BEHAVIORAL AND BRAIN SCIENCES (2004) **27,** 603–647
Printed in the United States of America

What to say to a skeptical metaphysician: A defense manual for cognitive and behavioral scientists

Don Ross
Department of Philosophy, University of Alabama at Birmingham, Birmingham, AL 35294-1260 and School of Economics, University of Cape Town, Rondebosch 7701, South Africa
dross@commerce.uct.ac.za
http://www.commerce.uct.ac.za/economics/staff/personalpages/dross/

David Spurrett
School of Philosophy and Ethics, University of KwaZulu-Natal, Durban 4041, South Africa
spurrett@ukzn.ac.za **http://www.nu.ac.za/undphil/spurrett**

Abstract: A wave of recent work in metaphysics seeks to undermine the anti-reductionist, functionalist consensus of the past few decades in cognitive science and philosophy of mind. That consensus apparently legitimated a focus on what systems do, without necessarily and always requiring attention to the details of how systems are constituted. The new metaphysical challenge contends that many states and processes referred to by functionalist cognitive scientists are epiphenomenal. It further contends that the problem lies in functionalism itself, and that, to save the causal significance of mind, it is necessary to re-embrace reductionism.

We argue that the prescribed return to reductionism would be disastrous for the cognitive and behavioral sciences, requiring the dismantling of most existing achievements and placing intolerable restrictions on further work. However, this argument fails to answer the metaphysical challenge on its own terms. We meet that challenge by going on to argue that the new metaphysical skepticism about functionalist cognitive science depends on reifying two distinct notions of causality (one primarily scientific, the other metaphysical), then equivocating between them. When the different notions of causality are properly distinguished, it is clear that functionalism is in no serious philosophical trouble, and that we need not choose between reducing minds or finding them causally impotent. The metaphysical challenge to functionalism relies, in particular, on a naïve and inaccurate conception of the practice of physics, and the relationship between physics and metaphysics.

Keywords: explanation; functionalism; mental causation; metaphysics; reductionism

1. Introduction

Philosophy progresses with a tide-like dynamic. Every wave, no matter how strong it seems while rolling in, is followed by a backwash, often nearly as powerful. This makes philosophical development difficult to identify except in long retrospect. For scientists who try to take philosophy seriously in their work, this is bound to be frustrating

Philosophers have been deeply involved in the development of cognitive science. The classical essays by Hilary Putnam (1963; 1967a; 1967b; 1975b) and David Lewis (1972) that articulated and promoted functionalist understandings of mind are among the foundational documents in the literature of the field. Among other things, they showed how and why the study of information processing as conducted in early Artificial Intelligence (AI) could and should be integrated with psychology more generally. And however far in sophistication the cognitive science community has since moved from the narrowly computational models of the 1960s and 1970s, it is hard to see how it would have gotten where it is now without them. So philosophers

Don Ross is Professor in the Center for Ethics and Values in the Sciences and the Department of Philosophy at the University of Alabama at Birmingham and in the School of Economics at the University of Cape Town. He is the author or editor of ten books and over forty articles on the foundations of the behavioral sciences, game theory, the philosophy of science, and the economics of developing countries. The first volume of his two-volume *Economic Theory and Cognitive Science* will appear from MIT Press in 2005.

David Spurrett is Associate Professor of Philosophy at the Howard College Campus of the University of KwaZulu-Natal, where he is also coordinator of the mind *and* world working group. He is the author of around twenty publications in the areas of philosophy of science, cognitive science and metaphysics, appearing in journals including the *Pacific Philosophical Quarterly, International Studies in Philosophy of Science,* and *Language Sciences.* In 2003, he was awarded the Vice-Chancellor's Research Award at the University of Natal and a President's Award from the National Research Foundation (South Africa).

 0140-525X/04 $12.50

do not exaggerate when they claim that their discipline has contributed crucial bricks to the edifice of contemporary cognitive and behavioral science.

By *functionalism* we understand any position that assigns serious ontological status to types of states or processes individuated by reference to what they *do* rather than what they are *made of* – that is, by reference to their effects, rather than (necessarily) their constituents. Functionalism of this sort was never without its critics, of course. From our perspective, eliminative materialists (e.g., Churchland 1981) have been the most important of these, and their arguments with mainstream functionalists have been immensely helpful in the effort to see how the neurosciences and robotics best integrate with the more rationalistic projects derived from AI. However, avowed eliminativists have always been a fringe, playing against a relatively monolithic functionalist consensus. For most of the past 30 years cognitive scientists could be assured that the main currents in the philosophy of mind, especially regarding causation and explanation, were running in a direction sympathetic to their activities. This has involved more than encouraging cheerleading, amounting to something of working scientific value. It has helped to guide choices amongst research directions by clarifying just where and how cognitive science might strive for serious integration with nearby research programs in, for example, neuroscience and the physics of dynamical systems theory (see sects. 3.1 and 4.2 below) without simply collapsing into them.

We regret to report, however, that the backwash has set in. Were a cognitive scientist to stroll into a typical discussion amongst the "purer" philosophers of mind at a professional seminar in 2004, she would find that functionalism is under siege in such settings. Instead, "new wave" reductionism[1] is the horse on which increasing numbers of philosophers are placing their bets.

We are, of course, being melodramatic here, and deliberately so. Good philosophers are rightly cautious about changing their minds or investing in fads, and worthwhile philosophical activity is not best seen as a war of "isms." Nevertheless, as philosophers we are concerned by the rise of a new scholasticism in philosophy of mind that, in the stated pursuit of a return to "real" metaphysics, threatens a loss of contact with empirical cognitive science. Our aims in this paper are, first, to substantiate this concern (thereby justifying the melodrama), and second, to offer grounds for resisting the arguments that inspire it. We think that metaphysics – "real," professionally done, metaphysics – is an important part of all science, including cognitive science. But we also think that what is recently being promoted under this banner is based on an unhealthy disregard for the actual practice of science, and that too little philosophical discussion of metaphysics shows adequate concern for this. To the extent that some philosophers allow their discussions to drift away from relevance to and coherence with scientific activity, the short-term course of cognitive science will not be much affected. However, since we would deplore a situation in which the conversation between philosophers and cognitive scientists wound down into separated silos, we think that a corrective with two aims is in order. One aim is to address philosophers themselves concerning the fundamental errors we diagnose in the new scholasticism. The other is to provide cognitive scientists with a manual for answering philosophers who try to convince them that there is something wrong with their metaphysics. After all, to the extent that cognitive scientists respond to philosophers by just shrugging and going off to another room, the conversation winds down; to the extent that they argue back, it continues.

Our discussion is organized as follows: In section 1.1 we review the standard arguments for functionalism in the special sciences, and offer an account of the rise of the functionalist consensus. In this section we also briefly describe the recent threat to functionalism. In section 2 this threat is examined in greater detail, looking first (sect. 2.1) at an influential argument against functionalism, and then (sect. 2.2) at the form of reductionism which increasing numbers of metaphysicians think is preferable. In section 3 we say a little more about the argument in favor of functionalism from the perspective of the special sciences (sect. 3.1), and also argue that the reductionism suggested by the metaphysicians would be disastrous for those sciences (sect. 3.2 and 3.3). Section 4 contains the metaphysical meat of this paper. In it we distinguish some different ways of taking metaphysics seriously (sect. 4.1), consider the relationships between explanation and causation (sect. 4.2), distinguish two senses of causation (sect. 4.3), and clarify a number of considerations relating to the nature of physics and the relationship between physical science and the metaphysics of causation (sect. 4.4). In section 5 we review our argument and offer a conclusion.

1.1. Functionalism, philosophy, and the behavioral sciences

Functionalism has a strong claim to being part of the methodological and ontological underpinning of any special science. By "special" science we have in mind any science not concerned with justifying, testing, or extending the generalizations of fundamental physics, and hence most science, including (see sect. 4.4) most of physics. Functionalism offers one way in which special sciences can defend their significance against Rutherford's claim that "there is physics, and there is stamp collecting" (Birks 1963). What potentially distinguishes a special science from stamp collecting is that it is organized around a distinctive taxonomy of phenomena and a set of processes, at some level of abstraction from fundamental physical processes which are non-redundant, amenable to scientific treatment, and to which a fully realistic attitude can be justified.

The original considerations that led to the development of functionalism were, as it happens, primarily drawn from issues in the sciences of behavior: a response simultaneously to a simplistic behaviorist equivalence of behavior and psychological state, and to the apparent "chauvinism" of expecting that the specific mechanisms which accounted for psychological states in humans should be regarded as reductive explanations of those states in general.

With respect to simplistic behaviorism, the case for functionalism runs as follows. Traditional behaviorism identified mental states with dispositions to particular behaviors, and hence expected that mental states – insofar as these were of scientific significance at all – could be read directly off surface behavior,[2] so reference to behavior could, and should, replace reference to mental states. One objection to this program pointed out that if mental states can, as seems likely, interact with one another, then there will be neither fixed nor simple pairings of mental states and dispositions to particular behaviors. This line of thinking suggests a place for intermediate causal roles played by (at least

initially) unobservable states between *stimulus* and *response.* Note that in the first instance these hypothesized intermediate states are characterized extrinsically, by reference to the difference they make to observable states and relations between those states.[3] At this stage, at least, it is possible to be agnostic about what it is that makes the difference in question, even while being confident that some difference is being made. This space for agnosticism about *what* plays the functional role in question relates closely to the second, and for present purposes more philosophically contentious, motivation for functionalism.

In this case the contrast is provided not by behaviorism but by reductionism. In the heyday of type-type reductionism it was expected that particular types of special-science states would pay their ontological and causal way by being reduced to types of some science closer to fundamental physics, applied to the same systems. So, perhaps, the biological properties of a system could be reduced to its chemical properties, and from there chemistry could be reduced to physics. Classic statements of versions of this view include Oppenheim and Putnam (1958) and Nagel (1961). The mental state of being in pain would turn out to be, or already was, reducible to the fact of having activated C-fibers (Place 1956; Smart 1959),[4] in the same way that "temperature" was supposedly reducible to mean kinetic energy of molecules (Nagel 1961). On this view, the type "pain" was to be considered *reducible* to the type "activated C-fibers" when some biconditional bridge law was found enabling statements about pain to be translated into statements about C-fibers, and vice versa. The other way of motivating functionalism is, then, to note that the proposed reduction is open to a charge of "chauvinism" (Block 1980b), because even if the biconditional linking pain and C-fiber activity in *humans* held, there presumably could be, or already were, agents physiologically different from humans that nonetheless experienced pain. So, in what came to be the standard jargon, even if something involving C-fibers were that which *realized* pain in people, the *role* of pain could be realized in different ways in other types of agent.[5] This, in a nutshell, is the multiple realization argument against type-type reductionism for psychological states, and, by implication, an argument for a science of psychology that spans differences in realization.

The multiple realization argument can be deployed in various ways as a positive argument for the functionalist project. In the hands of, for example, Fodor (1974; 1975; Block & Fodor 1972) it is used to make clear that many, at least, of the special sciences are concerned with entities and processes that are to some degree abstracted from the details of physical realization. As noted above, practitioners of those sciences can afford to be agnostic about the physical details that realize the relevant kinds and processes, because the distinctive descriptive and explanatory contribution made by their work depends for the most part on extrinsic, functional relationships between role properties. A simple and classic illustrative example of the argument here is Fodor's treatment of the notion of a mousetrap (Fodor 1968), conceived in functionalist terms as a device that takes as input a live mouse, and produces as output a dead one. Clearly, a wide range of devices and designs are capable of realizing the mousetrap role.

The immediately preceding discussion has referred to the concepts of roles and realizers in stating and partly defending functionalism. This distinction (which is stated and clarified in slightly different terms in Block 1980a) also allows two importantly different ways of being a functionalist. The difference turns on whether one is inclined to identify the functional states with the *role* they play, or with what is the *realizer* of that role in a given case. Put another way, a functionalist might think that "pain" or "money" pick out either *the property of having some other (physical) property that* realizes pain *or* money, or that, properly analyzed, they pick out *C-fibers firing* or *dollar bills.*[6] Saying that the description of the functional state picks out the role indicates commitment to the view that even though their realizers could be very different, humans and, say, Martians could be in the *same* mental state when in pain. On the other hand, tying the function to the realizer entails that humans in pain and Martians in pain are in different states, perhaps different types of pain, just *because* the realizers of the roles in each case are different.

Note that both flavors of functionalism are entirely compatible with materialism, or physicalism. A physicalist functionalist of either type will be committed to the principle that physics is complete, or causally closed, that is, that there are no nonphysical, for example, vital, fundamental forces (Papineau 1993; Spurrett 2001b; Spurrett & Papineau 1999). Similarly, she will be committed to the thesis that if you fix all of the physical facts, then you have fixed all the empirical facts that there are. Often, although not necessarily, this aspect of functionalist thinking is marked by saying that functionalists accept *supervenience* – the idea that there are no changes without physical changes.

Note also that there is a genuine tension between the different ways of being a functionalist. From the perspective of realizer functionalism, the role variety rides roughshod over distinctions that need to be taken seriously. So the *equivalence* of a hundred dollar bill, a check for the same amount, and a bag of coins with a total value of a hundred dollars does not amount to much when we have to try and say something about why we can only use one of them in a vending machine, or why only one of them can be bounced by a bank. On the other hand, from the perspective of role functionalism, too much attention to the realizers amounts to abandoning the apparent unity of many apparently powerful and useful generalizations. Qua "money," it has to be granted, there is a deep sense in which any realization of one hundred dollars just *is* the same.

The importance of this distinction did not emerge immediately during the early articulation of functionalism. In the classic papers collected in Putnam (1975a), for example, role and realizer versions are run together in a way that is, in critical retrospect, problematic. Philosophers were quick enough to unearth the tension, however. By the 1980s, central debates in the philosophy of mind revolved around arguments between role and realizer functionalists.[7] However, for a number of years, up to the mid-1990s, the debates were preoccupied with the question of whether semantic *meanings,* as bearers of functional roles for beliefs, desires, and other "propositional attitudes," could or could not be individuated for the purposes of cognitive science just by reference to intrinsic properties (causal, computational, constitutional, or whatever), or were irreducibly relational. This running controversy was known among philosophers as the *internalism versus externalism* debate, and for a while it seemed as if the dispute between realizer and role functionalists turned mainly upon it. Fortunately we need not describe its details here, because by the mid-

1990s it was largely over, with the internalists – the believers in so-called narrow content – having mostly surrendered (see Fodor 1994; Ross 1997).

At that point, some thought that the philosophy of mind had made itself ready for thorough integration into cognitive science. In particular, the strong connection between externalism about semantics and the idea that narrowly computational models of thought need to be replaced or supplemented by more biological, environmentally situated, and robotic ones (Brooks 1991) made the prospects for positive philosophical contributions to the scientific project look promising. Some of that promise has been realized; we cite, for example, Clark (1997) and Rowlands (1999) with approval in this connection. However, from around these same years two ideas gained strength among philosophers that encouraged skepticism about, instead of participation in, mainstream cognitive science. The first of these ideas, the conviction that qualitative consciousness is beyond the reach of functionalist method (Chalmers 1996), or, on some formulations, *any* scientific method (McGinn 1991), is a manifestation of conservative metaphysics that we thoroughly deplore, for reasons given in Ross (forthcoming) and Dennett (2001a; 2001b). This will not be our concern in the present paper, however. The second basis for metaphysical party-pooping, which *is* our present subject, encourages even deeper skepticism because it challenges not just functionalism's adequacy in a particular domain, but its coherence in general.

For reasons we explain in section 2 below, realizer functionalism did not die with semantic internalism. As far as we know, the first recognizably contemporary expression of the worry that states picked out by reference to functional roles alone cannot *cause* anything appears in Fodor (1987). However, at that point the worry was deeply enmeshed in the internalist/externalist controversy, so its subversive potential was not clearly spotted. With the passing of internalism it popped clearly into wide view. In Kim (1998) it finds book-length and elegant expression,[8] and our experience as casual anthropological observers of fellow philosophers indicates to us that the majority of philosophers of mind are, although not unanimously persuaded by this version, inclined to take the worry very seriously and, if not agreeing with its conclusion, to accept the same basic picture of how things are in science, especially physics, when engaging with it (e.g., Elder 2001; Marcus 2001). In what follows we occasionally find (and cite) allies among "pure" philosophers, and it is no part of our project to argue that *nobody* should pursue these problems by primarily logical methods. Given, though, that we are here confronted with a piece of metaphysics claiming consequences for science, we take it as deserving evaluation with an eye to the science *and* the metaphysics. As Marras (2000) points out, and as we will explain, what was originally supposed to be a consideration against role functionalism but *for* realizer functionalism now looms as a skeptical threat to *all* functional explanation in any science.

Our aim in this paper is to comprehensively respond to the basis for this skepticism, from the perspective of behavioral and cognitive science. Doing this, however, requires some excursions deep into metaphysics. Some scientists will likely doubt that such excursions could be worthwhile trips for them to go along on. Hearing that philosophers are making themselves uneasy about the enterprise of cognitive science because of metaphysical itches, they may be inclined to respond pragmatically, saying "*we* feel fine, so *you* stop scratching!" Such responses, often heard when philosophers confront scientists with their metaphysical scruples, do not *just* express a macho attitude. It has been a widespread opinion among philosophers of science for decades that philosophy has no privileged epistemic perspective from which it legitimately can or should try to bend science to any prior ontological objective or methodology. We endorse this stance, in a fashion to be made more precise shortly. However, we *also* agree with Kim (1998) that, if metaphysics matters, then it had best be done seriously. We believe furthermore that metaphysics only matters if it matters to science; and, finally, we believe, and argue below, that metaphysics matters to science. Given all this, it of course follows that if the metaphysical presuppositions of cognitive science are causing genuine itches, then everyone ought to care about scratching in the right place.

We find it necessary to say something about these grand themes for the following reason. Kim's flagship argument against the recent externalist-functionalist near-consensus *does* have a scholastic aura about it; in particular, as philosophers take up Kim's challenge and gnaw away at the problems he has raised, they focus a great deal of their attention on subtle differences amongst variations on the definition of the *supervenience* relation. We will ultimately conclude that this really is a scholastic's response, in the bad sense of the word (if there is a good sense). But this generates two strategic concerns at the outset. First, this conclusion may lead some philosophers to suppose that we are trying to have what Kim calls a "free lunch," that is, simply refusing to take the demands of metaphysics seriously. Second, the fact that what we regard as a *scientifically interesting* metaphysical problem comes dressed in scholastic garb – it is even based on something *called* "the supervenience argument" – will lead too many scientists to conclude right away that we are engaged in an in-house philosophers' quibble that is not any proper business of theirs. These concerns present us with the following tactical burden. We must present the supervenience argument – the basic grounds for the new disquiet – in a way that does logical justice to it *and* captures the gripping intuitions behind it that we do not think you have to be scholastically inclined to appreciate.

So, here goes. Our talk about "scientifically interesting metaphysics" gestures at the following fact. It *is* a feature of scientific epistemology, as really practiced in laboratories and journals, that the various pieces of scientific inquiry must broadly cohere into a general world-view that, at least in its core, almost all signed-up members of the mainstream scientific professions can share. Furthermore, it is a legitimate job of the "serious" metaphysician to ensure that proposals for articulating and enriching this world-view are, at least potentially, genuinely enlightening, and not merely verbal or technical. By "genuinely enlightening" we mean that such articulations should actually be able to help scientists choose amongst theoretical and/or procedural alternatives in cases where the empirical facts remain sufficiently underdetermined to leave options open in pragmatically pressing (as opposed to just logically possible) ways. Now, what we have just said is not very precise, and so not very bold. But it is enough to help show why metaphysics *can be* (and the issues raised by Kim's supervenience argument *are*) scientifically interesting. Our bland claim makes a minimal commitment to the idea that, at some level of ab-

straction, the sciences need to "hang together." However, this commitment is in direct tension with the best motives for having special sciences, all of which turn on the facts that, along various dimensions both ontological and epistemological, different sciences do not hang together, and that we will deny ourselves important insights and generalizations if our respect for minimal metaphysics makes us work too hard to try to get them to do so. Kim's supervenience argument is aimed precisely at this tension, and with unusually limpid clarity. Though, as we shall see, the argument generalizes all the way across the sciences, it helps its clarity, but at the same time especially challenges cognitive and behavioral scientists, that it is focused directly on this tension as it arises within the domain of their work, which sits across the fault line between generalizing and special ontologies. So we think that a working cognitive scientist who is confronted with Kim's argument will and should then notice the tension every time she goes to write up some new results, and will and should feel itchy. The problem then, we will argue, is that Kim and other philosophers, instead of telling her where and how to scratch, counsel relief through professional suicide. We will be pleased to show that this is not called for.

Kim's argument is aimed directly at role functionalism. According to its conclusion, role functionalism is not a stable metaphysical position. Instead, it collapses into a choice between epiphenomenalism and reductionism about mental properties, objects, and processes. Kim assumes that epiphenomenalism would be a dire outcome, both metaphysically and scientifically, but then spends much of his book trying to make reductionism seem palatable. As we will show, he is not convincing. The foundational assumptions of cognitive science, along with those of other special sciences, deeply depend on role functionalism. Such functionalism is crucially supposed to deliver a kind of causal understanding. Indeed, the very point of functionalism (on role *or* realizer versions) is to capture what is salient about what systems actually do, and how they interact, *without* having to get bogged down in micro-scale physical details. Functionalist understanding is, furthermore, supposed to deliver all the goods of properly causal scientific work: permitting predictions, causal explanations, sustaining counterfactuals, enabling the planning of interventions, and so forth. But if reference to role properties can be shown to be causally redundant, as Kim's argument purports to show, then the appearance of causal relevance is a sham, and role functionalists, including most cognitive and behavioral scientists, most of the time are really only telling "just so" stories to one another.

