretty much already seen one, and it goes pretty much like the following: If FGS was not a part of physicalism, then doubled qualia would be compatible with physicalism. But doubled qualia are not compatible with physicalism. Therefore, FGS is a part of physicalism. This argument has the virtue of validity, but perhaps the second premise can be questioned. At least, no argument has been given for the second premise. I assume most physicalists will like the second premise, and perhaps this is argument enough. More can be said, of course. As I will discuss in Sect. 5, I think, for instance, that current practice in cognitive neuroscience is shot through with what looks like empirical support for and/or tacit acceptance of FGS. I also think that there are projects outside of cognitive neuroscience that are similarly FGS-friendly, a point I will discuss further in Sect. 6.

5 Reflections on neuroscience and a defense of fine-grained supervenience

On several occasions when I've talked to people about mental–mental supervenience and how it should be ruled out by physicalism, I've been confronted by something like the following response:

> Insofar as cognitive functions *are not* localized to specific regions of the brain, mental–mental supervenience happens all of the time and should be embraced by physicalists, not ruled out by the postulation of FGS.

I think this response is exactly wrong, although it will take some explaining to get the point across. The explaining will require a review of some cognitive neuroscience, with an emphasis on evidence for and models of two different ideas on how cognitive functions are related to the brain. These two different ideas often appear in the relevant literature under the headings of "localizationism" and "holism."[12] In brief, localizationism involves locating the neural bases for distinct cognitive functions in distinct regions of the brain. Holism allows that distinct cognitive functions can share brain regions. Localizationism is obviously consistent with FGS, since brain-region differences are obviously physical differences. However, as I'll argue in a bit, all of the evidence for holism is fully consistent with FGS as well.

My interest here is primarily in models of holism, for I will be keen on seeing whether evidence for holism constitutes evidence against FGS. Perhaps the most fruitful and popular way of modeling how distinct mental properties may share brain regions comes from connectionism. My interest here will be in a review of the main relevant ideas of connectionism to show that holism does *not* provide counterexamples to FGS.

Connectionism involves modeling cognition in terms of neural networks. A neural network is a set of units (the neurons) and the connections between them. Let us consider a relatively simple connectionist model. Each neuron can be in one of several states of activation. We might think of these as real-numbered values ranging from 0 to 1. What state of activation a neuron will be in at a given time can be influenced by the state of activations of other neurons that are connected to it. Values called "weights" may be assigned to the connections, determining how much influence the other neuron may have via that connection. What state of activation a neuron goes into at a given time is determined by that neuron's transition function, which takes into consideration the states of activation and weights of adjacent neurons. Thus, for example, a neuron's activation may be a sigmoidal function of the weighted sum of the states of activation of the neurons connected to it.

One typical kind of neural network—a three-layer feed-forward network—has input neurons, output neurons, interneurons (hidden units), and connections that allow for a flow of information in a single direction from input units to hidden units and hidden units to outputs. In a "massively connected" three-layer feed-forward net, every input unit is connected to every hidden unit and every hidden unit is connected to every output unit. The topology of the network—the number of units and connections between them—is set by hand. The *weights*, however, are set instead by an automatic procedure—a learning rule that optimizes the set of weights to allow the network to perform some particular task. One such example comes from a network constructed by Garrison Cottrell and his colleagues, as described by Churchland (1995).

The network has 4,096 input neurons configured as a 64×64 unit retina, with each retinal unit's activation coding increments of brightness. The hidden layer has 80 units. The output layer has 8 units: 1 unit codes for face vs. non-face, 1 for male, 1for female, and 5 units encode "names" for faces. The network was trained with a set of photographs of 11 different faces. Training involves starting with the network's weights

[12] For discussions of localizationism and holism, see Anderson (2007) and Mundale (2002).

initially set to random values. Initial performance is, as expected, quite poor. Application of the back-propagation learning rule involves measuring differences between the right answer and the actual given answer and then making small changes to the weights based on a measurement of degree of error. After thousands of trials that involve exposing the retinal units to bit-mapped photographs of faces, the network can be tested to see how well it recognizes stimuli as being faces, being female or male, etc. When tested on faces from the training set, performance is 100%. When tested on a test set of novel faces, the network performed about 10–15% less well than on the faces from the training set. Such performance is comparable to the performance of human subjects.