So, apologies for some coming scholasticism duly made, let's now get this dangerous supervenience argument onto the table.

2. The armchair strikes back

According to Kim (1998), the key challenge to role functionalism turns on what he calls the "causal exclusion" problem, which arises if he is correct that putative physical and mental causes for the "same" event can be shown to be in conflict. His problem, therefore, is to provide an answer to the question: "*Given that every physical event that has a cause has a physical cause, how is a mental cause also possible?*" (Kim 1998, p. 38). This is the problem of "finding a place" (p. 2) for mind in a physical world, given the causal closure of physics. The fact that Kim is concerned with the problem as a *metaphysical* challenge means that it will not do simply to point out the pragmatic benefits, or indispensability, of mentalistic explanations (including causal ones) without having a good metaphysical story to tell about *how* and *why* such explanations are legitimate (cf. Marcus 2001). This would be the strategy, discussed and rejected above, of ignoring the demands of metaphysics; asking, as Kim says, for a free lunch – keeping your comfortable intuitions by refusing to notice that they commit you to anything outside of cognitive science.

2.1. Kim's "supervenience argument"

Whether or not a cognitive scientist is in the habit of using the word "supervenience," chances are good that some of her daily working assumptions involve at least a loose version of the concept. Starting generally: one set of (e.g., mental) properties supervenes on another (e.g., physical or neurobiological) set if, roughly, something cannot change with respect to its supervening properties without undergoing some change with respect to its subvening ("base") properties. Materialist functionalism involves commitment to supervenience in this sense, insofar as it is reasonable to suppose that the particular role some entity realizes cannot change without some physical changes taking place *somewhere.* This relationship of covariance plus some kind of dependence (because physical changes need not lead to changes at the supervening level) is weaker than reduction, and does not commit you to anything like realizer functionalism (let alone internalism) unless you add that the relevant physical change has to occur *in the realizer.*

Kim's argument takes the form of a dilemma that "apparently leads to the conclusion that mental causation is unintelligible" (1998, p. 39). The dilemma has two horns: on one horn mind-body supervenience is allowed to fail, and on the other it is assumed to hold. For the purposes of formulating the dilemma, Kim (1998) defines the mind-body supervenience thesis as follows:

> Mental properties supervene on physical properties in the sense that if something instantiates any mental property *M* at *t,* there is a physical base property *P* such that the thing has *P* at *t,* and [nomologically] necessarily anything with *P* at a time has *M* at that time. (p. 39)

This definition is not perfectly general. Philosophers have generated a large literature that debates the merits and failings of alternative definitions of supervenience. What is at issue in these arguments is the appropriate *scope* to aim for in stating generalizations about functional role-fillers. At *least,* "pain" should apply generally and univocally to (most) people, and probably to creatures with which people share recent common (or, perhaps, any) ancestors. Perhaps almost any life form would need a trip-wire system that alarmed it by making it feel bad. If so, then Martians would have pain too, however different its realizers might be in them. Now, to help discipline arguments about this sort of thing, it is a useful strategy to first fix the *essential* conditions on pain; that way you hold your semantics fixed and can test the empirical facts independently. Philosophers fix essential semantics by considering various abstract possibility classes, or "possible worlds." Depending on how many of these classes you want to legislate supervenience relations as having to hold across, you get different logical def-

initions of the relation. Fortunately for our purposes here, Kim's dilemma arises for *any* such definition, so we will treat the version just quoted as exemplary.

Here is the first horn of the alleged dilemma. If mind-body supervenience, in general, were to fail, and we are committed to the causal closure of physics, then it seems as though we could not make sense of mental causation. Put another way, if the supervenience relation *does not* hold, and mental causes do have physical effects, then we would have to deny the causal closure of physics – we would be claiming a physical consequence of a non-physical cause. As materialists, or physicalists, we cannot do *this,* so it looks like the supervenience relationship has to hold. (Kim takes commitment to the causal closure of physics as being a "minimal" requirement for physicalism). So far so good – this is a standard motivation for endorsing supervenience if you are not willing to be a reductionist (e.g., Fodor 1987).

On, then, to the other horn. "Suppose that some instance of mental property *M* causes another mental property *M** to be instantiated" (Kim 1998, p. 41). By the mind-body supervenience thesis, *M* has a physical supervenience base *P,* and *M** has a physical supervenience base *P**. Kim asks us to grant that *P* causes *P**. But, then, since *M** is realized by *P**, why have an apparently separate causal claim to the effect that *M* caused *M**, especially when it seems as though once *P,* then *M** was going to happen anyway (following Kim 1998; see also Marras 2000)? More precisely we seem to have to make a choice: "*M** is instantiated on this occasion: (a) because, ex hypothesi, *M* caused *M** to be instantiated; or (b) because *P**, the physical supervenience base of *M**, is instantiated on this occasion" (Kim 1998, p. 42).

Kim notes that the apparent tension above could be relieved by accepting that "*M* caused *M** by causing *P**." But, given that both *M* and *M** have respective physical supervenience bases, we should ultimately grant that "*P* caused *P**, and *M* supervenes on *P* and *M** supervenes on *P**" so that the "*M*-to-*M** and *M*-to-*P** causal relations are only apparent, arising out of a genuine causal process from *P* to *P**" (1998, p. 45).

So: If you *deny* supervenience you seem to be abandoning materialism, which would be terrible,[9] and if you *affirm* it you get stuck with a choice between epiphenomenalism about the mental, or reductionism. The former is an awful option for cognitive science. Therefore, the only option is reductionism. This is genuinely amazing, because the very point of endorsing supervenience was originally to allow materialism *without* reductionism!

2.2. Kim's reductionist proposal

Kim's reductionism is not quite the standard ("Nagelian type-type") variety that people still learn in undergraduate metaphysics and philosophy of science courses (see Marras 2002). According to that model, you reduce some type *x* to some type *y* by justifying a "bridge law" to the effect that all of the causal and other law-like generalizations you can state in terms of *x* can be restated in terms of *y.* Instead, Kim proposes a reductionism that proceeds along the lines suggested by Armstrong (1981) and Lewis (1980). The details of the proposal involve a crucial step called "functionalization" that involves "enhancing bridge laws . . . into identities" (Kim 1998, p. 97).[10] Identities, unlike bridge laws, give ontological simplification, and promise to explain why it is that the bridge laws hold true. Functionalization is to be achieved by "priming" the to-be-reduced mental property (the proverbial *M*) for reduction, which means reconstruing it in extrinsic or relational terms, that is, specifying its causal relations to other properties. So *M* is now "the property of having a property with such-and-such causal potentials, and it turns out that physical property *P* is exactly the property that fits the causal specification" (Kim 1998, p. 98). It follows that *M* can be identified with *P,* which would solve the causal exclusion problem, because one property cannot be in competition with itself over causal relevance, and Kim thinks there is no problem about the causal capacity of physical properties.

It is, of course, an open question to what extent such reductions are possible, and how extensive the scope of any given functionalising reduction will be. The multiple realization argument (discussed above and, again, below) indicates that functionally individuated properties can have very diverse realizers, so functionalising reductions should be expected to involve some disintegration of the role properties. Kim himself seems comfortable with this, describing the upshot of his arguments as being that "multiply realized properties are sundered into their diverse realizers in different species and structures, and in different possible worlds" (1998, p. 111). This is supposed to save *something* of functionalism, albeit at the expense of relinquishing the capacity to say what it is that makes some apparently similar functional properties related or the "same" in cases where their realizations are significantly different. (We return to the question of just how much difference would count as significant in due course.) Kim's approach, interestingly, inverts the standard image of functionalism, traditionally regarded as a major form of anti-reductionism, because on his view "the functionalist conception of mental properties is *required* for mind-body reduction" and is even "necessary and sufficient for reducibility" (1998, p. 101). But is this functionalism at all? Marras (2000) thinks not, and argues that Kim has "in fact given up on functionalism" of which a central idea was that mental/functional properties retained their "identity and projectibility across heterogeneous physical realizers." Kim, who claims to take multiple realizability "seriously," concedes that to those who might want to "hang on to" functional properties as "unified and robust . . . in their own right," his proposal will be a "disappointment" but also maintains that the conclusion in question is "inescapable" (1998, p. 111).

Notice at once that if there *is* any sort of functionalism still alive in Kim's proposal, it is realizer functionalism, not role functionalism. So perhaps what Kim's argument, and his way out of it, shows is that if you want to try to be a *serious,* anti-reductionist, functionalist then you had, somehow, better be a role functionalist. As discussed in section 1 above, many have thought this since at least 1987; but initially the implausibility of semantic internalism was the main reason. Now it turns out that there's a more general reason: if you try to be a realizer functionalist, you'll turn "inescapably" into a reductionist, and you won't be able to do cognitive science (or biology, or economics, or . . .)! Or so we now aim to show. Remember, though, that showing we would be *in trouble* if we followed Kim, no matter how *big* the trouble, does not show that we are *not* in trouble. Acknowledging that is the price of taking metaphysics seriously.

3. Special sciences without functionalism

In section 1 of this paper we outlined the reasons for the establishment of a broad functionalist consensus in the behavioral sciences, and the special sciences more widely. Functionalism seemed – was *devised* to be – ideal for such sciences, insofar as it offered a justification for focusing on role properties and extrinsic relations, coupled with a well-motivated degree of agnosticism about the exact physical details of the systems studied. In section 2, though, we described Kim's supervenience argument, contending that functional causal claims, understood as being claims about properties which supervene on more basic physical properties, are epiphenomenal, and can only have their causal status saved by reducing them to physical properties.

It is not essential that anyone view this as a problem. One simple way to avoid the challenge Kim poses is to be an instrumentalist about functional claims. That means contending that metaphysical questions about the causal status of scientific claims are just not important, and that what really matters is whether science is, in some sense or other, "useful." It is not, after all, *compulsory* to worry about metaphysics. If you are indeed willing to say that, ultimately, the validity of some piece of science is determined on pragmatic grounds, then this is your stop, and you can disembark right now. In so proceeding you are allowing that you do not mind if the behavioral sciences are considered to be a kind of stamp collecting – a process of arranging the artifacts of our own epistemic limitations in interesting or useful-seeming ways. (As we argue shortly, in so doing, whether you like it or not – and more to the point, whether *he* likes it or not – you are agreeing with Kim, because the only place he leaves open for the special sciences is an instrumentally justified one.)

If you're still here then perhaps you want to be more than a stamp-collector. Perhaps you want a defensible functionalist conception of *pain* that generalizes across species, or of *competition* that generalizes across organisms and ecologies, or even of *mousetrap* that does justice to the varied assortment of gadgets you have around for the purposes of killing mice. In this section we aim to do two things: first (in sect. 3.1), extend section 1 above by developing stronger and more sophisticated arguments against reductionism in the special sciences; and second (in sect. 3.2), make clear that Kim's proposal *does* amount to turning special scientists into stamp collectors.

3.1. Explanation and causation

It is a manifest fact about science that the various special sciences are partly constituted by parochial types of causal relations. Indeed, this is one of the principal things making them *special.* These relations are, furthermore, reciprocal functions of the accepted explanatory schemata in the relevant sciences. This fact is, at least in the first place, sociological rather than metaphysical. One way of being what Kim derides as a free-lunch seeker is to take this as a brute fact in need of no explanation, supposing instead that the "specialness" of each special science, taken individually, is somehow self-justifying. *Part* of what is involved in heeding Kim's enjoinder to take metaphysics seriously is acknowledging the need to say something about the circumstances under which special-science accounts are genuinely explanatory, where it is presumed that a genuine explanation is not merely something psychologically satisfying to someone, but must cite explanans that are both true and informationally non-redundant. In this section, we will show that, in light of leading accounts of explanation from the philosophy of science literature, Kim's version of reductionism would disqualify many or most prima facie powerful special-science explanations.

Where special sciences are concerned, we can inquire about the explanatory value of a specific account at either or both of two levels. An account might be genuinely explanatory just relative to the particular ontological and causal structure of the science in which it is embedded, but remain mysterious from the perspective of the wider standpoint at which science as a whole is expected to "hang together." Kim, of course, contends that explanations citing mental causes have just this status unless we embrace his reductionistic version of realizer functionalism. The inquirer who takes metaphysics seriously seeks accounts of phenomena that are explanatory *both* relative to the ontological presuppositions of her special science, and to whatever wider metaphysical principles unite the sciences as a whole. The project of seeking explanatory generality of this sort is historically, actually and normatively, *part* of the business of science. That is to say, the naturalistically oriented metaphysics that we engage in is continuous with, rather than separate from, what "scientists" do. Our main criticism of Kim's proposal, to which we will devote section 4, is that the particular wider metaphysical perspective he takes for granted has no persuasive justification. At the moment, however, we are concerned with tracing the consequences of Kim's proposal for the special sciences, and for the cognitive and behavioral sciences in particular. But since we contend that one such consequence would be the disqualification of a whole class of important (putatively, at least) explanations, we cannot avoid some introduction at this point of general considerations from the philosophy of science. For the moment, these considerations are intended to facilitate our discussion of special-science explanations. In section 4, we amplify them in a general treatment of the demands of serious metaphysics.

Kitcher (1976; 1981) has argued that *ontological unification,* either within a special science or across two or more special sciences, consists in the justification of common *argument patterns* that hold within or across, respectively, the science(s) in question. This claim is then substantiated through detailed analysis of the concept of an argument pattern, which is a set of ontological and structural primitives featuring recurrently in the explanations given in the unified domain. Thus, for example, evolutionary biology is unified by its recurrent use of explanations that cite measurable effects of environmental or other selection on the distribution of varying and heritable properties within populations. A biologist does not, *qua* biologist, query the cogency of this sort of explanation in general, because accepting the soundness of its generic logic and its general ontological appropriateness is part of what makes her a biologist. We need not here endorse all the details of Kitcher's analysis in agreeing that this idea identifies one plausible element of the vector of (soft) constraints on explanatory unification. Over the course of his recently truncated career, the late Wesley Salmon explored another element of the vector, one lying more clearly and directly in the metaphysical tradition that seeks a basis for ontological monism in one fundamental kind of "stuff." That is, Salmon en-

deavored to show something enlightening about the ontologies of all sciences by reference to general microstructural relations that bind all real objects and processes. In the philosophy of science literature, Kitcher's and Salmon's approaches are taken as offering rival bases for identifying good scientific explanations in the shared context of scientific realism.[11] We agree with Salmon's (1990) view that although neither his approach nor Kitcher's may furnish a complete and ultimate analysis of explanation, they form a complementary pair of answers to a general question about what science wants and needs from philosophy of science.

In the context of our response to Kim here, we will be following a road that Kitcher and Salmon have mapped quite explicitly in dialogue with one another. Kitcher (1989) characterizes his work as analyzing "top-down" explanation, wherein we explain phenomena by fixing their roles in wider ensembles of regularities, and he contrasts this with "bottom-up" explanation, the sort analyzed by Salmon, which consists in identifying the causal-mechanical processes that generate a phenomenon being explained. Salmon (1990) endorses this idea of a "duality" of explanatory approaches, which he takes to apply across the board. Thus, to cite one of Salmon's examples, we provide a top-down explanation of industrial melanism in peppered moths by means of the familiar story embedded in population genetics and evolutionary ecology, and we would furnish a bottom-up account to supplement it if we added facts about the synthesis of proteins that lead up to the production of differently colored wings.

As will be clear from our discussion in the previous sections, Kim can be happy enough with this sort of duality in *explanation.* His difficulties turn on the fact that, according to his analysis, top-down accounts of the Kitcherian sort cannot be causal. Neither Kitcher nor Salmon would necessarily disagree with this, because the duality they endorse is epistemological rather than metaphysical. However, many typical explanations in the behavioral and cognitive sciences seem to be simultaneously top-down *and* causal.

Consider the following example, based on Hutchins (1995) that echoes many others found in the current cognitive-science literature on intentional action. Some navigation systems on large ships require two specialist "pelorus operators," one on either side of the ship, each reporting, with the aid of a special instrument (the pelorus), the angular position, or bearing, of visible landmarks. Pelorus operators do not select the landmarks; instead, they are specified by other members of the navigation team. Imagine a pelorus operator, recently ordered to "stand by to mark" the bearing of a particular landmark, and so expecting imminently to be asked to report the continually changing bearing upon being ordered to "mark." The actual response to the "mark" instruction will be constituted by a series of neural, nervous, and muscular events that the pelorus operator cannot directly access for description to himself, or subsequently report as distinct from one another (even if he knows on theoretical grounds that they must have been).

The operator's actions – including adjusting the orientation of the pelorus, maintaining a state of readiness to report the current bearing by frequently consulting the apparatus and what is visible through it, and rehearsing and reporting the reading – will largely consist of pre-prepared subroutines that can be executed as relatively autonomous wholes. These subroutines will be the product of training, guided by personal habits and primed by ritualized social cues. Some subroutines will be specialized at gathering information from the world (reading the instrument, decoding instructions about landmarks), some at controlling the information gatherers (lining the apparatus up on a landmark, given an external instruction), some at producing responses according to strict conventions (reporting the bearing when instructed to "mark," inter alia by producing the required phonemes in the required order). Others still will co-regulate the activity of those already mentioned – preventing the reporting system from being executed until the "mark" instruction has been decoded, and so on. The routines will therefore partly be coded as dispositions in particular synaptic firing pathways, amenable to being triggered by some small subset of those synapses.

Further, the pelorus operator's *entire brain* must, on balance, be so configured that the output of the instrument reading subroutines, when released by the decoded "mark" command, controls reporting behavior, preventing him from becoming enraged when remembering that "Mark" is also the name of a romantic rival, or abandoning his station to tie a shoelace, and so on. He must instead be neurally primed to check and report the bearing at the moment of hearing the mark command, and do nothing else. So, there's the setup. And now, action! The command "mark" is uttered and decoded, the visual position relative to the calibrations on the instrument consulted, the markings transduced and processed, the result slotted into the conventional template, the phonemes rehearsed and uttered "[Landmark X] 237."

This explanation is relentlessly causal, but it is very far from strictly bottom-up. The "subroutines" to which we casually referred are black boxes, top-down characterizations of networks of connections that include both triggers and inhibitory links. At every stage, we picked out these black boxes as pure role-fillers by reference to a rich conceptual network that we already know the operator must have learned. Perhaps, though, we were just being lazy, or deferring to our own ignorance: if we could have provided the whole bottom-up story, individual electrochemical event by individual event, would not we then have provided the *exclusive* causal story? Let us postpone that question for now. Notice that even if a full specification of ordered synaptic potentials *is* the exclusive causal story, then, as functionalists of all sorts have long emphasized, reciting the specification would be a poor explanation of what happened, because there is nothing systematically special about *these* particular synaptic sequences that ties them to bearing reports from one occasion to the next. Furthermore, at one point we had to cite the dispositional state of the pelorus operator's *entire brain*! But this state will likely never occur again, exactly, no matter how long the operator's career or how many bearings he reports. And knowing the state in one case would do very little to illuminate different cases: What would those neurons, let alone the operator, have been doing were an alternative landmark to have been specified? Or were the "mark" command to have been given a moment later? Thus, the strictly synaptic account would miss almost all of the counterfactuals relevant to behavioral explanation. The fundamental basis for this is the servosystematic nature of the control architecture at work here. If some synaptic paths wander away from the central task, then feedback generated from other regions concerned with attentional focus will quickly recruit backup or alternative resources. Restricting explanation to the *actual* microcausal chain misses this structural fact.

The account of the pelorus operator's action given above is an instance of what Jackson and Pettit (1988; 1990) call *program explanation.* Social pressures operating on the pelorus operator ensure that one of many possible overall configurations of his brain that keep him focused on his task will (likely) be in place at the moment for action looms. This in turn "programs for" one suitable chain of synaptic events or another, by virtue of the feedback mechanisms through which brains embedded in environments control behavior in general. Here is what Kim says about program explanation. First, he invokes one half of Salmon's duality in asserting that "to explain an event is to provide some information about its causal history." Then

> what can be done is to define, say, the "causal network" of an event which is closed under both causal dependence and its converse, and then explain the idea of explanation in terms of providing information about the causal network in which an event is embedded. Pointing to an epiphenomenon of a true cause of an event [thus] does give some causal information about the event. (Kim 1998, p. 76)

In offering this analysis, Kim does not disagree with Jackson and Pettit themselves. According to them, the "programming for" relation provides "causally relevant information" but is not itself a causal relation. That is, to them, knowing about the pelorus operator's role plus the changing position of the landmarks tells us that *some* causal process sufficient for a bearing of "237" being reported will unfold, but not *which* one.