Let us turn now to consider what it might mean to attribute mental states to such a neural network.[13] There are two general kinds of mental states to consider: occurrent states and abeyant states. Occurrent states, like percepts, are events of relatively short duration. Abeyant states, in contrast, are relatively more long-term, such as long-term memory or stored knowledge. When we attribute such states to networks, occurrent states are implemented by patterns of activation and abeyant states are implemented by the connection weights (Churchland and Sejnowski 1992, p. 165; Haugeland 1991, p. 84).

If connectionist networks constitute counterexamples to FGS, then they will do so either in virtue of occurrent states or abeyant states. And it is relatively easy to see that occurrent psychological states implemented in connectionist networks will not pose any special problems for FGS. Distinct occurrent mental states are implemented by distinct patterns of neural activation. For example, the pattern of activations that constitutes the identification of a face as female is clearly distinct from the one for a male.

If there is to be a problem for FGS, it will be posed by abeyant states. If we look to the facial-recognition network, we can see how the problem might arise. The network's knowledge of male faces is distributed across the connection weights, likely the very same weights across which the knowledge of female faces is distributed. Is this, then, a counterexample to FGS? Are distinct mental states instantiated in virtue of all and only the same physical states? To see that we don't have a counterexample to FGS, we need to appreciate that the distinct physical properties that give rise to the distinct abeyant states are physical dispositional properties of the network. In the case of the knowledge of male faces, this is implemented in virtue of the physical disposition to activate the "male" output unit in response to male faces. The distinct abeyant state of knowledge of female faces is implemented by the distinct physical disposition to activate the "female" output unit in response to female faces.

There is nothing problematic or even unusual about distinct physical dispositions being distributed across the same spatial regions. Consider, for example, the distribution of a sugar cube's solubility and fragility throughout its volume. For another example, consider the gravitational and magnetic properties of a chunk of iron. Both are distributed throughout the chunk, yet both are distinct physical properties and both

[13] Of course, some people will be loath to attribute *any* mental states to such neural networks. However, such people are of little interest here, since my interest here is in considering possible counterexamples to FGS construed in terms of the instantiation of mental states by such neural networks.

may be regarded as dispositions—the disposition to behave one way in a magnetic field and the disposition to behave another way in a gravitational field.

Now, it is one thing for the bases of distinct dispositions to be equally spatially distributed and another thing for distinct dispositions to share a supervenience base, for the first thing, but not the second, is consistent with fine-grained supervenience. If two allegedly distinct dispositions turn out to share a supervenience base, then the allegation of their distinctness turns out to be false. Such a view is consistent with a Quinean view of dispositions that identifies dispositions such as solubility or fragility of, for example, a sugar cube with the microphysical structure of the sugar cube (Quine 1960, pp. 222–225). If the sugar cube's solubility and fragility are to be identified with one and the same microstructure, then, by the transitivity of identity, these allegedly distinct dispositions are not distinct after all.

Applying a Quinean attitude toward dispositions to the cases of allegedly distinct pieces of knowledge construed as abeyant states of the neural network involves treating the allegations of distinctness as false. If all and only the same connection weights constitute the network's knowledge of what male faces look like and its knowledge of what female faces look like, then the correct view is that these attempted knowledge attributions fail to attribute distinct mental states to the network. Whatever knowledge the network has, it is one and the same piece of knowledge that constitutes its ability to recognize male faces and its ability to recognize female faces.

Turning to questions of qualia, we see that qualia can be attributed to neural networks but the main proposals view them as occurrent not abeyant representations and as such whatever was said above about consistency with FGS applies as well to qualia. As I spell this out briefly in Mandik (2007, p. 427):

> When Churchland discusses color qualia, he articulates a reductive account of them in terms of Land's theory that human perceptual discrimination of reflectance is due to the sensory reception of three kinds of electromagnetic wavelengths by three different kinds of cones in the retina (Land 1964). In keeping with the kinds of state-space interpretations of neural activity that Churchland is fond of, he explicates color qualia in terms of points in three dimensional spaces, the three dimensions of which correspond to the three kinds of cells responsive to electromagnetic wavelengths. Each color sensation is identical to a neural representation of a color (a neural representation of a spectral reflectance). Each sensation can thus be construed as a point in this 3-D activation space and the perceived similarity between colors and the subjective similarities between corresponding color qualia are definable in terms of proximity between points within the 3-D activation space.