Suppose a scientist explains an animal's hunting by saying that it is hungry – in advance of knowing enough to have "sundered" hunger by reduction into $\text{hunger}_{\text{lion}}$, $\text{hunger}_{\text{mantis}}$, $\text{hunger}_{\text{snake}}$, and so on. She would then be giving us a program explanation of the hunting. Based on his remarks just quoted, Kim would concede that this explanation "gives some causal information." Furthermore, he seems to have no grounds for denying that it gives the *right* causal information so far as prediction and generalization are concerned; for, as the example of the pelorus operator is supposed to show, the program explanation supports the relevant counterfactuals. So why are we supposed to still be worried about the causal exclusion problem? Here's why:

> I believe it is only this sort of extremely relaxed, loose notion of explanation that can accommodate Jackson and Pettit's program explanations. Explanation is a pretty loose and elastic notion – essentially as loose and elastic as the underlying notions of understanding and making something intelligible – and no one should legislate what counts and doesn't count as explanation, excepting only this, namely that when we speak of "causal explanation" we should insist . . . that what is invoked as a cause really be a cause of whatever it is that is being explained. (Kim 1998, p. 76)

Implicit in this response is a metaphysical restriction on what sorts of states can and cannot figure in "real" causal explanations. Kim interprets minimal physicalism (that is, commitment to the causal closure of the physical) as requiring that all properties that cause things must *be* (perhaps by reductive identification) physical properties. This, of course, invites us to ask what makes a property "physical." Kim does not provide an analysis, but merely a recursive restriction that ties the physical to the "micro." That is: "First, any entity aggregated out of physical entities is physical; second, any property that is formed as micro-based properties in terms of entities and properties in the physical domain is physical; third, any property defined as a second-order property over physical properties is physical" (Kim 1998, pp. 114–15). Then the idea is that as long as the domain of "real" causal explanations is restricted to explanations that cite only micro-based properties, we are guaranteed never to violate the principle that physics is causally closed. Now we want to know what "micro-based" means. Here is Kim's definition of a micro-based property:

> *P* is a *micro-based property* just in case *P* is the property of being completely decomposable into nonoverlapping proper parts $a_1, a_2, \ldots, a_n$, such that $P_1(a_1), P_2(a_2), \ldots, P_n(a_n)$, and $R(a_1, \ldots, a_n)$. (Kim 1998, p. 84)

Micro-based properties are thus macroproperties that are not shared by the micro-constituents of the macro-systems that bear or instantiate them. So $\text{hunger}_{\text{lion}}$ could be a macroproperty, though hunger in general presumably could not (see sect. 3.3).

Thus, in Kim's view, whatever macroproperties "really cause" molar behavior must be decomposable into individual, nervous system–based, properties. This does *not* amount to the absurd thesis that all *causal powers* at the macro-level *are* actually microproperties; Kim knows that cars can get people down the street whereas parts of cars cannot. Rather, what he is committed to is the thesis that a system's causally effective macroproperties derive their effectiveness entirely from interactions among causally effective microproperties that are both regular and intrinsic to the same system.[12] He answers worries about radical multiple realizability of mental properties, with which both parts of this commitment are inconsistent, by suggesting that the possibility of practically interesting psychology shows that, as a matter of fact, multiple realization is not out of hand:

> The idea that psychology is physically realized is the idea that it is the physical properties of the realizers of psychological states that generate psychological regularities and underlie psychological explanations. Given an extreme diversity, and heterogeneity of realization, it would no longer be interesting or worthwhile to look for neural realizers of mental states for every human being at every moment of his/her existence. If psychology as a science were possible under these circumstances, that would be due to a massive and miraculous set of coincidences. (Kim 1998, pp. 94–95)

Many cognitive scientists who see program explanations as playing ubiquitous and irreducible roles in their domain (along with those of other special sciences) do *not* agree that it must be "the physical properties of the realizers of psychological states that generate psychological regularities and underlie psychological explanations." Most will likely concede that similar neurophysiology (and other physiology) from one individual to the next makes it possible for people to share comparable natural capacities, saliences, and learning histories, which is a necessary etiological condition for cultural learning. However, the operations of the natural devices that do this learning are not *equivalent* to our molar selves. Mental states are individuated by a process of triangulating under equilibrating pressure from similarity of cognitive and perceptual apparatus, similarity of social pressures on our histories of self-construction, and shared ecologies (especially social ecologies).[13] The basis for an interpretation of some set of synaptic potentials in the pelorus operator's brain as being "the state of believing that the bearing to the landmark was 237 degrees at the time he was ordered to mark" is, *in part*, reference to his history as someone conditioned to perform social roles,

and, in particular, a role in a practice that has such-and-such conventions.

Explanations of this triangulating kind are pervasive enough across the behavioral sciences that their genus constitutes a recognizable Kitcherian argument pattern. We identify hunger-states by triangulating amongst physiological, ethological, and evolutionary-ecological factors; and then we furnish explanations of particular events in animal lives by supposing that hunger programs for displays of search and consumption. We identify productive activities in economics by triangulating amongst considerations of energetic output, behaviorally derived utility functions, and culturally evolved rules of exchange; and then we try to explain particular decisions of firms by supposing that production-possibility frontiers and profit-maximization functions (given some cost of capital) program for the appearance at particular prices of goods on the market. There has been no shortage of attempts to rigorously ground this loose argument pattern of triangulation in a generic but rigorous common logic – dynamic game theory, in which any of a variety of selection mechanisms sifting amongst rival strategies for allocating scarce resources lead to predictable shifts in the distribution of behavioral tendencies, is the current favorite candidate (Gintis 2000; Ross 2001). In these respects, the behavioral and cognitive sciences look no obviously worse off, no intrinsically less unified as a suite, than the various wings of physics and chemistry taken as a group. But Kim's contention that special sciences are only genuinely explanatory if they can survive a reductionistic re-interpretation does *not* depend on his finding that their typical explanatory attempts fail Kitcher's criteria. Clearly, for Kim, the unifying strategy championed by Salmon trumps Kitcher's: the epistemological duality is not mirrored at the ontological level. Scientists cannot reasonably be expected to share this intuition, however, and throw away what look like powerful explanations from one leading philosophical perspective, unless serious, professional-class metaphysical arguments show that Salmon was more obviously holding trumping aces than even Salmon himself thought.

We thus best press at the strength of the basis for Kim's "hyper-Salmonian" intuitions about explanation by asking, first, how they are supposed to make sense of actual explanations in the behavioral and cognitive sciences, and then, if that strains the prospects for accommodation, inquiring into the persuasiveness of their roots in general metaphysical analysis by itself.

The second of these tasks is taken up in section 4. In pursuit of the first question, let us first note that the triangulational approach to the individuation of mental states in psychology is compatible with two possible situations where the macro-micro relation is concerned. On the one hand, mentalistic psychology and neurophysiology might employ typologies that cross-classify across their putative micro-bases. In Kim's words:

> To say that a given taxonomic system cross-classifies another must mean something like this: there are items that are classified in the same way, and cannot be distinguished, by the second taxonomy (that is, indiscernible in respect of properties recognized in this taxonomy) but that are classified differently according to the first taxonomy (that is, discernible in respect of properties recognized in that taxonomy), and perhaps vice versa. That is, a taxonomy cross-classifies another just in case the former makes distinctions that cannot be made by the latter (and perhaps also conversely). (Kim 1998, pp. 68–69)

According to Kim, this amounts to a denial of supervenience as a one-way relation, permitting what Meyering (2000) calls "multiple supervenience" (see sect. 3.2). Kim says that "this is a serious form of dualism, perhaps an approach worthy of serious consideration." Kim's two uses of "serious" here must prevent us from regarding this as name-calling. On the other hand, we really do not think that "dualism" is quite the apt word here, because in this context it is clearly supposed to indicate views that deny the causal closure of physics. We will indicate reasons for doubting that acknowledgment of multiple supervenience implies such dualism, after first indicating just what multiple supervenience amounts to and why special sciences constantly traffic in it.

3.2. Multiple supervenience and special-science explanation

Meyering (2000) introduces the concept of multiple supervenience by means of an analogy with dispositional explanation, and referring to the imagined example of Mary, electrocuted while atop an aluminum ladder[14]:

> Dispositions, just like macro-properties, fail to produce causal effects independently of their categorical base. And yet their explanatory power clearly differs from, or exceeds, that of their bases. This becomes intelligible when we recognize that one and the same categorical base "realizes" more than one disposition. Even so, only one of those is usually relevant for a given event. Thus Mary's death is related to the electrical conductivity of her aluminum ladder. But the categorical base thereof (the cloud of free electrons permeating the metal) also "realizes" such diverse dispositions as the thermal conductivity or the opacity of the metal. (p. 191)

The key point here is that the categorical base on its own, given that it realizes more than one disposition, plays a less effective role in an explanation than does one particular disposition it realizes. Referring to the realizer is insufficiently precise compared to citation of the disposition, or role. So the "actual realizer state is not merely inessential because a different state might have realized the same causal role. Rather, it is inessential because the *very same realizer state* may yield a wide range of very different causal trajectories" (Meyering 2000, p. 193). (This nicely exemplifies why even Salmon came to recognize the need for duality in explanation: Kitcherian top-down explanations are often more informative than bottom-up ones, and *objective* informativeness is surely a *metaphysically serious* aspect of explanation on any reasonable account.) One way of describing the state of affairs Meyering considers is to say that there are supervenience relations (i.e., relations of covariance plus dependence) going in two directions at once here. On the one hand, the disposition supervenes on a particular set of microproperties, but the disposition could be realized by different micro-arrangements. On the other, the relevant microproperties realize multiple dispositions, and if a given disposition is picked out in *relational* terms, it turns out to supervene on the system of macro-relations. (The earliest explicit appearance of this idea in the literature is Dennett (1981), who argues that explanations in cognitive science often rely on "macroreductions.")

If one acknowledges the possibility of multiple supervenience, then one disagrees with Kim's supposition that all supervenience relations point unidirectionally to physics. This might suggest a basis for a quick answer to Kim's su-

pervenience argument, because if you reject its implicit premise that supervenience relations must all be "downward," then you will not get impaled on the first horn of Kim's dilemma (see sect. 2.1 above), because *this* kind of breakdown of supervenience has no consequences at all for the causal closure of the physical. To clarify this last claim, multiple supervenience does not imply the spooky idea that you could change the global psychological state of the world while making no physical changes *at all.* But it does imply that, even given ideal science, you could not necessarily predict which *particular* physical changes would have to accompany a given psychological change; that is, that these relations are not, in general, systematic. Avoiding Kim simply by abandoning supervenience, though, wins a cheap victory by burying more substantial issues at stake. Kim would presumably deny that an explanation citing upward-supervenient dispositions can be a *causal* explanation; and Jackson and Pettit, in shying away from regarding program explanations as causal, presumably agree about this. This brings us to what we think is the deepest bedrock beneath the new metaphysical unease with the special sciences, with which we grapple in section 4.

Meanwhile, however, let us press on by asking what the special sciences actually *do* that leads them to pick out entities, processes, and kinds that do not end up in neat supervenience relations with physics. Meyering offers the following suggestion:

> What gets studied in the special sciences is in fact huge systems of concatenated micro-systems which are systematically organized in such a way that their typical causal antecedents prompt typical patterns of causal processing to eventuate in typical effects, which in their turn serve as typical inputs for yet other causal sequences of events to take place. Regimented in this way the system produces emergent effects that have no salience at the level of physics, and yet constitute the preconditions for the recurrence of the sequence in question, or for the emergence of related processes which are significant at that same level of special science description. (2000, pp. 193–4)

For an example of "emergent effects which have no salience at the level of physics," consider the huge collection of physical particulars that happen to constitute a given stock market crash. Such an event is clearly of considerable importance to the group of special sciences we call economics. The claim being made here is that *without* the perspective provided by the special-science explanations in question there would be no way of picking out *that* collection of particulars as being an event at all. It just wouldn't be on anyone's list of "things to be explained," any more than the particular things counting as "money" would cry out to be classified together on grounds recognizable to physics. So, it would appear, if you want to have descriptions, let alone *explanations,* of phenomena where functional, and especially multiple, supervenience obtains, then you need to grant the irreducibility of the kinds which feature in such explanations. To be blunter still, we are faced with a choice between embracing reductionism, or being able to construct the explanations we do in fact construct.

Faced with this choice, some thinkers have supposed that there just *cannot* be anything wrong with our apparently causal explanations, and hence that Kim just has to be wrong. One version of this response argues that if Kim is correct about the mental causal exclusion problem (sect. 2.1 above), then all of the special sciences are in the same trouble. Taking it as more or less self-evident that *that* cannot be the case, they reason that Kim's problem is not a real problem at all. Burge (1993), Baker (1993), and Van Gulick (1993) all offer versions of this "generalization argument." Kim's response is twofold: he argues that if the problem *did* generalize, to reason that there is not a problem because we find the conclusion outrageous amounts to demanding a metaphysical free lunch, and he argues that the problem does not, in fact, generalize very far.

If it seems like the causal exclusion problem *should* generalize, it is because the supervenience argument looks like it should apply to *any* nonphysical property, including chemical, geological, biological, and other special-science properties. In the limit, this suggests that *all* causation should "seep down" to the level of microphysics. Kim argues that this supposition trades on vague intuitions about a hierarchy of "levels" of properties, which need to be handled more rigorously. Specifically, he argues that we should distinguish between the realization relation and the macro-micro relation, and, having done so, recognize that the "*realization relation does not track the macro-micro relation*" for the reason that both "*second-order properties*[15] and their first-order realizers are properties of the same entities and systems" (Kim 1998, p. 82). To supplement this argument Kim develops a notion of a "micro-based" property (discussed above in sect. 3.1 and below in sect. 4.2), so as to save the physical status of "micro-based macroproperties" such as hardness, transparency, conductivity, and the like, as well as the objects in which we standardly locate them such as tables, windows, and nerves. Kim's reflections here are, we think, partly salutary: physicalism "need not be, and should not be, identified with micro-physicalism" (1998, p. 117). (Clapp [2001] develops independent arguments against misleading "level" talk, in the context of a defense of nonreductive physicalism.)

Marras (2000) argues, however, that Kim's attempt to limit the extent to which the causal exclusion problem generalizes is of limited success. At best Kim's arguments show that the causal exclusion problem is not an *inter*-level problem, indicating that the only causation is microphysical. What his arguments do not show is that it is not an *intra*-level problem for every individual special science. The possibility left open by Kim is that every special science is ontologically confused, in virtue of classifying the world into types that cannot be reduced to physics. In the light of what has been said above, it should be clear that the causal exclusion problem generalizes, at least, to every case of multiple realization of a functional or relational property. (In a complementary contribution, Clapp [2001] shows that Kim's argument has the "unsavory consequence" that it makes *all* multiply realized properties, including most paradigmatic physical ones, illegitimate partly *because* most properties are associated with causal/functional roles.) So the question how many of the special sciences are threatened by Kim's arguments is the question how many of them trade in multiply realized functional kinds. We think that *all* of them do, but this is not the place to defend this claim by means of an enumerative induction. A few examples, then, will have to do a lot of work. Consider water.

On Kim's view "being a water molecule" is a straightforward physical property, which he regards as the "micro-based" property of "having two hydrogen and one oxygen atom in such-and-such a bonding relationship" (1998, p. 84). This assertion is either false, or runs in the face of the practice of chemistry. A sample of liquid water does not

consist only of H_2O monomer molecules, but also, at any moment, of various polymeric molecules such as $(H_2O)_2$, and $(H_2O)_3$, in a condition of statistical equilibrium involving rapid reciprocating transformations (Millero 2001; Ponce 2003;[16] van Brakel 2000). If we allow polymeric forms of H_2O to count as water, then water is multiply realized, and Kim is simply wrong about what kind of property "being water" is. Further, and more importantly, what chemists recognize as procedures for determining sameness or heterogeneity of substance, or establishing whether something is a pure element or a compound, are a variety of tests of which the most crucial involve attempts to separate a sample into its different constituents, and to determine whether it is hylotropic under phase shifts (Needham 2002; Ponce 2003). These procedures track relational or dispositional properties – what it is that a sample *does* rather than what exactly it is made of. Following an account of these procedures, Ponce (2003, p. 145) concludes that "within chemical thermodynamics, what makes a given chemical kind a chemical substance rests in part on how the substance reacts (i.e. what the substance does) in certain experimental conditions, and what distinguishes one chemical substance from the next is not primarily its particular chemical composition or microstructure, but rather certain of its macroscopic, dispositional properties." Water is, perhaps, an especially telling example, just because if multiple realization operates at chemical scales, then it seems more likely to manifest at larger scales, where the smaller scale variability could be inherited.

This is definitely what we see in cell biology, where strict (Kim-style) reduction to molecular biology seems impossible because key biological phenomena such as "signal sequences" are multiply realized and context dependent, and because functional roles specified in biological terms are indispensable. As Kincaid (1997) argues, many different sequences of amino acids function as signals (multiple realization), but whether any given sequence does so is partly dependent on context (because the same sequences in other contexts *do not* play the signalling role – i.e., multiple supervenience), and, furthermore, "signal sequences" cannot be defined without reference to *biological* functions (see also Hull 1972).

No matter how far Kim's argument generalizes, though, we will not follow those who try to call for a free lunch. We are simply after the interim conclusion that Kim's problem, if it is a problem at all, affects almost all of the special sciences. It could well seem as though what is being argued for here is a kind of anthropocentrism, or pragmatism, where if something seems to *us* (or to chemists, biologists, etc.) like a good or powerful explanation, then, whether or not it is amenable to being reduced, it should be regarded as legitimate.[17] Realism about special-science types does not require any such abandonment of metaphysical seriousness, however.

Macroscopic states need be neither anthropocentric nor pragmatically justified if there is some way of making sense of their being real, in the sense of "real" which involves it not being up to us whether an ontology respecting Occam's razor would have to recognize them. Dennett (1991b), confronted with demands to take a position on whether ascriptions of beliefs should be thought of in realist terms, or as merely instrumentally justified devices, answered by offering a "mild realism" in which the reality of basic physical states was unproblematic, and in which macroscopic *patterns,* understood in information-theoretic terms as structures that encode non-redundant, objective information by means of compression, could be considered real enough to settle the debate. One of us (Ross 2000) has argued elsewhere that Dennett's position should be modified into a more thoroughgoing pattern-realism,[18] suggesting that a pattern should be considered objectively real if and only if:

1. It is projectible under at least one physically possible perspective and,

2. It encodes information about at least one structure of events or entities *S* where that encoding is more efficient, in information-theoretic terms, than the bit-map encoding of *S;* and where for at least one of the physically possible perspectives under which the pattern is projectible, there exists an aspect of *S* that cannot be tracked unless the encoding is recovered from the perspective in question (Ross 2000).

So considered, it is a contingent and empirical matter whether any particular real pattern is reducible to another, and, crucially, the question of the reality of any pattern is not to be decided on anthropocentric grounds. This is so because patterns are required to be projectible under a *physically possible* perspective, rather than a perspective that is an artifact of human perceptual or cognitive capacities. So if there is a physically possible perspective from which some phenomenon recognized by our current working ontology could be more efficiently represented under an alternative ontology, then our current ontology is false, regardless of whether we are or are not, or shall ever be, aware of the existence of the alternative possible perspective in question.

What realist special scientists do on this view, then, is seek to find real patterns in particular domains of reality, domains defined by sets of particular structures and/or processes at some level of abstraction from fundamental physics. These patterns are what Meyering needs to cash out his talk of "huge systems of concatenated micro-systems which are systematically organized in such a way that their typical causal antecedents prompt typical patterns of causal processing to eventuate in typical effects" (Meyering 2000).

A defender of Kim's line can object that what we have just said about explanation, and the irreducibility, indeed even the objective *reality,* of irreducible functional properties, does not automatically make any headway against the *causal* exclusion problem. It is, after all, in the name of solving that problem that we are supposed to "give up" on these irreducible properties. That is, it is just these properties that we are supposed to learn "to live without" so as to preserve a coherent and univocal concept of causation. Looked at this way, our banging the table and complaining about how difficult it would be to live without the properties is not a good answer to Kim at all.

It is true that the possibility of non-reductive realism about special-science types does not make *direct* headway, because it does not say anything yet about how to show that special-science generalizations invoking irreducible properties could be *really* causal. But it is more than a *mere* request for a free lunch, because it is crucial to showing what is at stake for the special sciences in evaluating the importance of Kim's argument. We maintain that Kim's position is based on serious misunderstandings about how things are in the special sciences, and to make our more direct argument against him we need to outline and defend what we take to be a more defensible picture. Because of his inaccurate picture of special sciences, Kim does not seem to

think the costs of his proposal are intolerable. We aim to show that they are utterly intolerable, requiring that we regard almost all explanatory activity in the special sciences as confused.

3.3. Stamp collecting

In section 2.2 above we briefly outlined Kim's reductionist proposal, which he urges as the proper response to his supervenience argument for the instability of non-reductive physicalism. His proposal involves "sundering" the types referred to in special-science explanations in accordance with the particular reductive bases for them we discover empirically. We argue now that this effectively urges us to abandon functionalism entirely, which goes against Kim's claim to the effect that his brand of reductionism is consistent with taking multiple realizability "seriously" (Kim 1998, p. 111).

Here, to recap, is why Kim's proposal is supposed to include elements of functionalism. The process of reduction he describes gets started with a role property (pain, say) and proceeds via the discovery of the particular physical realizers of that property to a series of reductive identifications, "sundering" the role into as many realizers as turn out to be empirically warranted. One immediate difficulty here is that without access to the role properties scientists would not know where to start looking for realizers, or what the realizers were supposed to be realizers *of.* That is, as we argued above, if they *started* from physical particulars and were prohibited from making reference to role properties, it is not clear that there would be any way at all for them to tell a collection of particulars that was the realizer of a functional property from one that was not, or to tell what manner of functional property it realized. This would be mission impossible: trying to look at some huge mass of physical detail, and hoping to be able to say at some point "Ah ha! It's a stock market crash, and it started at *this* moment, and the proper boundaries of the physical event constituting the crash are *here."* Kim's proposal, in other words, requires that his metaphysically justifiable types of science are parasitic on the very types he argues are epiphenomenal.

A defender of Kim's position may point out here that Kim does allow that by "grouping properties that share features of interest to us" it is possible that "important conceptual and epistemic needs" could be served (1998, p. 110). Perhaps, then, what we are calling parasitism is what he would call serving an important conceptual need. This is an unsatisfying answer, however, because it leaves those hunting real causal relationships using others whose work is epiphenomenal as trackers. It also makes clear, as we suggested above, that on Kim's view the only justification for functionally motivated special-science work is indeed instrumental: by doing that kind of science you help the reductionists figure out where to start digging, so as to dismantle the foundations of that very work. If we set this point aside and continue, matters only get worse for the special sciences of Kim's future world.

As we have seen, it is when empirical work turns up diversity in the realizers of some functional property that we are supposed to dismember the role property into its parts. Let us assume that Kim's hunch, or hope, that realizers are likely to turn out to be species specific is right – then perhaps we would sunder pain, irrespective of how well it paid its way as a single notion in behavioral science, into $pain_h$, $pain_m$, $pain_o$ (for, say, human, Martian, octopus). If we did this, we would be proceeding as though we had discovered (so far) that pain was actually *three* things. How, though, would we decide whether this were the case, or whether we had really found out that there were no such thing as pain in general? Or, perhaps, that only *one* of the three was pain (in which case *which* one?), and that the other two were something else? (See also Marras 2000; 2002.)

Looked at another way, had our scientists somehow managed, despite the parasitism worry noted above, to *start* with a set of realizers (not having to work out from raw physical data what is a realizer of a function and what is not), it is not at all obvious that they would group the realizers in the same way as they would group them given access to the role properties too. It could well be, for example, that the realizer of Martian flatulence was structurally more like the realizer of human pain than the two pain realizers were like one another. In this case scientists working with only a collection of empirical descriptions of realizers might be expected to group the realizers quite differently, if they were to group them at all. There would be nothing to stop them supposing, like neoplatonist medical thinkers, that walnuts might be therapeutic for some brain conditions because they *look* rather like brains. Again, it seems, reference to role properties by Kim's rules has to be an instrumental necessity arising from the fact that the reductive relations are unknown at the outset of any inquiry. Worse, the inquiry proceeds by making the role properties obsolete. If we call the three imagined realizers of the pain role different versions of pain, it seems we are doing so out of a kind of nostalgia for when we thought (if we ever did) that pain was in some sense *one* thing, rather than out of clear-headed recognition of what, by Kim's lights, we subsequently discovered.

The most important reason why the costs of going down Kim's road are prohibitively great is thus that it requires, in the end, giving up on the prospect of a unitary psychology, and in fact on any unified science referring to functionally individuated kinds. (Economics and biology are obvious instances.) As noted, Kim is willing to allow that by "grouping properties that share features of interest to us" it is possible that "important conceptual and epistemic needs" could be served. But he is also adamant that, in the end, functional properties with diverse realizers are properties "we will have to learn to live without" (1998, p. 106). In other words, the only justification for unitary sciences having as their objects functionally individuated kinds is *instrumental,* because multiply realized properties turn out to be not metaphysically acceptable.

To hammer this point home, let us examine a real example. Consider hunger and satiety. Hunger is multiply realized (perhaps, therefore, a property Kim thinks we may have to "live without"), by several mechanisms with distinct effects on different parts of the brain. We can be stimulated to eat by, inter alia, the mechanical sensation of an empty stomach, glucose level monitoring by the liver, the sight of others eating, the smell or taste of novel food, and stress, not to mention combinations of these and other factors. One of the various realizers of satiety, or of "contra-realizers" of hunger, it seems, is hormonal, particularly but not exclusively via peptides occurring in the gut. These play a role both in modulating other gut secretions, hence participating in the control of digestion, and in sending an

"enough" signal to the brain. Whether these hormones do in fact realize one or the other of these functions, or neither, or both, at any given time, depends on relational factors, so we here have a case of *multiple supervenience.* They might be active, yet we stop eating for other reasons (an artificially filled stomach triggered our mechanical sensors), or continue eating despite their action, perhaps because people around us are eating, or because we are anxious, or because the novel dessert is more attractive than the unfinished main course.

A special science which studies what Meyering describes as "huge systems of concatenated microsystems" or what we have suggested should be thought of as real *patterns,* has a shot at tracking typical patterns produced in consequence of the systematic organization of those systems. Such scientists get to make explanatory, and *predictively powerful* statements like:

> When dietary variety is produced by providing a meal or diet composed of several foods, animals generally become hyperphagic relative to single-food meals or diets. (Raynor & Epstein 2001)

At the risk of laboring one point we have been emphasizing, it is worth drawing attention to the term "variety" in the above quotation. Pattern-hunting special scientists, such as behavioral scientists interested in motivated behavior, in this case eating behavior, are able to justify broad-scope predictive generalizations referring to "variety." In humans, the variety in question is strikingly multi-modal, which is to say that the effect is stronger if the foods differ in more than one way, including taste, color, shape, smell, texture, and presentation. "Variety," though, just *has* to be multiply realized (there are different ways of being different) and multiply supervening (structural features of the food may ground various dispositions, only some of which contribute to "variety" in a given context), and so, by Kim's lights, it is one of those properties we're going to have to learn to live without.

That, however, just cannot be acceptable. We hereby bet the farm that any *possible* life form which metabolizes and is faced with resource scarcity will have *something,* and in all likelihood several things, playing the role of hunger, and that some of the generalizations of, inter alia, our psychology, ecology, and microeconomics, will apply to it.

Kim does not have a direct argument against multiple supervenience. It is off his radar insofar as it is more powerfully antireductionist than anything he seems willing to consider. We have not yet shown him to be wrong. What we have done is shown just how bad it would be for the special sciences were Kim's position to be generally endorsed. Now we need to look at how to disarm his argument for the view that anything much needs to be changed in the special sciences at all.

4. Taking metaphysics seriously

We have stressed repeatedly that answering Kim requires taking metaphysics seriously. This does *not* mean respecting any particular *a priori* hunch about the objects of any special science. Rather, it means acknowledging such demands on the structure of scientific inquiry as transcend the disciplinary boundaries of individual special sciences, with the aim of productively applying these demands to guide interpretations of the relationships amongst hypotheses generated across separate sciences. Part of our diagnosis of what is wrong with Kim's approach is that he is mistaken about the relationship between one metaphysical problem and the work of physics, so we begin this section in 4.1 by distinguishing a number of metaphysical questions relevant to the issues at hand. In section 4.2 we return to multiple supervenience, and the related questions of what to count as physical, and how to draw the macro-micro distinction in a way consistent with the account of realism offered in section 3.2. In section 4.3 we distinguish two ways in which *cause* has been understood in the history of philosophy, and argue that Kim equivocates between them. Finally, in section 4.4 we argue that Kim erroneously supposes that physics provides us with the answer to a metaphysical question, and furthermore that he is seriously mistaken about how things are with physics.

4.1. What metaphysics demands (or: How to pay for lunch)

Clearly the metaphysical question bothering Kim is the following: *What explains the fact that the supervenience relations that do in fact hold, hold at all?* Kim thinks the answer to that question would be a solution to the causal exclusion problem (sect. 2.1 above); and his own reductionist proposal (sect. 2.2) is supposed to show how the supervenience relations hold because some stronger, reductive relationship holds between physical facts and functional (special science) facts: the causal capacity of special-science properties can be inherited from the unproblematic causal capacities of the physical properties with which we find they are identical. We agree with Kim that mere *invocation* of supervenience cannot answer the metaphysician's question about the place of mind in a physical world. If the special sciences that deal in supervenient types are not to be isolated from the rest of our scientific ontology, we must indeed be able to explain why the particular supervenience relations (both general and specific) that in fact hold, hold at all.

As just indicated, the causal exclusion problem, considered very generally, is a problem about the *unity* of our scientific worldview, as briefly introduced in section 1.2 above. In the context of the naturalistic, broadly empiricist conception of knowledge and reality presupposed here, the task of the metaphysician, if she has any task at all, is to systematically investigate the ways in which relatively separated and special tracks of scientific inquiry "hang together" to imply a whole greater than the sum of their respective parts. This is important not just because people like having unified world-views. Principled, if always necessarily tentative, answers to metaphysical questions are required to help scientists make sensible bets on which special-science kinds they should be trying to explain and which ones they would be better advised to try to explain *away.* The study of unification as a distinctive enterprise can be predicted to covary in importance with the extent to which individual sciences develop specially. In the heyday of positivism, the demand for unification was typically given the strongest possible reading by philosophers who supposed that special-science generalizations should be logically derivable from more fundamental generalizations, and/or that all special-science types should be logically constructible from fundamental types and relations. Insisting on such "strong unification" amounts to asserting reductionism as a meta-

physical hypothesis, which is just what, in disagreeing with Kim, we are here rejecting. As indicated above, the history and practice of actual sciences, especially the behavioral, cognitive, and life sciences, honors no such reductionist constraint. However, the history and practice of science *does* demonstrate consistent concern for unification in a weaker sense. To the extent that the conclusions of a given special science are isolated from those of all other special sciences, in the sense that their generalizations and/or ontological typologies are strictly "brute facts" from all available exogenous perspectives, we find ourselves with a mystery or set of mysteries (Friedman 1974), and science is never content with mysteries.

We can try to describe the generic ambition for unification a bit more precisely by distinguishing three specific kinds of project that might collectively constitute it:

1. Identifying a unifying ontological structure that justifies the argument patterns accepted across all of the sciences.
2. Saying something genuinely enlightening about the ontologies of all sciences by reference to general structural relations of some kind.
3. Identifying the "glue" that holds all objective relations in place.

Notice that none of the three metaphysical problems we have identified is necessarily about *causation* although all can be read as having something to do with it. For the time being (although see sect. 4.3 below) we remain agnostic about whether the "glue" might be something worth calling causation. In recent philosophy of science the first problem has been most strikingly associated with the work of Philip Kitcher, the latter two with that of Wesley Salmon. We discuss relevant details of their respective positions shortly.

Talk of "binding ontologies together," or of the metaphysician's "universal glue" is unabashedly metaphorical. Positivism was, among other things, an attempt to explicate unification without resorting to superficial metaphor, but like most similarly motivated projects in the history of philosophy, it failed because it committed itself to claims that were too strong and specific to fit the full complexity of actual science. We will not, then, be able to avoid metaphor here – "glue," indeed! – in trying to say what metaphysical explanation aims at. What we can do, and will, is as far as possible allow the dominant analyses in recent philosophy of science (Salmon's and Kitcher's) to constrain what it is that we do with our metaphor.

4.2. Supervenience and physical causation

We saw in section 3.2 how the prospect of a breakdown of ubiquitous one-way supervenience struck Kim as tantamount to dualism. That should not be surprising, because we also saw in section 2.1 that the first horn of the dilemma forming the supervenience argument has it that denying supervenience just *is* denying the causal closure of physics. Kim demands something stronger than general commitment to supervenience in the form of a principle requiring that there be "no changes without physical change," though. He wants (see sect. 1.1) "narrow" supervenience, where the supervening properties of some entity must supervene on its internal, or intrinsic properties. Failures of *this* kind of supervenience do not by themselves imply anything about whether physics is causally closed. One way such failures can arise, consistently with the closure principle, is from cross-classifying taxonomies, which in turn can arise from triangulational individuation of mental states, as discussed in section 3.1 above.

Though triangulational individuation is *compatible* with cross-classification, it does not imply it. Social-ecological properties relevant to mental state individuation in the case of the pelorus operator (sect. 3.1) *could* be micro-based in Kim's sense. Perhaps cognitive scientists could work adequately with a system of mental-state classification sensitive to two or three micro-based taxonomies of properties among which it banned conflicts with its own coherence rules.[19] However, cognitive scientists are just not, as a matter of fact, trying to regiment their macroproperties in the way relevant to this scheme. (As Wallace [2003] argues, physicists faced with the logically identical issue in relating quantum-level properties to macroproperties do not try this either.) So this apparent possibility for reconciling Kim with cognitive science is worth pursuing only if Kim's independent metaphysical motivations for needing some such reconciliation are truly pressing. As we now argue, they are not.

Kim provides no direct analysis of the concept of a physical property. Instead, as we have seen, he assumes that the domain of physical properties is antecedently clear, and then analyzes putative *non*-physical properties as micro-based macroproperties. Cognitive and behavioral scientists might imagine that this way of proceeding reflects consensus among metaphysicians, appealing to some well-established analysis of the physical. This is not the case. Recall, first, that the distinction between "top-down" and "bottom-up" prevalent in philosophy of science is drawn – by Kitcher and Salmon – by reference to the logic of *explanation,* not by appeal to a brute concept of the "physical." We need to ask, then, what makes something a macro-state, relative to some other set of states that are micro-states. It would be circular in this context to say that M is a macro-state relative to micro-states $m_1, \ldots, m_k$ just in case M is specified in terms of properties that supervene on the properties in terms of which $m_1, \ldots, m_k$ are specified; and Kim, given his project of showing that supervenience does not explain the relationship between the mental and the physical, would have to agree. Because we want to test Kim's picture against the prevailing metaphysic in general philosophy of science, we need to relate the macro-micro distinction directly to differences in kinds of explanations. This can be done following Kim's own lead, as given in his remarks on the relationship between explanation and causation quoted in section 3.1 above. Let us say that M is a macro-state relative to $m_1, \ldots, m_k$ just in case the (mere) information that M obtains fails to carry information picking out a particular member of $m_1, \ldots, m_k$ as causally relevant. Let us add that M is a *scientifically reputable* macro-state just in case restrictions on the set of micro-states, one or more of which must have obtained, can be stated in a scientific vocabulary more general than that of the special science that generalizes over states of type M. The point of this way of restricting the scientifically reputable macro-states is to reflect the weak unification requirement that special sciences cannot be metaphysically comfortable in complete isolation. Thus: an individual's performing an action that constitutes a move in a social game (e.g., the perolus operator's uttering "[Landmark x] 237") carries the information that some set of dispositions selecting the relevant action, encoded by the potentials along some synaptic pathways in that individual's

brain, was available to be triggered, and was in fact triggered, by some state of affairs encoded as an instance defined according to the rules of the social interaction by some other set of synaptic pathways in that same individual's brain. By virtue of what might knowledge of the social action carry information about such generic sorts of brain processes? By virtue of the actual and particular content of some empirical theory of mental architecture and its relations to neural structures, on the one hand, and behavior, on the other. Similarly, to pick up the final example from section 3.3, to say that "hungry things are more likely to eat" compresses information about a range of multiply realizable states and mechanisms, and is arguably not further compressible.

Notice that this way of analyzing the macro-micro relation is strictly relative to a particular special-science context; we have said nothing yet about what might make some state or property "intrinsically" or "absolutely" micro. This is because commitment to the non-isolation of special sciences does not imply commitment to the idea that all special sciences admit of hierarchical analysis in terms of one basic science. Kim, however, must suppose that there are "intrinsically" micro-states, since only this could justify his implicit restriction on scientifically reputable macro-states being, as it is, stronger than the one just given. The point is not that he must suppose that psychological states reduce directly to such states; rather, the claim is that if there are no such states to end a potential regress and do the "real" causal work, then Kim's supervenience argument would lead to an antinomy rather than to a disjunction with a preferred horn as he supposes. Our key question, then, is: *Does (serious) metaphysics lend support to this intuition*? Nonphilosophers might be disappointed, although not surprised, to hear that the answer is complicated. It requires examining some details of the tensions and complementarities in the two generic perspectives, the Kitcherian and the Salmonian, on the scientific realist's epistemology.

Kim and his supporters can, prima facie, draw strong support from Salmon (1984; 1999), whose most general goal is to articulate and defend a realist interpretation of the point and nature of science. In particular, according to Salmon, science aims to describe the causal structure of the world. In the end, we will raise grounds for doubting that "causal" is an unambiguously perspicacious word here. So let us say for now that the essence of this sort of realism seems to us to be crucial to any sort of realism worth having, and describe that essence thus: Science aims to tell us how the world is structured, that is, how its various processes and classes of entities constitute a single working machine.[20] In trying to describe how a machine works, a natural approach is to try to lay out its various internal processes and indicate how they influence each other. Salmon aims to justify a picture of science that, as a whole, is engaged in this project. It is a virtue of such ambitious realism that it must go beyond mere affirmation of an independently existing world and wrestle seriously with Hume's epistemological challenge, to wit: How could anyone know, by any amount of observation, which links between processes are causal and which are not? Salmon's answer here is that we can observe something that is precisely *diagnostic* of causation. That is, we can see that certain processes transmit information about their antecedent stages whereas others do not. Only the former are genuine processes. Following Reichenbach (1957), we can put this in terms of the transmission of marks. In the absence of specific structure-preserving (and, ultimately, structure-constituting) activity, entropy will eliminate marks on objects that carry information about their histories. A structure is, by definition, something that resists entropy, even briefly. Therefore, wherever marks are preserved we have structure. The goal of science is to discover the structures in nature. We can discover such structures because, as fairly sophisticated information-transducing and processing systems, we can detect, record, and systematically measure mark-transmitting processes.

This is a terrifically powerful and, we think, deeply inspiring idea. It captures the core component of scientific realism – that science describes mind-independent natural structure (and activity) in an ontologically systematic way – while respecting the essence of empiricism. This latter constraint is that science has no place for inherently hypothetical events or processes that are in principle beyond our capacity to physically detect, for example, events on the other sides of space-like or time-like singularities, such as the interiors of black holes or the far side of the big bang, or events outside of our collective light-cone. One of us (Ross 2000) has exploited this idea to suggest a general metaphysic of existence; so we could hardly think it more important as a *metaphysical* insight. What, though, does it have to do with causation?

4.3. Two notions of causation

Salmon takes the idea described in the preceding section to be, first and foremost, an analysis of causation. Is it? As our remarks immediately above make clear, it is certainly an analysis of something[21] quite fundamental. But its primitive notion is information-transmission (in the physical and mathematical, not pragmatic, sense of "information"), not causation. It therefore amounts to a semantic proposal to treat causation as an information-theoretic concept. Should we accept the proposal? Because Salmon recognizes Hume's challenge to the effect that causation cannot be picked out by some observational procedure independent of the analysis itself, this evaluation must proceed pragmatically. What effect would accepting the semantic proposal have on our broader conception of science, and of particular sciences? In particular, will it justify Kim's intuitions about intrinsically micro-causal relations?

Kitcher (1989) provides a detailed critique of Salmon's analysis, which we will summarize. First, we must reproduce Kitcher's gloss of Salmon's analysis:

> (CP) P is a causal process if and only if there are space-time points c, e such that P links c and e and it is possible that there should be a modification of P (modifying a characteristic that would otherwise have remained uniform) produced at c by means of a single local interaction and that the modified characteristic should occur at all subsequent points from c to e without any subsequent interaction. (1989, p. 462)

This rests the idea of a causal process on the prior idea of a *causal interaction,* demanding an analysis of causal interactions in non-causal terms. Here is Salmon's analysis of causal interaction:

> (CI) Let P_1 and P_2 be two processes that intersect with one another at the space-time point S, which belongs to the histories of both. Let Q be a characteristic that process P_1 would exhibit throughout an interval (which includes subintervals on both sides of S in the history of P_1) if the intersection with P_2 did not

occur; let R be a characteristic that process P_2 would exhibit throughout an interval (which includes subintervals on both sides of S in the history of P_2) if the intersection with P_1 did not occur. Then the intersection of P_1 and P_2 at S constitutes a causal interaction if:

1. P_1 exhibits the characteristic Q before S, but it exhibits a modified characteristic Q' throughout an interval immediately following S; and
2. P_2 exhibits the characteristic R before S, but it exhibits a modified characteristic R' throughout an interval immediately following S. (Salmon 1984, p. 171).

We then have a case of causal interaction between P_1 and P_2 at t if and only if there exist S, Q, R at t satisfying CI. Two features of this analysis are crucial to Kitcher's criticism. First, it depends essentially on counterfactuals: we need to be able to pick out characteristics that *would have* carried on inertially in the absence of the interaction. Second, it makes the concept of a macro-level cause depend on the idea of a micro-level cause. (This is just what Kim assumes is unproblematic.)

These two features interact to generate the main criticism.[22] First, note that macro-processes typically involve vast ensembles of interactions. To use Kitcher's example, if a batted baseball breaks a window, then we have, along with the interaction between the bat and the ball, interactions between the ball and gusts of wind, the ball and changes in the Moon's gravitational field, these changes and the window, and so forth. We thus need to be able to pick out the *relevant* counterfactuals to identify the macro-cause, namely,

(A) If the bat had not intersected P_1 [the process that is the history of the ball's space-time coordinates], then the momentum of P_1 would have been different;

(B) If the momentum of P_1 after its intersection with the bat had been different, then the momentum of P_1 just prior to its intersection with P_2 (the window) would have been different;

(C) If the momentum of P_1 just prior to its intersection with P_2 had been different, then the momentum of P_1 just after the intersection would have been different (specifically, the window would not have broken!) (Kitcher 1989, p. 471).

But how do we know to pick out *these* counterfactuals? By reference, it would seem, to what we already know about the general causal structure of the world! Notice also that if we have these counterfactuals picked out, then we might be tempted to analyze the causal process just in terms of *them;* a detour through informational considerations would seem redundant.

A defender of Salmon could reply here that his analysis takes as its proper object only an *ideal* causal process, which would be a micro-process such that, given the restricted predicates available in its physical description, S is exhaustively and exclusively *defined* by a finite set of characteristics $Q, \ldots, Q'$ and $R, \ldots, R'$. Now, however, it seems that we must know the causal structure of the world in order to pick out the class of ideal interactions S.

We do not know if these technical problems can ultimately be solved. For our purposes here, this matters less than the general complaint Kitcher draws on the basis of them. That is: Salmon's analysis

> requires that we provide an account of the way in which the causal structure of the macroscopic world results from the stringing together of elementary processes. Even if we already had such an account, the emerging picture of our causal knowledge is one in which the justification of *recherché* theoretical claims about idealized processes seems to be fundamental and our ordinary causal knowledge derivative. (Kitcher 1989, p. 469)

Now, we do not think that this constitutes a serious objection to Salmon's substantive accomplishment, where that is interpreted as articulating the kinds of real structures in the world that science aims, in the limit, to discover. However, we *do* think that Kitcher's point has force against the idea that an analysis such as Salmon's, even if it can be made technically bulletproof with respect to its intended sphere of application in fundamental metaphysics, can be pressed into service as an analysis of the elaborately macroscopic, feedback-driven processes cognitive and behavioral scientists seek to characterize when they talk about mental causation, and the similarly complex causal patterns characteristic of science in general.