On such a neural network model of qualia, distinct qualia will be implemented in ways fully consistent with FGS.

Lest anyone think current cognitive neuroscience is wholly holistic, it is worth pausing to appreciate just how much is actually quite localizationist. Examples of mental processes localized to specific cortical regions include the visual perception of color and motion (Mandik 2007). Other aspects of vision with distinct localizations include visual recognition of object identity and the visual perception of an object's spatial

location (Mishkin et al. 1983). Further examples include the localization distinct kinds of linguistic competence to Broca's and Wernicke's areas (Bechtel 2001). Koch (2004) has even found evidence of neurons that fire specifically in response to Bill Clinton.

Evidence for localism is evidence for FGS. However, it is worth emphasizing that FGS does not require localism. Holism can be consistent with FGS. Holistic implementations of multiple cognitive functions by a single neural system is consistent with FGS as long as the different functions arise in virtue of physical differences in the neural system.

6 FGS outside of neuroscience

One need not look only to cognitive neuroscience to detect commitments to FGS. We can find tacit, if not explicit, allegiance to FGS in various lines of research in the philosophy of mind that are quite removed from any explicit appeal to cognitive neuroscience. Below I briefly review two: Fodor's (1975) Language of Thought theory of cognition and Rosenthal's (2005) Higher Order Thought theory of consciousness. There is something especially interesting about finding FGS lurking in Fodor's theory, since Fodor is such a high-profile critic of reductive physicalism.

In brief, the language-of-thought hypothesis (LOT) holds that having distinct thoughts not only involves having distinct physical states, but also that these states are composed of distinct re-combinable physical components. So, for example, if a person thinks the distinct thoughts BIRDS FLY and INSECTS FLY, this involves having distinct physical representational states—in these cases a state that represents birds, a state that represents insects, and a state that represents flying. Thinking that birds fly involves a relation of a person to two distinct inner physical items, the bird-representation and the flight-representation. Thinking the distinct thought that insects fly involves a relation to a distinct set of inner physical items, in this case the insect-representation and the flight-representation. The crucial case to consider to see exemplification of FGS concerns when a single individual at a given time instantiates distinct mental properties. Here, LOT supplies distinct physical properties. If one thinks, at a given time, both that birds fly and that insects fly, distinct physical representations—representations of birds and representations of insects—account for the distinctness of the thoughts. (Additionally, distinct tokenings of the flight-representation may be involved.)

Since much of the discussion in this paper concerns consciousness, it will be useful to briefly consider a philosophical account of consciousness. Rosenthal identifies conscious states with mental states that one has a thought about. Higher-order mental states are mental states that are about other mental states. Conscious states are plausibly states one is conscious of and Rosenthal explains what it means to be conscious of a mental state in terms of thinking of a mental state. Rosenthal is no dualist, so the states in question—both lower- and higher-order—are supposed to be distinct physical states. There are several ways in which we can come up with exemplifications of FGS consistent with HOT. One would involve a subject undergoing two distinct states, one of which is conscious and the other of which is not. The two states would be physically distinct states, and the conscious one would be further distinguished by a third state, the higher-order thought about it.

Further work of Rosenthal's (2005) that provides exemplifications of FGS is his homomorphism account of sensory qualities. Sensory qualities are properties of sensory states in virtue of which the states resemble and differ from one another. States of persons, being states, do not bear first-order resemblances to objects. However, the set of resemblances and differences between the sensory states is homomorphic to the set of resemblances and differences between the objects in the external world that we perceive with our senses. Just as a specific color may be identified in virtue of a set of relations to other colors, a sensory quality may be identified in virtue of a set of relations to other sensory qualities. Again, since Rosenthal is no dualist, these states, qualities, and relations are supposed to be physical. The homomorphism theory can be used to construct an exemplification of FGS as follows: A perception of, say, the Japanese flag involves several sensory qualities. To name a few: one corresponding to redness, another to whiteness, and still others for the circularity of the red spot and the rectangularity of the flag itself. The homomorphism theory entails that the distinct mental properties—in this case, the distinct sensory qualities—are instantiated by a given individual at a given time in virtue of distinct physical properties.[14]