We will now present an alternative interpretation of Salmon's achievement, intended to shed light on what we see as the equivocal nature of the concept of causation. We take our cue here from Redhead's (1990; drawing on Kuhn 1971 and Russell 1917) discussion of causation and physics. Redhead notes that classical physics, in which forces played a crucial role, has given way to forms of physical theory in which forces have been eliminated. Redhead, we think justly, accuses those metaphysicians who wish to retain forces of anachronistically clinging to a distinction between natural and forced motion. In general relativity, says Redhead:

> There is no such thing as a non-natural motion. To most physicists the old-fashioned idea of cause arises from the idea of our interfering in the natural course of events, pushing and pulling objects to make them move and so on. In modern physics there are just regularities of one sort or another. (Redhead 1990, p. 147)

This attitude represents a principle that seems to us to be well justified by induction on the history of science. The central concepts of traditional metaphysics, including the Aristotelian distinction between natural and forced motion, are *folk* concepts. "Eliminativism," in the usual sense of that word in the philosophy of mind, is the thesis that folk concepts tend to be progressively eliminated from scientific practice. We do not endorse eliminativism in *that* sense. The concept of mind, for example, may enable us to pick out and generalize over real patterns in nature. Furthermore, the folk concept of agent causation may be *biologically necessary* in the sense that no functioning agent could get by without it. Science will therefore have things to tell us about both minds and agent causation. However, there is no compelling reason to think that folk intuitions about which patterns, if any, must be *general* should survive as the scope of scientific knowledge widens. The concepts and axioms of Euclidean geometry pick out and organize some real patterns – the class of (approximate) physical isosceles triangles, for example. But Euclidean geometry is not general, in the sense of describing most space adequately. We may gloss Redhead as suggesting that the concept of causation has its uses in describing the doings of agents, and perhaps in a range of other special inquiries, but that these uses do not generalize to physics.

Redhead is clearly asserting that metaphysicians *should not* use the concept of causation in talking about general physical relations. This is a stronger conclusion than we en-

dorse. Suppose that a scheme developed from Salmon's proves empirically adequate and logically perspicacious for bringing us closer to an analysis of the universal glue that metaphysicians seek. Suppose furthermore that Salmon's use of the term "causal structure" to describe what he is analyzing sticks, and not simply out of semantic inertia, but because – after all – the idea of analyzing causation as, at bottom, an informational relation is not *silly* or *pointless.* Then it would be right to say that the concept of causation had generalized. However, it would have done so along only one or a few of the dimensions that compose its historical semantic vector. Other such dimensions, those peculiar to the concept's origins in describing the interventions of agents, would have been discarded. Alternatively, we might end up (on a similar outcome in the philosophy of science) with "causation$_1$" and "causation$_2$." Our argument will not require a preference among these or other semantically plausible scenarios. The claim we need is merely a bit of conceptual history: that causation has its origins as a folk concept associated with agency, and that the concept as it figures in realist fundamental metaphysics, such as Salmon's, is intended to have no such associations, because it must shed them if it is to do the work Salmon wants from it. We will then see that Kim's challenge to functionalism depends on these very associations.

Before going further, it will help briefly to substantiate this conceptual history. Prior to the modern period we find no concept equivalent or isomorphic to the kind of causal notion analyzed by Salmon. Aristotle's efficient causation, considered apart from his wider metaphysic, maps best onto Redhead's "pushing and pulling" of objects by agents. The full Aristotelian story, with its multiply composed causes and deep teleology, is an elaboration of the folk notion modeled on the execution of a plan for intervention by an agent. The rise of science disturbed this picture. Famously, the rationalists were led to continual controversy amongst themselves over how to relate agent-causation to mechanical accounts; thus we have Descartes' immaterial will that nevertheless exerts mechanical effects, Malebranche's occasionalism, Leibniz's pre-established harmony, and so on. We suggest that Hume's attempt to analyze causation away is best interpreted in light of this history. Following Hobbes, but with much greater sophistication, Hume sought to explain all mental activity as mechanical (Ross 1991). Furthermore, on his account all mental activity had its ultimate impetus outside the mind, in the sources of impressions. Hume's was thus the most thoroughgoing denial of Aristotelian agency observed in Western philosophy to that point. But the elaborated folk notion of causation he inherited from his tradition was rooted in the idea of the intrinsically active agent. Attempting to drain *this* element out of the concept of causation, Hume found almost nothing positive left in it; and so, in his hands, it becomes merely a superstitious overinterpretation of regularity.

Hume thus set the philosophy of science along a trajectory that, with respect to its treatment of causation, finds maturity in the analyses of Reichenbach and Salmon. On this whiggish reading of the history, Kant represents a regressive step, trying to regiment the folk notion within the necessary operations of the understanding, and positivism a recovery of the Humean path from *within* the framework of Kantian metaphysics (Friedman 1999). Had this been the only major development in theories of causation after Hume, then it would be appropriate to describe the modern history of causation as a steady re-analysis away from the original folk notion and towards an idea – whatever its exact content – that could find its conceptual gravity wholly within the framework of fundamental physics. In evaluating efforts like Salmon's *as analyses of causation,* we would then be asking, in effect, whether the concept ultimately finds a role within that framework, or is fated for elimination.

However, this post-Humean development is not *all* that has happened to causation in recent philosophy. With the rejection of positivist and behaviorist accounts of mind in the 1960s and 1970s, functionalists reasserted the metaphysical significance of the mental, in a way that was not a reworking of Kant's attempted compromise with empiricism. Functionalism – when it does not drift toward epiphenomenalism – seeks to give the mental a real and distinctive causal role. Most important for present purposes, it understands that role in a way that resurrects the folk idea of causation, since the minds defended by functionalists are analyzed precisely as the ontological basis for agency. Thus, while one tradition within the philosophy of science continued the project of trying to drain the agency out of causation, a parallel department worked assiduously at putting it back in! The contemporary metaphysical muddles represented by Kim, against which we are taking issue, are consequences of this double development.

As discussed in section 1, functionalists have disagreed significantly amongst themselves over *how* mental causation could best be rehabilitated. The attributionist school of thought, following Dennett, has articulated a metaphysics of mind according to which the componential analysands of mind have been progressively distanced from the micro-causal gears of behavior – brains and nervous systems. This is the perspective whose consequences for behavioral and cognitive science are discussed in section 3. Philosophers in this camp can excuse themselves from any particular commitments with respect to the fundamental metaphysics of causation-in-general, *except* insofar as some particular account of such causation turns out to be essential for respecting weak unification constraints on all sciences. Describing as they are an unabashedly macroscopic set of phenomena, and with no ambitions in the direction of reductionism,[23] they can respond to demands for explanation of mental causation in the same way as any special scientist asked to discuss the kinds of processes picked out by the scope constraints of her discipline. Ask a geophysicist about "geological causation" and you will be told about tectonic plates and flows of undersea lava and so forth. Similarly, a functionalist cognitive scientist might address issues related to mental causation by talking about feedback mechanisms, servosystematic control architectures, modules built by natural selection, neural networks simulating Von Neumann computers, and so forth. To a philosopher who regards the topic of mental causation as *essentially* a part of fundamental metaphysics, such answers will look like cases of changing the subject. However, they are no more illegitimate than the geophysicist's similar answer to the similar question. The point here is just that a scientist's "taking metaphysics seriously" does not imply slavery to semantic legislation by metaphysicians. Saying that special sciences are sensitive to metaphysical issues is not to say that special sciences *are* exercises in (highly specific) metaphysical inquiry. If, at the end of the day, metaphysicians convinced us that the con-

cept of *causation* descended from Hume is more confusing than helpful, then either cognitive scientists will find other ways to talk about evolved macroscopic patterns in behavioral control built by interactions of genetic and cultural evolution, or, alternatively, we will collectively "decide" to let the Aristotelian semantic heritage triumph, regard cognitive scientists as having provided a naturalistic analysis of agent causation, and conclude that *causation* is a concept *restricted* to application in cognitive science and other disciplines that study agents. From the present state of play, it seems to us, either future semantic trajectory is possible; but neither threatens functionalism.

To return to the general point and summarize it, the history of philosophy incorporates a tension between two quite different notions of causation, both of which survive because both have been intended to serve legitimate but differing projects. On the one hand, special sciences are partly constituted by parochial types of "interaction-transmission" relations, where by such relations we refer to Salmon's "glue" without prejudging the details of the relationship between this and any causal concept as used in any particular special science. As Kitcher has emphasized, parochial, special science-relative varieties of the interaction-transmission relations are reciprocal functions of accepted explanatory schemata in the relevant sciences. Aristotelian agent causation, or folk psychological causation, is one such special interaction-transmission relation. Contemporary functionalism has significantly revised the Aristotelian or folk notion, particularly in denying the coherence of the idea of a unified "Cartesian" will with direct causal capacities of its own, but there is a clear lineage relationship nevertheless. Those whom Kim charges with being "free-lunchers" stop here, content to point out how useful this notion is. Kitcher's analysis of the importance of unification is one aspect of paying for lunch, in that it aims simultaneously at explaining and transcending (without abandoning) the cluster of parochial interaction-transmission relations. That is, Kitcher takes the existence of this cluster *as a fact for metaphysics to explain.* This is why Salmon, who was engaged directly in constructive metaphysics, can view his project and Kitcher's as complements.

Salmon's project, however, simultaneously continues to appeal to the *other* philosophical tradition with respect to the concept of causation and its use. As we explained above, the origin of this tradition lies in Hume's conviction that the folk concept of causation, as rooted in the kinds of program explanations peculiar to invocations of agent causation, fails to generalize. That concept therefore fails to be a suitable candidate for universal glue. Salmon is in pursuit of such glue, and this pursuit is the core of the serious metaphysician's job. What potentially confuses matters is that Salmon thinks his candidate for glue preserves enough traditional associations to make "causation" the appropriate name for it. We argued above that this is a semantic decision on his part. It is an understandable and not unreasonable decision, but it is not forced science and it does not commit a follower of Salmon to thinking that an analysis of some concept of causation deployed *outside* the project of seeking universal glue must be illegitimate. As we explained, the substance of Salmon's effort can and should survive a revision of his semantic decision. Whether or not what Salmon gives us is best described as an analysis of *causation,* his work demonstrates the need for understanding the structure of the world in terms of objective informational properties if one is to reconcile realism and empiricism. This involves no retreat from understanding Salmon's work as a deeply illuminating contribution to fundamental metaphysics; and the semantic revision need in no way obscure the fact that the contribution evolves out of earlier inquiries, such as Hume's and Reichenbach's, into the nature of causation. Ultimately, if Salmon's work has shown us the way toward a successful account of *universal* glue, then all parochial special-science causal relations must be susceptible to analysis in terms of it. This is what the definition of existence in terms of information-transmission defended in Ross (2000), and cited back in section 3, is supposed to achieve. Here, we do not depend on the unqualified success of that analysis. Our ultimate goal – not yet accomplished – is to show that Kim's whole project relies on a metaphysical intuition about causation that is itself less secure than the role-functionalist explanatory processes it seeks to undermine. We therefore need only demonstrate alternatives to aspects of this intuition, not definitively to replace it with a better one.

However, the fact that we have had to reinterpret Salmon's project in a special way might still seem worrying. All the weight of our answer to Kim, it may appear, rests on this reinterpretation. So anyone finding it unpersuasive, or fearing that, *whatever* we choose to call Salmon's glue-candidate, it will not be compatible with the cognitive scientist's parochial concept of causation, will be unsatisfied. Here we must work on locating the burden of argument. There is, first of all, no question that Salmon analyzes what he calls causation in terms of information-transmission. So if Salmon-style analysis is to support Kim's reductionist picture, then we should be able to find an "information-transmission exclusion problem" analogous to the causal exclusion problem. Defending such an analogy would require a justified intuition to the effect that a map of all the real causal-transmission paths in the universe has to have a simple, low-dimensional geometry, so that if information borne to a receiver by, say, the collisions of particles, were fully transduced and analyzed, then all information borne to that receiver down all other available paths would be redundant. Can any of the work in fundamental metaphysics we have discussed here ground such an intuition?

Perhaps. Suppose someone thinks that the sort of information-transmission relation necessary for performing Salmon-style analysis of special-science causal relations just *is* the causal relation delivered by physics. In that case, the fact that science does observe a rule to the effect that special sciences are not allowed to contradict the generalizations of physics, conjoined with the view that the Salmon-style analysis in question is approximately correct, would lead straight to the intuition just presented, and hence to an information-transmission exclusion problem. Kim indeed seems to believe that the serious metaphysician's master-concept of causation, to which all special-science causation concepts are then answerable, comes from physics. This, at least, would explain his convictions that micro-*physical* causes exclude mental ones, that supervenience must be a one-way relation, and that allowing some program explanations to count as causal amounts to a form of dualism. A similar assumption underlies Jackson's and Pettit's convolutions to the effect that program explanations may be "causally relevant" but cannot be causal. Our reinterpretation of Salmon's project allows it to go forward unencumbered by this assumption that physics supplies one concept of causation for every legitimate purpose. This is a good thing for that

project, because the assumption is *false* of physics as we find it.

4.4. Physics and the physical

In common with much metaphysical philosophy of mind, Kim's arguments trade on a particular image of how things are with physics. This image includes commitment to the view that there is no controversy about whether the apparently causal claims of physicists are indeed causal,[24] and that the distinction between physics and the "special" sciences is simple and exclusive. Much of the bite of the causal exclusion problem arises from the contrast between physics thus understood, and the special sciences. Both assumptions, though, are false: physics in general is not inquiring into ultimate causes, and much, perhaps all, of physics consists of a large collection of special sciences.

We do not deny that there is a metaphysically important sense in which physics is fundamental. Physics is alone among the disciplines in being required to aim at generalizations that hold across all materially possible worlds (by which we mean: all worlds that could actually exist within the boundaries of the singularities, space-like and time-like, that limn in-principle observable space-time). This leads to the fundamental asymmetry in the structure of the sciences we mentioned above, according to which no other discipline may violate currently accepted generalizations of physics, but physics itself respects no corresponding limitation. This entails a (weak) version of the principle of the causal closure of the physical: no special science may traffic in information-transmission relations or, therefore (again, presuming the adequacy of a Salmon-style analysis), traffic in parochial causal relations, that are spooky according to physics. But Kim, as we have shown, needs something stronger than this. The intuition that brings Salmonesque metaphysics into the service of his causal exclusion problem requires that physics *supplies* the form of general causal relation that must then generalize. But it does not.

Working physicists sometimes talk about causes. But, as Cartwright (1983; 1989) has argued, this is because most working physicists, most of the time, are not in search of nomic generalizations holding across the whole scope of materially possible reality. They are, therefore, working within special sciences *within* physics. If we ask, as our engagement with Kim has now forced us to, "What is a physical cause, *in general*?" we must answer the question by reference to the part of physics that seeks generalizations across the whole scope of the science, that is, fundamental physical theory (general relativity and quantum mechanics and electrodynamics). What turns up, by way of examples, is nothing.

In section 4.3 we referred to Redhead's remarks on the elimination of forces from physics. Redhead goes further than that, and argues that much physics has very little to do with causes. Instead of causal laws, he maintains, physicists are interested in finding "laws of functional dependence" such as Boyle's law, where pressure and volume co-exist in certain specifiable ways without it making sense to say that one causes the other. Galileo's law describes the behavior of falling objects, but does not identify any general "cause" of the displacement of objects. Acceleration, for example, just *defines* the kinematic relationship expressed by the law. A standard move is to say that the law measures a "force," and that *that* causes objects to fall. But, as Redhead (1990) notes, because the notion of *force* derives directly from the Aristotelian analysis, this move adds no content to Galileo's law beyond transference of an anthropocentric metaphor. Redhead intends skepticism about the idea of causation as a scientific concept altogether, but we need not go this far for the sake of the argument here. Nor need we claim that if some special branches of physics invoke parochial causal notions of their own, these must be equivalent to the Aristotelian folk notion – we claim that there is a plurality of special interaction-transmission relations, not just a folk notion and a scientific one. What is important here is the factual point Redhead makes about the practice of physics, which is that it does not feature the use of any *general* sort of thing – forces, fields, charges – that is a characteristic kind of cause, picked out by physicists in contrast to other possible general kinds of causes. If this is plausible where classical physics is concerned, it is surely that much more persuasive when we attend to contemporary physics.

In fact, the message obtained from careful attention to physics is about as bad for Kim's hunch as can be imagined. Loewer (2001) joins us in noting that Kim requires a "generation and production conception of causation," but then writes:

> The fundamental laws (for example, Schrödinger's law) relate the totality of the physical state at one instant to the totality at later instants. The laws do not single out parts of states at different times as being causally related. If S' is the microphysical state in a region R at time t' and t' is a time prior to t, then nothing less than the state S of the region R^* that fills the backward light cone of R can be said to produce S'. We cannot say that one event (one part of the physical state) produces another part since the laws do not connect parts in this way. (p. 323)

It gets still worse. Batterman (2000) argues that most *theoretical* (as opposed to purely manipulative) activity in physics consists in searching for what physicists call *universalities.* By this they do not mean, as a philosopher likely would, metaphysical principles necessarily holding everywhere, but *physical* facts that allow them to extract "just those features of systems, viewed macroscopically, which are stable under perturbations of their microscopic details" (p. 129). In particular, they search for suitably abstract topological characterizations of systems in which basins of attraction emerge that corral microphysically heterogeneous processes around the universalities – for example, renormalization group fixed points among Hamiltonians in Hamiltonian-space descriptions of fluids, gases, magnets, pendulums, and other diverse systems that display "critical" behavior with respect to phase-states. Thus, far from invoking generation-and-production causal relations at the micro level to explain functional dependencies among physical states, physicists look for principled physical reasons for *ignoring* most of the aspects with which such relations might be identified.

A follower of Kim might object that this is all just epistemology. If physicists can extract useful generalizations by ignoring lower-level causal detail, just like psychologists do, then this is all to the good; but it does not, and could not, show that necessary causal work is not actually being done "down there" where micro-events are generating and producing other micro-events. However, this interpretation is at best gratuitous, and at worst a contributor to confused physics. Physicists do *not* begin by identifying a micro-level of generation-and-production relations and *then* find bases for abstracting away from some of these relations. They in-

stead engage in measurement, manipulation, and re-parameterization of whole systems until universalities emerge. Wallace (2003) argues that failure to *shake off* the intuition that "down there," under the level of system-level patterns that show stability within restricted measurement time-scales, lies a realm where all measurement values are definite independently of scale, leads to apparent paradoxes and pseudo-problems. For example, there is a widespread belief that the established quantum formalism is incompatible with definiteness of measurement at the macro-level – the famous problem of Schrödinger's cat – and so needs to be supplemented with empirically unmotivated parameters for finding connection principles that temporally link multiple worlds, or link multiple observers into continuous minds, so as to allow for superpositions of micro-states without corresponding macro-superpositions (a cat being simultaneously dead and alive). Such mangling of the formalism to make it square with our old metaphysical hunches has a severe cost in terms of physical theory: it "almost inevitably spoils the relativistic covariance of the theory." We best dissolve these insoluble dilemmas, Wallace suggests, by dropping the hunches and thinking of "reality" as measurement-scale-relative patterns in structural properties of quantum states "all the way down." (Reading Wallace's argument alone, one might worry that his point is itself just a qualitative philosophical hunch, but this is not so. Recent work by Nottale (1993; 2000), for example, gives formal details motivated from within physics.)

The above survey of physical ideas is not intended to represent a settled picture of or a committed prediction about the metaphysical implications of contemporary physics. The point, rather, is this: Physics supplies no "master-concept" of causation that is motivated independently of some particular explanatory program. Physics does not encourage, and may well even actively *discourage* – as Wallace effectively claims – the Salmonesque interpretation of causation-as-metaphysical-glue on which Kim relies, unless that glue is reinterpreted structurally. By "structurally" we intend reference to networks of informational relations as suggested earlier, or to topological structures of fractal space-time following Nottale (1993; 2000), or to something else that features a key property incompatible with Kim's hunches about physics, namely, that basic measurements are indexed by global rather than local analytic procedures, in the strictly mathematical sense of these terms. (That is, measurement values are not indexed to neighborhoods of points.[25]) What all such interpretations have in common is that, far from undergirding the one-way local supervenience that Kim transforms into reductionism as a general metaphysical principle, they suggest it to be a scientific anachronism. There is indeed a deep tradition of "causation as universal glue" to which Kim implicitly appeals, but that tradition has found its most sophisticated contemporary expression in analysis of causation as a special type of information-transmission or other global-structural relation.[26] One cannot derive a basis for insisting that physicalism implies one-way supervenience from analysis of this concept. So Kim's particular reductionist image, from which follows all the trouble for special sciences identified in section 3, ends up resting on a vague and unmotivated hunch about an unanalyzed class of absolutely general causal relations.

The upshot of our discussion to this point is negative: We have shown that there is no rational basis for thinking that special sciences that traffic in multiply realized and multiply supervenient functional kinds should expect to have to endure conceptual revolutions under pressure for general unification of science, or because they presently fail to track real or non-redundant causal relations, thereby missing the genuine explanations of the phenomena in their domains. We thus hope to have shown cognitive and behavioral scientists how to answer, at least in general terms, the complaints of skeptical metaphysicians. However, we noted at the outset that a more straightforward way of doing that, by a simple appeal to epistemic pragmatism, has always been available to special scientists, and is no doubt the sociologically dominant response. The justification for all the work we have been asking behavioral scientists to do in working through our argument depends on our claim that metaphysics *should* be taken seriously, that is, can actually contribute something *positive* to cognitive and behavioral inquiry. The promises made at the beginning of the discussion, therefore, have not been discharged until – now that we have swung a wrecking ball at the structures of conservative metaphysics of mind – we say something about how to build a scientifically useful structure above the rubble. To this we therefore turn in our concluding section.

5. Conclusion

The general consequences of the preceding discussion for behavioral and cognitive scientists can be consolidated as consisting of one negative point and one positive one. In order to state the latter from a clear basis, we begin here by summarizing the negative upshot. Recall that we identified a general and important metaphysical task as that of identifying whatever it is that holds all objective relations in place, metaphorically calling this a kind of glue. Recall also that we have distinguished two different senses of *cause* discernible in the history of philosophical reflection on causation, and relevant in different ways to metaphysics and reduction. Kim's conviction that the special sciences are answerable to physics amounts to the conviction that the particular causal claims produced by physics amount to statements about the metaphysical glue. He thinks that physical causal claims are already metaphysically unimpeachable, and that that is the reason why the causal claims of special sciences have to answer to them. But physics is itself largely composed of special sciences, physicists are not best seen as in the business of discovering causes, and the primacy of physics does not consist in the fact that physicists are, simply in virtue of being physicists, automatically doing fundamental metaphysics. Kim is thus multiply mistaken.

The intuition among philosophers that "down there," in physics, lies an unproblematic and univocal concept of causation that can directly inform metaphysics runs very deep. As we have seen, even Jackson and Pettit, whose work has been motivated by a concern to defend and articulate the basis for the strong autonomy of special sciences, succumb to this picture when they assume – without argument or even much discussion – that program explanations, no matter how important they may be to scientific explanation, cannot be causal. We have argued, however, that the metaphysical ground is nowhere close to being sufficiently settled or uncontroversial to drive a set of conclusions as logically strong, or as troubling for scientific practice, as the new reductionists imagine. Indeed, we suggested in the last

preceding section that current trends in physics and the philosophy of physics make their commitment to a *localist* conception of "causal glue" look like an increasingly poor bet.

We thus claim to have shown cognitive and behavioral scientists how to see off metaphysicians who are skeptical about the explanatory adequacy of their hypotheses and conclusions on grounds that these rely on ultimately reducible or eliminable causal mechanisms. They can say: "We're scientists, not metaphysicians. We aren't trying to explain *in general and at one analytic stroke* how our macro-phenomena relate to micro-phenomena. This doesn't mean that we dismiss metaphysics as irrelevant; we'd worry if you were right that we're positing scientifically isolated mystery processes. But we don't need to integrate ourselves with other sciences by identifying mental causes with nonmental causes – there isn't any single scientific concept of causation to govern this. We'll stay integrated by piecemeal connections as we go, and if a metaphysician offers us a more general unifying principle that actually sheds potential light on our subject – mind and behavior – then we're all ears." In the light of our discussion's length and complexity, however, we can't exactly claim that this gift has come for free. Many may be inclined to think that all we have done is provided a tediously unnecessary justification for doing something they could always do, but without work: ignore philosophers. A scientist who believes that *all* metaphysics is gratuitous to her activity will not thank us for buying her a lunch she had no interest in eating.

We have operated from the assumption, however, that metaphysics can and should be taken seriously *as a part of, and for the sake of, science.* We thus owe some demonstration of payoffs in these terms. There are, we think, two. First, cognitive and behavioral scientists *do,* like most special scientists outside of physics, invoke and rely on distinctive causal concepts; but these are frequently implicit, and this implicitness can and does complicate debates over investigative methods and interpretations of conclusions. We think we are now in a position to say something enlightening about the causal concepts at work in cognitive and behavioral science. Second, special sciences, despite being separable by definition, *do* lean on one another in a variety of practically significant ways. We will be able to say something useful about the details of that, too.

Virtually all models in the broad domain of cognitive and behavioral science rest on the idea that nervous systems, in interaction with environments, are engines that "produce" behavior and perhaps – a recurrently controversial point – representations. Behaviorists, Gibsonians, some connectionists, and many neuropsychologists have motivated their research programs by, and thus apparently made their importance hostage to, the conviction that representations are the wrong sorts of things for carrying ultimate causal efficacy. Those who make more robust use of representational structures typically counter their skeptical *scientific* critics by noting that computer programs manipulating representations are undeniably causally significant, and that anti-representationalist projects are simply refusing to avail themselves of helpful resources. Still, it is usually conceded, the causal sources of behavior cannot be representational "all the way down"; somewhere, somehow, there must be a level of activity analogous to that of electrical circuits in computers that fully explains the "mental" patterns. To suppose otherwise is to allow the possibility of dualism or magical emergentism or some similarly irresponsible license for seceding from the legitimating sphere of real science. Anyone who doubts that real scientists worry about these issues, in the course of criticizing, defending and building new work upon serious models, need merely review a sample of back issues of this journal.

The set of assumptions that drives these debates is distinguished from Kim's only in being (typically) a bit less explicit. They arise because the legacy of two concepts of causation in tension is a *general* inheritance, alive in psychology as well as philosophy. Functionalist analyses of representations as making irreducible differences to behavior are piecemeal vindications of the scientific significance of old-fashioned agent causation. This sort of causation *is* disturbingly unlike the kind of modern causation by mechanical bumping or (later) magnetic or gravitational pulling that all of science is still often imagined to be "ultimately" about. It seems to us that in much of cognitive science, explanation by allowing mental control of physical behavior *is* viewed as a kind of pragmatic compromise: ever so useful for getting work done, but *someday*. . . .

Suppose, though, that the appropriate way of dissolving the tension is to allow a refined and sophisticated kind of agent-causation as a parochial special-science type, while giving up on the modern, generic, kind of causation. Such positive news for cognitive scientists should be as surprising – if comforting rather than threatening – to some cognitive scientists as to Kim and his followers among philosophers. Yet so far as other sciences – emphatically including physics – *and* careful metaphysical speculation are concerned, it is just as plausible to suppose that the fundamental ontological structures governing all of science are global and structural as to suppose that they are local and mechanical. Contemporary cognitive and behavioral science is dominated by accounts of feedback-driven servosystems and hypotheses about how natural and cultural selection can build and maintain them. It is very natural to suppose that such complex dynamics must be "built out of" simpler processes in an additive way. This, however, is a metaphysical assumption, derived from meta-reflection on the history of science, which is not now standing up well under concerted pressure.

Here is an alternative metaphysical image: the dynamic patterns studied by the cognitive and behavioral sciences are instantiations, at particular scales of metric identification and measurement, of *more global* dynamics characteristic of the physical universe in general. Recent work in mathematical physics, information theory, and analytical metaphysics shows how to make this claim relatively precise and non-fuzzy from the perspective of mathematics,[27] but it *is* more than a little boggling with respect to prevailing intuitions about explanation. However, the current adventures of the conservative metaphysicians can help to remind us that explanation by reference to "ultimate" collisions of particles is no less *logically* puzzling; we simply grew accustomed to it during the long march from the days of Galileo and Kepler. Let us be clear: we are not here *asserting* that, as matters have turned out metaphysically, the world is made of informational topologies (or some other kind of globally structuring manifold) "all the way down" and demanding that everyone sign up. Scientists are justified in their prevailing pragmatic intuition that metaphysical inquiry does not generate clean, resolute satisfactions of

this sort, and that this is related to the grounds on which they should keep some distance from it in their pursuit of lasting explanatory accomplishment. However, metaphysical frameworks guide science as *constraints* whether we like it or not. Furthermore, for reasons we will now briefly discuss by explicit reference to the causal concepts of cognitive science, some such constraints are, at any given time, necessary for scientific progress. It is equally crucial that these constraints be allowed to evolve, to the extent of complete replacement over time. Such constraint management is sufficiently delicate, and sufficiently important, to justify philosophical activity. We will illustrate the point by reference to some actual recent debates and activities in cognitive science.

We have already mentioned one way in which implicit localist metaphysics influences activity in cognitive science: it leads those who develop representationalist models to constrain them by the idea that they must be amenable to vindication through implementation in some set of "lower-level" local mechanisms. Often, all this amounts to is that a few speculative paragraphs on possibilities for such implementation get tacked onto the backs of papers describing representationalist models; and this is hardly a problem, if it is a problem at all, over which to get worked up. However, it expresses the fact that most scientists *do* feel a responsibility not to leave their models isolated from the wider, unified explanatory project. Vague speculations about implementation are, at their limit, lip-service acknowledgments of this. The principle becomes important to science when it is taken truly seriously. The leading expression of genuine commitment to the principle that has recurrently characterized work in behavioral and cognitive science is *restriction* of modeling approaches to domains of explanation that are taken to be *already* unproblematic from the perspective of localist implementation.

Glimcher (2003) has recently given us a history of neuroscience from this explicit perspective. The Sherrington program for explaining all "determinate" behavior by reference to passive reflexes, each of which responds in isolation, according to fixed condition-action rules, to a finite menu of possible stimulations, is as perfect an instance of commitment to Kim-style localism as can be found anywhere in science. Because Sherrington doubted, empirically, that all behavior is determinate in this way, he was a dualist; note, then, that his dualism was not a *directly* metaphysical thesis, but a *scientific* response given an implicit metaphysical constraint on hypothesizing. More interesting, as Glimcher shows, is that grounds for doubt about the capacity of pure reflexology to explain even the plausibly "determinate" behavioral patterns were made evident (by Graham Brown) during Sherrington's lifetime, and more systematic critiques in the mid-twentieth century, both theoretical and experimental, by von Holtz, Mittelstaedt, Weiss, and Bernstein accumulated decisive refutation. Nevertheless, Glimcher argues, the most emulated and productive neurophysiological investigations *right now* – connectionist models of learning using backpropagation, and Shadlen et al.'s (1996) celebrated work on visual perception of motion in monkeys are Glimcher's examples – continue to honor Sherrington's localist paradigm. The contemporary work of course invokes a range of new mechanisms Sherrington could not have imagined; but the commitment to input-driven, localized, non-hierarchically governed processes remains in place.

Glimcher's point is not that these studies do not merit celebration or emulation. His point, rather, is that neuroscientists, despite knowing that localism in their domain is not generally true, and despite their not being willing to allow dualism, concentrate their best energies on such phenomena as *can* best be modeled in localist terms. We now suggest that this should be interpreted not as a retreat from unification with other special sciences but as an indication of extremely serious commitment to it *given a prevailing, mostly implicit, localist metaphysic.*[28]

With Kitcher, Friedman, Kincaid, and other philosophers we have cited approvingly in the course of this paper, we agree that special scientists are right to care about avoiding completely isolated explanations. Here an overt normative principle is in order: If what science mainly delivered were a chaos of scattered descriptions of unrelated phenomena, we would be justified in feeling crushingly disappointed in it. This would be science as, at best, a pure under-laborer to engineering, not a collective project for increasing our understanding of the universe. Like Glimcher, we intend no criticism of cognitive scientists who respect this norm by doing powerful work that preserves unity by leaving wider metaphysical assumptions unchallenged. However, explanation is no virtue if we do not care whether explanations are, in addition to being comprehensible, *true.* This implies that we must not leave metaphysical presuppositions unchallenged in practice unless philosophical reflection convinces us that the presuppositions in question are actually justified. The justification of a metaphysical presupposition should rely mainly on its fruitfulness in science, so commitment to localism was a healthy restriction for a long time. But in some sciences – in physics, in economics, in many parts of biology and cognitive science – that time has passed. Recent experience and reflection suggests that explanation of local phenomena as instances of global dynamic structures is a viable alternative route to unification.

This is just what Kim and the conservative metaphysicians deny. Kim's commitment to fundamental metaphysics is expressed as the insistence that one does not solve the mind/body problem by offering particular accounts of intentional processes in non-intentional terms. Rather, one must try to explain how in general "mental properties and physical properties are related, and hopefully also explain why they are so related" (Kim 1998, p. 5). We agree. We also agree that functionalism cannot be vindicated as providing such an account by mere appeal to supervenience. But functionalism *could* license agent-causation as a legitimate, special-science-parochial sort of causation after all, if more general accounts of informational or other topological dynamics can show it to be non-mysterious – and the prospects here look promising (see Juarrero 1999). Surely we have, up to an important standard of generality, explained how mental properties and physically non-problematic properties are related if we produce a broad account of feedback-driven servosystems and the ways in which evolution has built nervous systems that support them. If dynamic systems theory *is* a way of doing metaphysics – and that is what we are suggesting – then servosystematic control without localist reduction is not isolated as a basis for explanation (and the same goes, in spades, for evolution). Here, perhaps, is the source of the technical tools through which the special problem of mental causation and the general questions about universal glue find a common logic of address. But the working problem

of mental causation, as we see it, is the very old problem of how agency is possible. *Causation,* in this context, means something special: the processes, whatever they are, by means of which thoughts and decisions, beliefs and desires, make a real difference in the world. We see no reason to believe that there is any more "general" a way of addressing this problem than by the approach of contemporary behavioral science, with its plethora of servosystematic control processes grounded in neuroscience and ethology.

Kim's problematic is what you get if you reify the folk and the post-Humean concepts of causation. On the one hand, you find yourself wanting to show how interventions by agents can make a difference to what actually happens in the world. But then, on the other hand, you insist that these interventions must be micro-processes, or decomposable into micro-processes, that agents must not just turn out to be programs. Well, we think it is overwhelmingly likely that agents are programs, and they are not anything else. Some mental states are reliable bearers of information about other mental states, even though no particular state in the supervenience base of one is a reliable bearer of information about any other particular state in the supervenience base of the other. If something is a running series of such states, then it is a program, something that acts and exists *by* compressing information. Indeed, living systems are only possible at all thanks to the fact that some of their states, including mental states in those with brains, extract and emit useful (accurate) information in compressed form. That they *can* do this is empirically evident. It is also not, pace Kim, mysterious: thanks to the dynamics of servosystematic feedback structures, multiple supervenience is possible (and actual).

This contradicts nothing that physicists either presuppose or tell us. Physicists, like all scientists, study patterns of compressed information at whatever scales they can be found, not, by elision, some non-existent level of "ultimate" micro-banging and colliding. This is good news, we take it, for most cognitive scientists. But the full news is even better. When we "do" metaphysics in the naturalist's way – by standing back and looking hard at collections of special sciences in abstraction – then moving our attention from the cognitive sciences to the physical ones does not involve a discontinuous leap from spooky or redundant causal relations to good old-fashioned mechanical ones. We see instead convergent dynamical accounts that can be swapped across the boundaries of the many special sciences in a profoundly interdependent intellectual market.

ACKNOWLEDGMENTS

Previous versions of this paper have been presented at conferences in Dubrovnik, Stellenbosch, and Ghent. We thank the audiences on those occasions. We are also grateful to John Collier, Daniel Dennett, Harold Kincaid, James Ladyman, Ausonio Marras, Veronica Ponce, Alex Rosenberg, and Cosma Shalizi for critical feedback and comments, to the editors and referees of this journal for comments and criticisms of an earlier version of the paper, and to Nelleke Bak for careful reading of the text.

NOTES

1. The phrase is due to Bickle (1998). However, we will not here be engaging with Bickle's interesting thesis, which has enough direct empirical content to be a piece of cognitive science in its own right. The philosopher who has done most to inspire the backwash is Jaegwon Kim, and his most influential argument, as given in Kim (1998), will be the target of our discussion.

2. Or, at least, so philosophers often say. As a claim about actual behaviorist psychologists, this claim is largely nonsense, flatly untrue of, for example, E. C. Tolman or Karl Lashley. But the importance given to knocking over this straw man in the history of the rise of functionalism is indisputable, and is what is of relevance to us here. We would encourage more footnotes like this one in the philosophical literature, however.

3. Functionalism, thus understood, can be a kind of behaviorism – just one allowing for some intermediate behaviors between stimulus and response. Our own favored variety of functionalism is in fact of this behaviorist sort; but this will not play any direct role in our argument in this paper.

4. We use this example because it is standard in the literature we are describing. We should point out, though, that the philosophers who introduced it knew from the outset that it was at best neurologically implausible, and were using it as a place-holder for some imagined future reduction of a psychological to a neurological state.

5. The same argument structure has also been used (e.g., Horgan 1997) to argue for functionalism within a species (on the basis of significant neural and other differences between conspecifics), or, over time, within a single individual.

6. There are various particular ways of being a realizer functionalist in the broad sense indicated here. One particularly strong way is via the "functional analysis" strategy associated with Armstrong and Lewis, as discussed in section 2.2. Another way, canonically defended in Pylyshyn (1984), requires that types be individuated either by reference to intrinsic properties of members of the type, or by reference to intrinsic properties of independently specifiable sets of tokens of the type.

7. On our broad conception of realizer functionalism; see note (6) above.

8. Kim (1998) is not the first expression of the problem, merely an elegant, sophisticated and up to date version of it. Yablo (1992) is a clear and widely cited statement of the issues as of the early 1990s, and the papers in Kim (1993) show many of the lines of argument and thinking that lead up to Kim (1998).

9. We assume throughout that we are talking amongst people who would regard the admission of supernatural causes into science as the end of the world.

10. Talk of "enhancing" is somewhat sloppy, as Marras (2002) points out, but the details need not detain us here.

11. The philosophical literature on explanation is enormous, and so some philosophers might object to our announcing that we can boil it down to consideration of just two approaches. A few meta-comments on that literature are therefore in order. It divides naturally into two piles. The first pile, concerned directly with the way in which the search for explanation descriptively and normatively guides scientific activity, really *does* mainly revolve around the dialectic established by Kitcher's and Salmon's long argument with each other (see Kitcher & Salmon 1989). The second pile, highlights of which include van Fraassen (1980), Garfinkel (1981), and Achinstein (1983), concerns the logic of explanatory statements. Both piles descend from the classic work on explanation in philosophy of science by Hempel (1965), that, in the way of positivism, saw these two concerns as indistinguishable. To a post-positivist of whatever stripe, however, they are distinct, and to a considerable extent orthogonal. That is, just about any combination of views from the first and second debates can be made compatible (see, e.g., Kincaid 1997, whose work depends on subtle recombinations of them). For our purposes in this paper, only the first set of issues about explanation are directly relevant. Non-philosophers are cautioned, however, against taking our summary as a mini-survey of the whole literature on the subject.

12. As Batterman (2000, p. 118) notes, "Kim's argument won't go through unless the causal properties of the macroproperties *just are* the resultant (or 'sum') of the microstructural properties."

13. For a sample of the literature urging this perspective, see McClamrock (1995); Wilson (1995); Clark (1997); and, especially influential with respect to what we say here, Dennett (1991a). Pet-

tit (1993) provides the most systematic, though very cautious, investigation of these ideas.

14. According to Menzies (1988) this line of argument was suggested by Lewis.

15. That is, a property possessed by an object (such as dormativity) in virtue of its having some more basic (e.g., chemical) properties.

16. We are especially indebted to Ponce's treatment here.

17. Clapp (2001) successfully argues that some leading defenses of the autonomy of special sciences, such as Fodor's (1974), are guilty of this lapse of metaphysical seriousness. We should therefore note explicitly that none of our arguments in this paper depend on the idea that the kinds of any special science must be preserved as kinds just *because* people find it useful to think with the concepts they represent. Indeed, on one interpretation this is what "taking metaphysics seriously" in our sense here *means.*

18. Ross finds it necessary to make Dennett's idea more systematic because Dennett's own account, as Ross explains, leaves too many doors open to emergentist and, in other places, instrumentalistic, readings. On the other hand, Dennett's paper surpasses Ross' in anticipation of a theme to which we will shortly be devoting much attention: the relationship between reductionism and a scientifically unsophisticated understanding of causation. See, especially, Dennett's footnote 11, and compare this with sections 4 and 5 of the present paper.

19. Stich (1983) devoted a book to arguing that this sort of picture, intended as a way of reconciling a plausible cognitive scientific typology of states with folk psychology, could not work. Kim must believe that Stich is wrong about this.

20. Cartwright (1983; 1999) has famously argued that the world is *not* a single, working machine, but is instead "dappled," by which she means ontologically disunified. Dupré (1993) has urged a similar thesis. For reasons given in Spurrett (1999; 2001a) we reject this conclusion. The fact that science is never finished, and therefore never completely unified, may mean that its current description of the world at any given time will always be of a world that is "dappled"; but to derive *as a metaphysical conclusion* the claim that the world *is* dappled is to simply abandon the regulative ideal that informs Salmon's project, and, for that matter, Kim's. Answering Kim *this* way *would* simply amount to shrugging off the significance of realist metaphysics, another way of trying to have lunch for free.

21. The "something," we would say, is indeed fundamental structure; Ross (2000) takes it to be the network of Schrödinger-style negentropic relations. That network is *our* favorite candidate for universal glue.

22. Kitcher develops, at length, additional criticisms based on counterexamples to Salmon's technical criteria for distinguishing genuine causal processes from pseudo-processes. We will not incorporate these into our summary here, because they contribute little to the issues relevant to our discussion, and because even if Salmon's apparatus is repaired so as to block the counterexamples, Kitcher's main critique is unaffected.

23. It is common outside philosophy for Dennett to be called a "reductionist" because he analyzes intentionality and consciousness without recourse to any entities or processes incompatible with the causal closure of physics. However, Dennett in fact denies, like us, that there is any *general* relation between physics and special sciences stronger than global supervenience; and in the context of most debates in philosophy of science, this makes him as anti-reductionist as the recent tradition allows. Thus, for example, when Kincaid (1997) defends anti-reductionism, he feels he needs to spend a few pages showing that he need not go as *far* in that direction as the "radical" Dennett. Ross (2000) explains in detail the sense in which Dennett's anti-reductionism is radical.

24. Philosophers typically grant that our current physical theories are open to revision, so the point here is *slightly* more complicated. Still, for philosophers of mind an ideal physicist is generally *assumed* to be making unproblematically causal claims, whereas an ideal economist, say, would need to do additional philosophical work over and above her economics to justify thinking of her claims as causal.