LOT and HOT are not unique. Exemplifications of FGS similar to those in connection with LOT can be constructed instead in terms of Evans's Generality Constraint (Davies 1991; Evans 1982). In the domain of philosophical accounts of consciousness, exemplifications of FGS can be spelled out in theories other than HOT. First-order representationalist theories such as Tye (1995) and Dretske (1995) would do just as well. I will not take the time here to spell out further details.

Pointing out the widespread tacit commitment to FGS serves two purposes. The first is to drive home the point that FGS is not to be given up lightly. The second is that, in some cases at least, there are deep tensions in these theories insofar as, like Fodor's, they involve commitments to nonreductive physicalism. For, as it has been a major goal of this paper to show, FGS and nonreductivism constitute an untenable pairing. Given the first point, the second point leads us to see that if one member of the pair is to be given up, it should be nonreductivism.

7 Concluding remarks: what's the big deal about brains?

The point of discussing commitments to FGS by people not explicitly committed to neural reductionism is to show that FGS is not simply something held by people with a prior conviction that everything mental will turn out to be neural. However, I very much want to urge the point that everything mental *will* turn out to be neural and it is now time to consider the question "Why the brain?" Perhaps the relevant question is better put this way:

> Assuming that the considerations concerning mental–mental supervenience, doubled qualia, and the regress argument conclusively prove that all mental

[14] While the main features of Rosenthal's accounts of consciousness and sensory qualities may perhaps be adopted by dualists, Rosenthal's own allegiance to physicalism and intention that the accounts be consistent with physicalism are clear. See in particular Rosenthal 2005, p. 195.

properties must reduce to physical properties, why think that the physical properties that they reduce to will be *neural* properties?

Why, indeed? It seems to be an open question whether distinctively neural properties are essential to the instantiation of mental properties. One can buy into reductive physicalism and reject neural reduction bases in favor of chemical or thermodynamic reduction bases, just to name a few. Perhaps, then, systems that have no distinctively neural properties—nonetheless have certain chemical or thermodynamic profiles that suffice for mentality. Perhaps. But I doubt it. I hope that I may be forgiven for being so brief about this, but I think there are three reasons (at least) for thinking that the physical reduction of the mental should be a neural reduction.

The first reason for believing in neural reduction is that no uncontroversial examples of entities that implement consciousness or cognition exist without brains, or at least, neural networks. It is uncontroversial that alert human adults have mental states. It is also uncontroversial that they have brains. Things are much more contested for the brainless.

A brain is a part of a body and so a ghost, being disembodied, would presumably constitute a brainless realization of mentality. I think it safe to say that there is no non-controversial evidence for ghosts. Sometimes philosophers discussing multiple realizability suppose it possible for there to be aliens or artifacts that instantiate mentality brainlessly with heads instead full of goo or microchips. However, such thought experiments do not sufficiently grapple with the question of how to be sure when we have an instance of brainlessness on our hands. A head full of goo or microchips is not necessarily a head without a brain unless brains are necessarily made out of neither goo nor chips. But if the possibility remains that there could be goo-brains or chip-brains, imagining implementations of mentality realized in different kinds of stuff is not necessarily imagining brainless implementations. Thus even if goo-heads or chip-heads were discovered to implement mentality (which, so far, has not yet happened) they still would not count as uncontroversial confirmations of the mind-endowed lacking brains.

The second reason for believing in neural reduction is that there is no reason to doubt that it is in virtue of their brains (or their brains plus something else) that creatures like us implement consciousness or cognition. Let us entertain briefly how unpromising non-neurocentric theories have been.

Mental properties are had by an organism either in virtue of the whole organism or one of its proper parts. Further, it is easy to accumulate evidence against the first disjunct. For example, traumatic amputations do not necessarily rob amputees of their mentality. Further evidence along these lines may be obtained by comparing the relative effects on mentality of lobotomies and appendectomies. That the seat of our soul is some proper part of us is old news, but the appendix never had a chance and the Aristotelian coronary hypothesis was rejected long ago. What we know about where drugs need to go to go to work and what brain injuries impair what mental functions has tipped the scales pretty clearly in favor of neurocentrism.

Now, these sorts of considerations, while they lead to the view that the brain is *important* for mentality, they leave open the question of whether the implementation of mentality is *exhaustively* neural. In opposition to what we might call the neural

exhaustion thesis (the thesis that the mental is exhaustively neural), we have various embodied, embedded, and externalist proposals for including the body and even chunks of the environment of the organism as part of the supervenience base of the organism's mental properties. While it is far beyond the scope of this paper to refute the theses of the embodied and extended mind, the third point in favor of neural reductionism is relevant to these issues.

The third reason for believing in neural reduction is that no *reductive* research program has been as productive as neural ones. There have been, in recent decades, three major proposals that have been physicalistic without reducing mentality, à la behaviorism, to the behavior of whole organisms: classic computationalism, connectionism, and (certain versions of) dynamic-systems theory (see Eliasmith 2003 for a review of these three positions). Classicism got wedded, in many people's minds, to nonreductive physicalism, largely due to the influence of Fodor (1974) and Putnam (1967). Dynamic-systems theory included proposals of a specifically neural character, (Freeman 1991) while others looked like warmed-over behaviorism (van Gelder 1995). Either way, dynamic systems theory was confronted with some devastating objections (Eliasmith 2001; Glymour 1997; Grush 1997). The main point here, though, is not any knock-down refutations of non-neurocentric research programs. The point here is that neurocentric research programs have been massively productive both in theory and in application.

To summarize: (i) no uncontroversial examples of brainless mentality exist, (ii) organisms that have mentality and brains have mentality in virtue of their brains, and (iii) neurocentric reductionist research programs have been massively more productive than non-neurocentric reductionist research programs. So, if mental properties are going to reduce to physical properties, then (i)–(iii) give reason to believe that the physical properties in question are going to be neural. And why should we think that mental properties are going to reduce to physical properties? The answer is that, if we are going to be physicalists at all, contemplation of doubled qualia and related scenarios leads us to embrace formulations of physicalism that include fine-grained supervenience. Further, once fine-grained supervenience is included in the definition of physicalism, the only way to avoid a nasty regress is to embrace a formulation of physicalism that is reductive.

Acknowledgements For helpful remarks on earlier versions, I thank Murat Aydede, William Bechtel, Lawrence Davis, Anthony Dardis, Chris Eliasmith, Rick Grush, Bryce Huebner, Brian Keeley, Adam Pautz, Tom Polger, John Post, Jesse Prinz, David Rosenthal, Peter Ross, Eric Thomson, Anders Weinstein, Chase Wrenn, Jeffrey Yoshimi, Tad Zawidzki, and two anonymous referees. For keeping my "that" and my "which" nonidentical I thank Rachelle Mandik.

References

Anderson, M. (2007). The massive redeployment hypothesis and the functional topography of the brain. *Philosophical Psychology, 21*(2), 143–174.

Bechtel, W. (2001). Linking cognition and brain: The cognitive neuroscience of language. In W. Bechtel, P. Mandik, J. Mundale, & R. S. Stufflebeam (Eds.), *Philosophy and the neurosciences: A reader*. Oxford: Basil Blackwell.

Block, N. (1978). Troubles with functionalism. *Minnesota Studies in the Philosophy of Science, 9*, 261–325.