25. We owe this insight to Andrei Rodin.

26. The logic of this, and comparison of the causation concept's role in different branches of science, is made formally explicit in a recent paper by Thalos (2002).

27. We allude to the earlier references to the work of Nottale (1993; 2000).

28. This point has also been vividly argued by Dennett (1991a).

Open Peer Commentary

Metaphysics, mind, and the unity of science

David Boersema

Department of Philosophy, Pacific University, Forest Grove, OR 97116.
boersema@pacificu.edu

Abstract: Ross & Spurrett's (R&S's) rebuttal of recent reductionistic work in the philosophy of mind relies on claims about the unity of science and explanation. I call those claims into question.

Ross & Spurrett (R&S) have written a spirited, and I believe fundamentally correct, rebuttal of recent work in metaphysics that seeks to undermine the anti-reductionist, functionalist consensus of the past few decades in cognitive science and philosophy of mind. Their rebuttal focuses on challenging metaphysicians' treatment of causality and their conception of physics, including the relationship between physics and metaphysics. Calling such recent work in metaphysics "the new scholasticism," R&S decry its "unhealthy disregard for the actual practice of science" and its tendency to "drift away from relevance to and coherence with scientific activity" (target article, sect. 1, para. 5). Although much of R&S's article focuses on the recent work of Jaegwon Kim in particular (Kim 1998), in this short commentary I will address the more global concerns of explanation and the unity of science. Also, although I agree with much of what I take R&S's project to be here, I will emphasize in my comments those aspects that I find suspect or at least in need of clarification.

R&S claim that: "The goal of science is to discover the structures in nature. We can discover such structures because, as fairly sophisticated information-transducing and processing systems, we can detect, record, and systematically measure mark-transmitting processes" (sect. 4.2, para. 5). This brief claim points to and points out several issues that beg for elaboration. While bemoaning the strident reductionism found in the work of recent metaphysicians (or what I take as the metaphysicians' commitment to a unity of science), R&S apparently do not deny a unity of goals of science (to discover nature's structures). The unity of science that is questionable, then, I assume is: (1) unity of methods, (2) unity of values, or (3) unity of content (i.e., epistemic unity, axiological unity, or ontological unity). Unity of methods I take as a commitment that the various sciences, in the final analysis, do or ought to investigate the world following the same (or similar enough) procedures, processes, and methods in order to provide accurate, reliable, and perhaps replicable information. I infer that R&S reject the notion that all of the various sciences do or even ought to do this. We hope clinical trials on the effectiveness of placebos will involve double-blind studies, but we don't expect cosmologists trying to determine more accurate parameters of black hole event horizons to involve such studies. But what reductionist would think otherwise?

Unity of values I take as a commitment that the various sciences, in the final analysis, do or ought to demand the same (or similar enough) standards, aims, and so on for (1) the worthiness of scientific information (i.e., unity of epistemic values such as quantifiability of data, levels of accuracy or precision, etc.) or (2) the social significance of scientific information (i.e., unity of social/ethical values such as providing predictable control of applications, emancipatory value of information, etc.). The epistemic values and the social/ethical values that we attach to and demand of scientific investigations are not uniform. Given the consequences of a mistake in the interpretation of data, we vary our expected level of confidence in experimental results. But again, what reductionist would think otherwise? Therefore, if R&S's concerns about reductionism are concerns about the unity of science, it is not clear that such concerns are ones about epistemic or axiological unity of science because it is not clear that reductionists are committed to any such unities. Is the concern then really just ontological unity? If it is, I confess that I share some of their reticence, although it is not obvious to me that most recent metaphysicians are guilty of being committed to such unity. We do demand that the content of the "special sciences" (i.e., not physics) not violate the content of physics because we take physics to tell us about the basic components and constituents of the world. However, it is not clear that the demand that the content of, say, cognitive science not contradict the content of physics is the same as the demand that the content of cognitive science be formally derivable from or eliminable into the content of physics. The difference between the two, I take it, actually lies not in any ontological commitments (to unity or otherwise) but rather in what counts as being explanatory of the phenomena being investigated. It is just this notion of explanation where I find another point that begs for elaboration. Quickly, before turning to that, however, I will repeat the present concern, which is just what sense of unity of science that R&S find so questionable (at best) or reprehensible (at worst). I find an underlying commitment to unity in their rejection (target article, Note 20) of Cartwright's and Dupré's criticisms of such unity (Cartwright 1983; 1999; Dupré 1993).

In their account of explanation, R&S draw heavily and directly from the work of Kitcher's (1981) unification model of explanation and Salmon's (1984) causal model. In combating a commitment to a reductionist unity of science view, however, they take not causality but information transmission (in the mathematical sense of Shannon & Weaver 1949) as the primitive notion for explanation. I find such a move to be fecund and philosophically more fruitful than a reliance on a causal model, but its very virtue is one I would take a reductionist to find as missing the point. The information-transmission model of explanation retains the virtue that R&S want, which is to provide an objective measure of explanation that does not make any ontological commitment to any reduction to physics. The old stand-by about the bank robber Willie Sutton demonstrates that what counts as a good explanation is relative to the aims and goals of the inquiry. ("Willie, why do you rob banks"? Willie: "Because that's where the money is.") However, I take it that for Kim and other metaphysicians, it is not *a good* explanation that is sought, but *the correct* explanation (with emphasis on *the* correct explanation, not *a* correct explanation). Reductionists, I suspect, would find R&S's discussion of explanatory accounts to be beside the point, because any explanation that is not finally cashed out in physical terms is not correct, regardless of how good it is toward salving any particular inquiry.

Ontological disunity and a realism worth having

Steve Clarke

Centre for Applied Philosophy and Public Ethics, Charles Sturt University and Australian National University, Canberra ACT 2601, Australia.
stephen.clarke@anu.edu.au
http://www.csu.edu.au/faculty/arts/cappe/people/clarst/clarst.htm

Abstract: Ross & Spurrett (R&S) appear convinced that the world must have a unified ontological structure. This conviction is difficult to reconcile with a commitment to mainstream realism, which involves allowing that the world may be ontologically disunified. R&S should follow Kitcher by weakening their conception of unification so as to allow for the possibility of ontological disunity.

According to Ross & Spurrett (R&S): "Science aims to tell us how the world is structured, that is, how its various processes and classes of entities constitute a single working machine" (sect. 4.2, para. 5). They consider this claim to be "crucial to any sort of realism worth having" (sect. 4.2, para. 5). R&S's crucial claim sits very awkwardly with a consideration that is usually taken to be part and parcel of a serious commitment to realism. This is the requirement that the world be conceived of as existing independently of our thinking about it – the realist requirement of mind independence. The committed realist will be on the lookout for unwarranted presuppositions that we bring to our interpretation of the world and will attempt to get by without such presuppositions. The assertion that science aims to tell us how the various aspects of the world collectively constitute a "single working machine" looks like it is based on the presupposition that the world must be a single working machine. From the perspective of a mainstream realist who is committed to a conception of the world as mind independent, this is an unwarranted presupposition because it seems possible that the world is not a single working machine.

R&S do not do much to unpack the phrase "single working machine," and it may be thought that the above line of reasoning could be evaded if their commitment to a conception of the world as a single working machine was interpreted in a sufficiently nebulous way. However, R&S appear to disqualify themselves from adopting this line of defense by explicitly identifying Nancy Cartwright's (1999) "dappled world" thesis – the view that the world is ontologically disunified, lacking unifying laws, kinds, or other universal ontological categories – as a thesis that is incompatible with the claim that the world is a single working machine (target article, Note 20). Their conception of the world as a single working machine involves the assumption that the world must have a unified ontological structure.

R&S cite recent work, owing to Spurrett (1999; 2001a), that takes issue with Cartwright's claim that there is strong evidence that the world is ontologically disunified (Note 20). I agree with Spurrett that Cartwright (1999) has not done enough to warrant this conclusion. Nevertheless, it surely is possible that the world is ontologically disunified. We do not have to insist that the world is ontologically disunified to have grounds to doubt the claim that the world must have a unified ontological structure. We can be "agnostic dapplers," to invoke Lipton's (2002) terminology, remaining open to the possibility that the world is ontologically disunified, as well as remaining open to the possibility that it is ontologically unified. If we adopt this sensible open-minded attitude, then an insistence that the world must be ontologically unified remains in tension with the realist ambition to depict the world as it is, independent of our presuppositions about it, because we remain open to the possibility that we are living in a disunified world.

R&S associate the claim that there is a unified ontological structure to the world with the work of Philip Kitcher, and they draw heavily on Kitcher's unificationist account of explanation in an effort to identify a form of explanatory unification that is suitable to their conception of science. Their reliance on Kitcher is unfortu-

nate because Kitcher has long recognised that the presumption that the world must be ontologically unified is a weakness of his unificationist approach to explanation. In his words: "it looks as though the approach must defend the *prima facie* implausible thesis that the world is necessarily unified" (Kitcher 1989, p. 496). Kitcher's initial response to this problem was to "recommend rejecting the idea that there are causal truths that are independent of our search for order in the phenomena. Taking a cue from Kant and Peirce, we adopt a different view of truth and correctness" (1989, p. 487). This is *a* solution to the problem created by the possibility of ontological disunity, but it is not a solution that genuine realists, which R&S purport to be, should be happy to endorse. In effect Kitcher is proposing that we compromise realist ambitions by adopting a Kantian position in which order is, at least in part, projected onto the world. Kitcher (1989, 1994) is quite explicit about the Kantian flavor of his views.

Kitcher has recently undergone a change of heart. He now tells us that his "grand project of articulating the most unified vision of nature that we could achieve . . . is mistaken" (Kitcher 1999, p. 347). In 1989 Kitcher was a grand unifier, but the 1999 Kitcher is an advocate of "Modest unificationism." Modest unificationism involves accepting that "the world may be a disorderly place, that the understanding of its diverse phenomena may require us to employ concepts that cannot be neatly integrated" (1999, p. 339). Modest unificationism involves looking for unity where we can find it, while accepting that there may be limits to the amount of unity that is there to be found. It is a position that should be congenial to genuine realists because it does not involve presuppositions about the ontological structure of the world.

R&S begin by observing that "Philosophy progresses with a tide-like dynamic." The low tide of logical positivism was more than half a century ago, but it seems that the high watermark of realism has not been reached, if their article is any guide. Their conviction that science should aim to describe the world as a single working machine appears to be an unwarranted remnant of the strong unificationism characteristic of the heyday of logical positivism. Kitcher has abandoned a similar conviction, and I can only urge R&S to follow his lead. Mainstream realism is compatible with the weak unificationism that Kitcher (1999) now advocates but not with the form of unificationism that R&S currently favor.

ACKNOWLEDGMENTS

I thank Daniel Stoljar and Seumas Miller for helpful comments.

Reduction, supervenience, and physical emergence

John Collier

Philosophy Programme, University of KwaZulu-Natal, Durban 4041, South Africa. **collierj@nu.ac.za** **http://www.nu.ac.za/philund/collier**

Abstract: After distinguishing reductive explanability in principle from ontological deflation, I give a case of an obviously physical property that is reductively inexplicable in principle. I argue that biological systems often have this character, and that, if we make certain assumptions about the cohesion and dynamics of the mind and its physical substrate, then it is emergent according to Broad's criteria.

Reduction is ambiguous in three ways. It may mean inter-theoretic reduction, the reduction of fundamental kinds of things (substance, traditionally), or that certain particular entities (objects, processes, or properties) can be eliminated without any loss of explanatory power in principle. I will ignore inter-theoretic reduction. The reduction of the number of fundamental kinds of things is best called *ontological deflation.* I will assume the closure of the physical (physicalism), and I will assume that all scientific explanation is in some sense causal and that explanatory power is lost only if the causal nature of a higher level entity is not in principle completely reductively explicable.

Despite supervenience, if explanatory reducibility fails in principle for some entity, then it is emergent. If there is no possible argument (deductive or inductive) from the parts, their intrinsic properties, and their relations to the full causal powers of the entity itself, then reductive explanation fails in principle. I will show that this holds for certain obviously physical properties of some systems under certain specific conditions. I will further argue that this helps to identify a class of systems for which reductive explanability fails. In these cases, even if physicalism is true, they are emergent. This idea of emergence fits C. D. Broad's criteria (Collier & Muller 1998).

The planet Mercury was found in the 1960s to rotate on its axis three times for each two times it revolves around the sun. This was extremely surprising because it had been thought that it would be in the same 1:1 harmonic as our moon-earth system. There are several more complex harmonic relations in the solar system. It is well known that the three-body gravitational problem is not solvable analytically, but it can be solved numerically, in principle, to any degree of accuracy we may require for any finite time (this is true for any Hamiltonian system). However, these cases involve the dissipation of energy through tidal torques unless the system is in some harmonic ratio. We would like, ideally, a complete explanation (possibly probabilistic) of why Mercury is in a 3:2 harmonic. Because of the high mass of the sun and the proximity of Mercury to the sun, the high tidal torque dissipates energy reasonably quickly in astronomical time; therefore, Mercury is likely to end up in some harmonic ratio in a finite amount of time. The central explanatory problem then becomes: why a 3:2 ratio rather than a 1:1 ratio, like our moon, or some other harmonic ratio?

We cannot apply Hamiltonian methods, because the rate of dissipation is roughly the same as the characteristic rate of the phenomenon to be explained. If the dissipation rate were small, then we could use an approximate Hamiltonian system; if it were large, we could use a step function. We are left with the Lagrangian. It is well known that these are not always solvable even by numerical approximation. I will give an intuitive argument that the Mercury's harmonic is such a case. Each of the possible harmonics is an *attractor.* Why one attractor rather than another? If the system were Hamiltonian, then the system would be in one attractor or another. In principle we could take into account the effects of all other bodies on Mercury and the sun (assuming the universe is finite, or at least that the effects are finite) and decide with an arbitrarily high degree of accuracy which attractor the system is in. However, given the dissipative nature of the system, it ends up in one attractor or another in finite time. If we examine the boundaries between the attractors, they are fractal, meaning that every two points in one attractor have a point between them in another attractor, at least in the boundary region. This is as if the three-body gravitational problem had to be decided in finite time, which is impossible by numerical approximation (the problem is noncomputable, even by convergent approximation). Therefore, there can in principle be no complete explanation of why the Mercury-sun system is in a 3:2 harmonic. There is approximately a one-third chance of 3:2 capture, one-half chance of a 1:1 capture, and the rest of the harmonics take up the rest of the chances. The chances of a 3:2 capture are good but not that good. The system is obviously physical, but it has a nonreducible property. This property fits Broad's notion of emergence.

How does this apply to the mind? It is highly likely that there are nonlinear dissipative processes in the brain in which the rates of the processes are of the same order as the rate of dissipation. There are also likely to be huge numbers of attractors. The larger the number of attractors, the lower the probabilities of capture in any particular one generally; therefore, a complete reductive explanation seems highly unlikely. This case is certainly true for many biological processes (as in development and in evolution; see Brooks & Wiley 1988; Kauffman 1990). The brain is, after all, biological. We must explain backwards from the attractors that are

formed, that is, downwards from constraints on the constituent physical processes of the order found in the attractors that "win" (Campbell 1974).

But the situation is worse. Certain properties hold a system together (called *cohesion* in Collier 1986; 1988; Collier & Hooker 1999; Collier & Muller 1998). Cohesion is the unity relation for a dynamical system (previous references; Collier 2002). The unity relation is the basis of the identity of an entity. If the property of cohesion is nonreducible, then the object is nonreducible (not the *kind* of object; that can vary). It is certainly possible that the cohesion of the mind, if there is such a cohesive thing, is of this sort. Kim's arguments address ontological deflation (and kinds of objects), not emergence in particular dynamical systems. It is quite possible for an entity to be physical in every respect but not to be reducible in any way that is relevant to complete scientific explanation, even in principle.

ACKNOWLEDGMENTS

I thank INTAS and the Konrad Lorenz Institute for Evolution and Cognition Studies for support while I was doing this research.

Supervenience: Not local and not two-way

James Ladyman

Department of Philosophy, University of Bristol, Bristol BS8 1TB, United Kingdom. **james.ladyman@bristol.ac.uk**

Abstract: This commentary argues that Ross & Spurrett (R&S) have not shown that supervenience is two-way, but they have shown that all the sciences, including physics, make use of functional and supervenient properties. The entrenched defender of Kim's position could insist that only fundamental physics describes causal relations directly, but Kim's microphysical reductionism becomes completely implausible when we consider contemporary physics.

Ross & Spurrett (R&S) point out that the definition of supervenience as (roughly) no change in the supervening properties without a change in the subvening properties, does not imply realizer functionalism (or internalism) unless the relevant subvening change has to occur in the realizer (target article, sect. 2.2). However, they go on to cite Kim (1998), defining supervenience such that if the mental properties of something are to be different, there must be a difference in the physical properties *of that thing.* This appears to rule out externalism, according to which mental properties depend on relations to the environment. If a change in relations does count as a change in the realizer, because relational properties are included in the subvenient base, that reconciles this definition of supervenience with externalism and allows the causal exclusion argument to proceed but with realizer functionalism, not role functionalism, as its target. It seems that Kim's causal exclusion argument relies on local rather than merely global supervenience, but it also seems that local supervenience is less plausible, and certainly the completeness of physics does not entail local supervenience.

A confusing thing about this article is the notion of multiple supervenience and the role it plays in R&S's attempt to reconcile the causal closure of physics with the causal efficacy of supervenient and functional properties. R&S argue that there is two-way supervenience, but they do not show that there is a modal rather than merely an epistemic dependence of, say, physical properties on functional ones. Nothing they say defends the implausible claim that there can be no change in physical properties without a change in mental properties. Rather, they argue persuasively for multiple realizability and the indispensability of functional properties in science.

As R&S diagnose it, Kim's causal exclusion argument threatens to reduce the special sciences other than physics to stamp collecting. To this diagnosis it may be objected that nothing is being taken away from the special sciences by denying that the properties to which they refer in their theories are causally efficacious. After all, the supervenient properties are realized, and the realizers are causally efficacious. Hence, in any concrete case, someone who uses, say, the language of mental states to talk about behaviour and its causes could be regarded as *referring* to physical tokens of the supervenient types, and there are causal connections between those physical states, albeit ones that are of no salience to us. Therefore, according to this response, in "S's belief that *p* caused them to do X," the referent of "S's belief" is a physical state that really does cause the physical state that tokens S's doing X. Saying that beliefs cause actions is elliptical for saying that beliefs are tokened by physical states that cause physical states that token actions. Therefore, it may be argued that the special sciences are tracking a rich causal structure, and therefore doing real science and not mere stamp collecting, but that structure is being described indirectly by means of supervenient properties. Psychology, say, may issue predictions and systematise data in a way that would be epistemically inaccessible to physics, but mental causation is really between physical realizers of mental states. However, this need not be instrumentalism because it may be conceded that supervenient properties are real features of the world and not mere constructs, while maintaining that they only have causal power vicariously.

R&S point out that much of physics is not fundamental and describe properties that are supervenient on atomic and subatomic realizers. Suppose that physics does describe the world by means of supervenient functional properties and that temperature and pressure are examples. There is no doubt that describing the macroscopic properties of a gas in these terms allows for reliable predictions in terms of laws. However, someone of Kim's persuasion could argue that an increase in the pressure of a gas at constant volume does not cause anything; rather, the increase in temperature is a consequence of many microevents that happen to be amenable to a more convenient description than listing them all (and note that there is a physical story to be told about how the universal properties of differently realized macrostates arise). Temperature is a coarse-grained functional property and summarises the statistics of a multitude of microevents. It is a real property but not a causal one. On this view, there is physics, there is stamp collecting, and there is some physics that is stamp collecting.

Which brings us to fundamental physics, which presumably describes the domain where the real causal action is happening in the movements and interactions of microbodies. That quantum phenomena have led to the return of the spectre of action at a distance to physics is well known. This is particularly apposite to metaphysics when local supervenience claims are at issue because arguably what quantum nonlocality requires is not action at a distance *per se,* but the denial of local supervenience. Entangled states of joint systems are just those that violate the principle that the joint state of the whole should supervene on the states of the parts, and, as is well known, Bell's theorem tells us there is no consistent way of attributing states to the parts from which the properties of the joint system can be recovered (without action at a distance). Furthermore, things only get worse for the advocate of microcausation as the only real causation. Quantum field theory does not apply at arbitrarily short-length scales, and researchers in quantum gravity are exploring theories that dispense with spacetime altogether and then try and recover it as an emergent feature of something else. Kim, or anyone who similarly thinks that the real causal processes are only at the fundamental physical level, would then be faced with claiming that there are no true causes in space and time. At that point, if not before, it is surely right to conclude with R&S that the causal explanations of the special sciences are as genuine as those of even fundamental physics.

ACKNOWLEDGMENT

I thank Finn Spicer for comments and discussion.

Causation, supervenience, and special sciences

Graham Macdonald

Department of Philosophy, University of Canterbury, Christchurch, New Zealand; and Department of Philosophy, University of Connecticut, Storrs, CT 06269-2054. **graham.macdonald@canterbury.ac.nz**

Abstract: Ross & Spurrett (R&S) argue that Kim's reductionism rests on a restricted account of supervenience and a misunderstanding about causality. I contend that broadening supervenience does nothing to avoid Kim's argument and that it is difficult to see how employing different notions of causality helps to avoid the problem. I end by sketching a different solution.