Braddon-Mitchell, D., & Jackson, F. (1996). *Philosophy of mind and cognition*. Oxford: Blackwell.
Chalmers, D. (1996). *The conscious mind: In search of a fundamental theory*. New York: Oxford University Press.
Churchland, P., & Sejnowski, T. (1992). *The computational brain*. Cambridge, MA: MIT Press.
Churchland, P. M. (1995). *The engine of reason, the seat of the soul: A philosophical journey into the brain*. Cambridge, MA: MIT Press.
Clark, A., & Chalmers, D. J. (1998). The extended mind. *Analysis, 58*, 7–19.
Davidson, D. (1970). Mental events. In L. Foster & J. Swanson (Eds.), *Experience and theory* (pp. 79–101). Amherst, MA: University of Massachusetts Press.
Davidson, D. (1973). The material mind. In P. Suppes, L. Henkin, A. Joja, & G. C. Moisil (Eds.), *Logic, methodology and the philosophy of science* (pp. 709–722). Amsterdam: North-Holland Publishing Company.
Davies, M. (1991). Concepts, connectionism, and the language of thought. In W. Ramsey, S. Stich, & D. Rumelhart (Eds.), *Philosophy and connectionist theory* (pp. 229–257). Hillsdale, NJ: Lawrence Erlbaum.
Dennett, D. C. (1978). *Brainstorms*. Cambridge, MA: MIT Press.
Dretske, F. (1995). *Naturalizing the mind*. Cambridge, MA: MIT Press.
Eliasmith, C. (2001). Attractive and in-discrete: A critique of two putative virtues of the dynamicist theory of mind. *Minds and Machines, 11*, 417–442.
Eliasmith, C. (2003). Moving beyond metaphors: Understanding the mind for what it is. *Journal of Philosophy, 100*, 493–520.
Evans, G. (1982). *The varieties of reference*. Oxford: Oxford University Press.
Fodor, J. A. (1974). Special sciences, or the disunity of science as a working hypothesis. *Synthese, 28*, 97–115.
Fodor, J. A. (1975). *The language of thought*. Cambridge, MA: Harvard University Press.
Freeman, W. J. (1991). The physiology of perception. *Scientific American, 264*(2), 78–85.
Glymour, C. (1997). *Goethe to van gelder: Comments on 'dynamical systems' models of cognition* (Electronic version). Department of History and Philosophy of Science, University of Pittsburgh PhilSci Archive. Retrieved June 5, 2008, from http://philsci-archive.pitt.edu/archive/00000139/.
Grush, R. (1997). Review of port and van gelder's mind as motion. *Philosophical Psychology, 10*(2), 233–242.
Haugeland, J. (1991). Representational genera. In W. Ramsey, S. P. Stich, & D. E. Rumelhart (Eds.), *Philosophy and connectionist theory* (pp. 61–89). Hillsdale, NJ: Lawrence Erlbaum.
Hofweber, T. (2005). Supervenience and object dependent properties. *The Journal of Philosophy, CII*(1), 5–32.
Koch, C. (2004). *The quest for consciousness*. Englewood, CO: Roberts & Company.
Land, E. H. (1964). The retinex. *Scientific American, 52*, 247–264.
Lycan, W. (1987). *Consciousness*. Cambridge, MA: MIT Press.
Lynch, M., & Glasgow, J. (2003). The impossibility of superdupervenience. *Philosophical Studies, 113*, 201–221.
Mandik, P. (2007). The neurophilosophy of consciousness. In M. Velmans & S. Schneider (Eds.), *The Blackwell companion to consciousness* (pp. 418–430). Oxford: Basil Blackwell.
McLaughlin, B., & Bennett, K. (2006). Supervenience. *The Stanford encyclopedia of philosophy* (Electronic version, Fall 2006 ed.). Retrieved June 5, 2008, from http://plato.stanford.edu/archives/fall2006/entries/supervenience/.
Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neurosciences, 6*, 414–417.
Mundale, J. (2002). Concepts of localization: Balkanization in the brain. *Brain and Mind, 3*, 313–330.
Putnam, H. (1967). The nature of mental states. In W. H. Capitan & D. D. Merrill (Eds.), *Art, mind, and religion*. Pittsburgh University Press.
Quine, W. V. (1960). *Word and object*. Cambridge, MA: MIT Press.
Rosenthal, D. (2005). *Consciousness and mind*. Oxford: Clarendon Press.
Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*, 417–457.
Tye, M. (1995). *Ten problems of consciousness*. Cambridge, MA: MIT Press.
van Gelder, T. (1995). What might cognition be if not computation? *Journal of Philosophy, 92*, 345–381.
Wilson, J. M. (2005). Supervenience-based formulations of physicalism. *Nous, 39*(3), 426–459.