The problem of whether properties proprietal to the special (nonphysical) sciences can exert any distinctive causal influence on the world is not new, as shown by the long history of the debate concerning the possibility of genuinely emergent properties. In this context, emergent properties are not just those properties that are possessed by wholes but not by the parts making up those wholes. The controversy concerns whether emergent properties exert a causal influence and have causal powers, which cannot be understood as arising simply from the conjunction of causal powers of the properties of the parts. Jaegwon Kim has sharpened the debate by making explicit some of the assumptions underlying the controversy, in particular the supervenience of special science properties and the causal closure of physics. In simple terms, one set of properties, the A-set, is supervenient on another set of properties, the B-set, when there can be no change in the A-set set without a change in the B-set. If mental properties supervene on physical properties, then one cannot change one's mind without there being a physical change (the precise modal force of "can" and "cannot" is left to one side, but it is important that there is modal force). For our purposes, the causal closure of the physical is the claim that any cause that has a physical effect must itself be physical. Kim's claim is that these assumptions lead to the conclusion that the only way that the supervening properties can be causally efficacious is if they are identical to physical properties. Reduction is the way to go.

Ross & Spurrett (R&S) have done us a service by taking these metaphysical arguments seriously and noting the dire consequences of Kim's conclusion for cognitive scientists. They argue that Kim's pessimistic conclusions rest on a too-restricted view of supervenience and a muddled picture of causation. With regard to supervenience, the complaint is twofold: the picture presented by Kim assumes that there is a one-way dependence of the mental on the physical, and the physical properties forming the supervenience base are localised *intrinsic* physical properties. These two aspects are sometimes treated as though they are the same, but they clearly differ, and differentiating between them is crucial to evaluate their argument against Kim. Putting to one side the issue of intrinsicality (the question is whether relational properties only hold in virtue of the intrinsic properties of the relata), the objection to localised supervenience is that our mental properties can change without necessitating *local* physical change, say, in our brains or bodies. Let us say this is true. It does not follow that supervenience does not hold. What follows is that the set of physical properties can include nonlocal (environmental) properties – supervenience must be broadened. For some reason, R&S think that broad supervenience entails "multiple supervenience," or a two-way dependence of the mental on the physical and *vice versa.* It is difficult to see why this follows. Broadening supervenience just enlarges the base; it does not, by itself, show that there is a two-way dependency relation (or not in any way that goes beyond the covariation implied by supervenience).

The argument from multiple supervenience is taken from Meyering (2000), who uses the fact that a categorical base can realize different dispositional properties (the thermal conductivity-electrical conductivity example) to support the conclusion that there are "emergent effects that have no salience at the level of physics." But unless this is taken to mean that the lack of "salience" is ontological (rather than epistemological), it is difficult to see that Kim's main claim is endangered. Meyering's argument shows that when a categorical base supports different dispositions, *which* disposition is triggered in a particular case depends on the context, the initial conditions. Again, this shows only that specific causes require specific contexts, an unremarkable conclusion, and not one that by itself supports anything Kim would deny. Perhaps it is the terminology ("multiple supervenience") that is getting in the way of understanding R&S's main point here, which seems to concern the essentially *relational* nature of the properties forming the subject matter of the special sciences. With this I have much sympathy (see also Millikan 1999).

The charge that Kim's reductionism rests on a misunderstanding of causality in physics is more convincing, but it is unclear what the consequence is. That is, let us say that Redhead is right that physicists concerned with "fundamental" physics do not use causal terminology and that Loewer is right that deep down it is a matter of the state of the universe at one time compared with the state of the universe at a different time. What are the consequences for our understanding of, say, chemical interactions? Or to our understanding of how our visual system works or how bats echolocate? R&S claim there is no single concept of "cause" that will do the work required by Kim's argument, that he assumes some kind of "folk" concept that incorporates agent-causality, and this is appropriate, if at all, only in some domains and not others. In particular, it is not applicable in physics.

The problem here is that the domains are not separate. R&S are themselves convinced of what could be called the multiplexity of interactions between various facets of the ontologies subsumed by the different sciences. Take a particular chemical process like photosynthesis; this will have a causal explanation invoking chemical and physical properties, and it will also be biologically explainable in terms of the exercise of the functions of those properties. Is the functional explanation also a causal explanation? If it is, one has to show how the two causal explanations of this phenomenon work in harmony and that the invoked causes do not "compete." One does not achieve anything by disuniting the notion of cause operating in the two explanations – or so it seems to me.

R&S are right in seeing Kim's argument and conclusion as radically revisionary of practice in the human sciences. Like them, I believe that the argument generalises to any science whose properties supervene on physical properties, and I agree with them that reduction is not necessarily the way to go. Given my disagreement with their diagnosis as to what has gone wrong, what is to be done? The basic mistake made by Kim is, I suggest, in thinking that only property identity will ensure the causal efficacy of the special science properties. Reduction captures causal efficacy without attendant problems of overdetermination or causal competition between the properties because it ensures the *coinstantiation* of reduced and reducing properties. However, this shows that there is a stopping place short of the reductionist solution and that is *property-instance* identity (coinstantiation) without property identity. Biological (for example) and physical properties can be distinct but still be coinstantiated. The analogy with determinable and determinate properties show how this works; "being colored" is a different property from "being blue," but any instance of blueness is (the "is" of identity) an instance of being colored. It is plausible to hold that where we have supervenience between sets of properties, then whenever the supervening properties are instantiated, they will be coinstanced with the supervened on properties (see Macdonald 1989; Macdonald & Macdonald 1986, for the original presentation of this solution. A more recent elaboration is given in Macdonald & Macdonald 1995). This ensures causal efficacy without overdetermination. It also permits nonreduction, given the distinctness of the supervening properties from the base properties. The remaining problem is: Why does supervenience hold? But that is a different problem (for some thoughts on this, see Macdonald 1992).

Functionalism without multiple supervenience

Ausonio Marras

Department of Philosophy, University of Western Ontario, London, ON, N6H 1T4, Canada. **amarras@uwo.ca** **http://publish.uwo.ca/~amarras**

Abstract: Multiple supervenience is a problematic notion whose role can well be served by a contextualized or properly restricted standard notion of supervenience. It is furthermore not needed to defend functionalism against Kim's charge that cross-classifying taxonomies imply a serious form of dualism; nor does Ross & Spurrett's (R&S's) Kitcherian account of the metaphysics of causation crucially depend on multiple supervenience.

Because *multiple supervenience* is meant to play a large role in Ross & Spurrett's (R&S's) account of the metaphysics and epistemology of special science explanations, it is important to be clear as to what kind of relation it is and how it is supposed to help us resist Kim's reductionist stance. The notion makes its appearance in the context of the authors' response to Kim's (1998) charge that nonreductionists who appeal to the "cross-classification thesis" with respect to the mental and physical taxonomies are committed to abandoning psychophysical supervenience and to embracing "a serious form of dualism" (for supervenience is required for upholding the "causal closure of physics," a minimal requirement for physicalism). Here is what the authors say to this: "According to Kim, [holding the cross-classification thesis] amounts to a denial of supervenience as a one-way relation, permitting what Meyering (2000) calls 'multiple supervenience'" (sect. 3.1, last para.). They then go on to suggest that there are reasons for doubting that multiple supervenience implies any sort of dualism that denies the causal closure of physics. Because, as they later point out (sect. 3.3), Kim never confronts the idea of multiple supervenience ("it's off his radar in so far as it is more powerfully antireductionist than anything he seems willing to consider"; sect 3.3, last para.), their response to Kim suggests that even if he is right in claiming that cross-classification implies the denial of "one-way supervenience," he nonetheless fails to appreciate that this leaves open the possibility of another kind of supervenience, *multiple supervenience,* which (by their lights) is consistent with cross-classification, as well as with the causal closure of physics.

I think there are problems with this response. First, what sort of relation do R&S understand multiple supervenience to be? By contrasting it to "supervenience as a one-way relation," they seem to imply that multiple supervenience is *not* a one-way relation, and by supposing that the possibility of multiple supervenience enables one to "reject [Kim's] implicit premise that supervenience relations must all be 'downward,'" or that they all "point unidirectionally to physics" (sect. 3.2, para. 2), they seem to imply that *multiple* supervenience may point *upwards,* in the opposite direction than the standard sort of supervenience entailed by multiple realization. I think this is a confusion. All supervenience, multiple or otherwise, is a "one-way," unidirectional relation from the higher (functional) level to the lower (realization) level if conceptualized as a *dependence* relation, and from the lower to the higher level if conceptualized in terms of a relation of *determination.* The only difference is that the *mapping* effected by standard supervenience is a *one-many* mapping (at least if multiple realization is involved), whereas in the case of multiple supervenience, the mapping is *many-one:* multiple higher-level properties supervene on the same base property. No doubt R&S must have meant something of the sort; for surely the "direction of determination" (or, conversely, the "direction of dependence") remains the same in both cases.

Second, the idea of *multiple* supervenience so characterized is, strictly, incoherent. Consider two distinct, nonequivalent higher-level properties M1 and M2, and suppose that something x exemplifies M1 but not M2 at $t1$ and M2 but not M1 at $t2$ (i.e., suppose that x has undergone a change with respect to its M properties). Multiple supervenience would have us suppose that there might be a base property, P, on which *both* M1 and M2 supervene. However, that is impossible: by definition, supervenience requires that there cannot be a change with respect to the supervening properties without a corresponding change with respect to the subvening properties. One could fix this by imposing certain restrictions, for example, by requiring that the supervening properties be coextensive (where none can be exemplified without the others being simultaneously exemplified), by relativizing them to a given context (as would be natural in "Twin-Earthian" cases) or interpretation scheme (as when the same physical process in a computer implements different programs), or by broadening the supervenience base so as to include the appropriate contextual conditions. However, then it is not clear that the notion of *multiple* supervenience does any work that cannot be done by the standard notion of supervenience, locally or nonlocally construed.

Third, multiple supervenience is, in any case, not needed to answer Kim's challenge from cross-classifying taxonomies. We can have cross-classification *either* when we can make distinctions in terms of the higher-level properties that we cannot make in terms of the base properties, *or* when we can make distinctions in terms of the base properties that we cannot make in terms of the higher-level properties, *or* both. Now it is clear that when we are dealing with higher-level *functional,* and, in particular, *mental* properties, it is the *second* of the aforementioned options that is the relevant one, for it is of the essence of functional/mental properties that they be (at least in principle) *multiply realizable.* However, that implies that there are distinctions that can be made by the *base* (or *physical*) taxonomy that cannot be made by the functional/mental taxonomy, and that is just to say that the former supervenes on the latter. Therefore, cross-classification, in so far as it pertains to the functional/mental taxonomy *vis-à-vis* the physical taxonomy, does not violate supervenience and thus entails no "serious form of dualism." Conversely, the first and the third options above do entail the denial of standard supervenience: they represent precisely the sort of situation envisaged under *multiple supervenience* (hence, my earlier claim that unrestricted multiple supervenience is not supervenience at all). Far from providing a way to meet Kim's challenge from cross-classification, multiple supervenience falls prey to just that challenge.

Fortunately, then, functionalism does not have to depend on multiple supervenience to prove its metaphysical credentials, nor do R&S's valuable insights about the autonomy of functionalist explanation in the special sciences. Indeed, what does all the interesting work in their defense of functionalism against Kim's epiphenomenalist challenge is the unfolding of the Kitcherian idea that the metaphysics of the attribution of causal powers cannot be divorced from the epistemology and methodology of explanation, whose holistic, unificatory, and highly contextual character has no reflection in Kim's "conservatively metaphysical" conception of causation. Whether this idea is itself ultimately defensible is, of course, another matter.

Really taking metaphysics seriously

Barbara Montero

Department of Political Science, Economics and Philosophy, The College of Staten Island of the City University of New York, Staten Island, NY 10314. **barbara@antinomies.org** **http://barbara.antinomies.org**

Abstract: Ross & Spurrett (R&S) fail to take metaphysics seriously because they do not make a clear enough distinction between how we understand the world and what the world is really like. Although they show that the behavioral and cognitive sciences are genuinely explanatory, it is not clear that they have shown that these special sciences identify properties that are genuinely causal.

Ross & Spurrett (R&S) claim to be taking metaphysics seriously, but I doubt metaphysicians such as Kim would agree. Taking metaphysics seriously means in part making a distinction between

how we understand the world and what the world is really like, that is, between explanation and ontology, and it seems that even if R&S have shown that the behavioral and cognitive sciences are genuinely explanatory, as I think they have, it is not clear that they have shown that these special sciences identify properties that are genuinely causal. As such, R&S's article fails to convince the serious metaphysician who is persuaded by Kim's causal exclusion argument that mental properties can perform real causal work.

Explanations in nonfundamental sciences, including much of physics, as R&S point out, are frequently not entirely bottom-up. Moreover, as they also argue, it is not at all clear how one could eliminate top-down explanations. For it does seem that when we substitute explanations in terms of neural states for explanations in terms of beliefs and desires, we lose the very phenomenon we are trying to explain. But is this a point about our cognitive abilities or a point about the way the world works? That is, is it an epistemological point or a point about the ontological nature of beliefs and desires?

R&S take the ineliminability of top-down explanations to show something about the way the world works because they take the connection between explanation and ontology to be tight. In fact, they claim that an explanation is not something that is merely psychologically satisfying, "but must cite explanans that are . . . true" (sect. 3.1, para. 1). If this were the case, the fact that the cognitive and behavioral sciences are not explanatorily irrelevant would also show that they could carry their ontological weight. However, although I agree that an explanation should be something more than merely psychologically satisfying, requiring that the explanans be true would rule out many, if not most, of our current scientific explanations from counting as explanations because many, if not most, of our current scientific explanations are probably false. For example, Newton's laws are taken to be explanatorily powerful yet are known to be false. And most likely, given the history of scientific theorizing in the hard and especially the soft sciences, it is likely that much of our currently accepted theories, which are employed to explain various phenomena, will turn out to be false. Thus, if we require explanations to cite explanans that are true, we have to admit that probably science is not explaining much of anything, which I would think R&S, being themselves good naturalists, would not want to do. Therefore, while requiring explanans to be true weds explanation to ontology, it does so at a high price. Once we give up the requirement that explanans must be true, however, we have a gap between scientific explanation and how the world really is, a gap that a savvy metaphysician such as Kim can attempt to pry open.

It is a distinct question whether Kim has pried open the gap between how science explains the world and how the world really works, showing in effect that we must be mistaken either in our belief that the special sciences traffic in causal properties or in our belief that explanations in the special sciences are in some significant sense irreducible. I happen to think that Kim has not done this. R&S claim that the causal exclusion problem turns upon there being a clear-cut notion of causation in fundamental physics. However, I do not think that it does. Kim can avoid talk of fundamental physics because the causal exclusion argument can be reformulated as a problem about the apparent overdetermination of the neural and the mental. Arguably, neurophysiology is causal (neurophysiologists, at least, do make causal claims), and it also seems likely that once we set the neurophysiological cause, one does not need to add anything mental to produce the desired effect. Therefore, it would seem that R&S's well-grounded skepticism about causal concepts in the domain of fundamental physics is beside the point.

Although this reinterpretation of the causal exclusion argument cannot be faulted for assuming that there is a clear-cut notion of causation in fundamental physics, it can be faulted for another reason. As I see it, while systematic causal overdetermination may be metaphysically profligate when the causes at issue are relevantly distinct, such as when a man is simultaneously shot and suffers a heart attack, and as such, his death is caused twice over, mental causes in as much as they are constituted by neural causes are not distinct in this way.[1] Is there any reason to say that the neurophysiological and not the mental does the real causal work? Certainly there is no more reason to say this than to say that aspirin does no real causal work and that only the ingredients of aspirin do. Since we need not reject aspirin's causal powers, we need not reject that the mental gives us real causal powers. Because of this, sciences trafficking in such causes are doing more than mere stamp collecting.

In responding to the causal exclusion argument in this way, am I trying to get a free lunch? I think not. The response does not reject the causal exclusion argument merely because it is general and thus, if successful, would not only render the mental causally profligate but also virtually all other phenomena save for those at the level of fundamental physics. Rather, the response provides a metaphysical distinction between properties that cause problematic overdetermination and properties that do not. As such, it seems to me to be a much more straightforward way to address Kim's causal exclusion argument and, at the same time, to take metaphysics seriously.

NOTES

1. Melnyk (2003) argues for this point.

The vessels and the glue: Space, time, and causation

Andrei Rodin

Department of Philosophy, Ecole Normale Superieure, 75230 Paris Cedex 05, France. **rodin@ens.fr**

Abstract: In addition to the "universal glue," which is the local mechanical causation, the standard explanatory scheme of classical science presumes two "universal vessels," which are global space and time. I call this outdated metaphysical setting "black-and-white" because it allows for only two principal scales. A prospective metaphysics able to bind existing sciences together needs to be "colored," that is, allow for scale relativity and diversification by domain.

If our world could be satisfactorily accounted for by a single science, then we would not need to distinguish a particular science of metaphysics or any other particular science. Because this is not the case and we have numerous sciences that cannot be reduced into one *trivially* (to say the least), we need metaphysics to work on gluing those sciences together, be the *glue* some kind of reduction to universal physical laws or something else. Aristotle invented metaphysics (which he called *first philosophy*) to bind physics (by which he meant broadly the study of all natural phenomena) with mathematics and logic (so afterwards the latter two disciplines could be considered as tools for the former). Because Aristotle's physics has branched into numerous disciplines, our need for the unifying science of metaphysics is even stronger than Aristotle's. A scientist calling for a *free lunch* has two options: either to take uncritically the nostalgic dogma of reductionism according to which in the distant future all sciences will collapse back into physics (to leave aside unification dogmas borrowed from outside of science), which is epistemologically irresponsible, or to give up the idea of unity of science, which turns science into a combination of mystery and stamp collecting. If no reasonable and testable reductionist hypothesis can be made *now,* then this is a job of metaphysicians to suggest tentative ways to glue sciences by means other than reduction. It goes without saying that working on binding sciences together a metaphysician must have a good understanding of what he or she is going to bind. Otherwise, the unifying efforts of a metaphysician will be simply ignored by the scientific community and for good reason. As Ross & Spurrett (R&S) show, this unpleasant situation is not uncommon even for the mainstream metaphysical discussion.

Now let me be more specific about the *glue.* R&S label as " localist metaphysics" and "localist paradigm" a generalised explana-

tory pattern of classical (and hence outdated) mechanics where the global dynamics is reconstructed from local interactions of point masses, and those interactions are interpreted in causal terms. Two remarks are in order here. First, a historical one. The Cartesian idea of explaining global dynamics in terms of strictly local *pushing* never worked well. Newton's gravitational *pulling* is a long-distance, not a strictly local, interaction. This made the gravitational force an extremely doubtful concept in the eyes of Newton's contemporaries (Leibniz [1890] expressed his misgivings on this point in the form of bitter irony.), and this concept was formally dispensed with by the introduction of the Lagrangian and Hamiltonian formalisms (which did *not* aim to meeting the Cartesian localist requirement, however; therefore, Redhead's point could be made even within classical mechanics without his reference to general relativity). Second, and more important, if we ask what binds things together when one applies the explanatory pattern of classical mechanics, then the answer that this is the local or pseudo-local interaction of point masses interpreted in causal terms will be only partial. Obviously this role is also played by *space* and *time.* Because *glue* is a localist metaphor we shall call space and time (in the Newtonian absolutist sense) *vessels.* Apparently the *vessels* work better across disciplines than the *glue* (localist causation): Although it remains a controversial point whether we can and should specify one type of causality working across all special sciences or specify parochial types of causality particular to given disciplines, or both, the idea that every material entity or process exists (or occurs) in the *same* physical space and time (or space time) sounds commonsensical. Moreover, it is interesting that tentative parochial space-time concepts, in particular biological ones, are also known, although they remain marginal (Vernadsky 1988).

Whereas causality (the glue) within the classical setting is local or pseudo-local, space and time (the vessels) are *global* in the sense that they supposedly allow for locating all possible point masses and all possible events in a *uniform* way (as R&S put it, "measurement values are not indexed to neighborhoods of points"; sect. 4.4, para. 7). Therefore, the interplay between the local and the global scales in the classical framework involves the vessels *and* the glue. However, this framework does not allow for any *intermediate* scale: we have global space and time comprising everything, and structureless point masses interacting only locally (therefore, the idea of long-distance interaction does not exactly fit the paradigm). For this reason, I doubt that "localist metaphysics" is an appropriate term to characterize the described setting. I suggest "*binary* metaphysics" to stress the fact that it allows for only two principal scales.

I cannot discuss here details of any tentative metaphysics that could be relevant to contemporary physics and other sciences. Apparently it should be *not* binary but allow for *scale relativity* and perhaps be also diversified by domain. The philosophical literature discussing space-time concepts of the fundamental physics is vast, but it is not always accurately considered in metaphysical discussions. (A common mistake is to limit discussion to *special* relativity, whereas only general relativity is a full-fledged theory of spatiotemporal dynamics). It is more difficult to say what is going on with the concept of causality in contemporary physics just because, as R&S note, it apparently does not play any essential role there. Perhaps this is too easy an answer. It is more useful to study the evolution of the concept attentively than just to say that it dies off. For (to put it in functionalist terms), the *role* played by the concept of causality, namely, the role of *glue,* apparently remains essential, and if classical causality dies off, then this or a similar role must be taken by something else. Reichenbach's early attempt to reconstruct causality in terms of "marks" and its development by R&S in terms of information processing are promising. I would like to note the fact that Reichenbach's suggestion about causality is hardly separable from his analysis of relativistic spacetime.

ACKNOWLEDGMENT

I thank Gillian Barker for improving my English and giving valuable advice.

"Causation" is only part of the answer

Matthias Scheutz

Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN 46556. **mscheutz@cse.nd.edu**
http://www.nd.edu/~mscheutz/

Abstract: Although Ross & Spurrett (R&S) successfully fend off the threat of Kim's "supervenience argument" by showing that it conflates different notions of causation, their proposal for a dynamic systems answer to the mind-body problem is itself yet another supervenience claim in need of an explanation that justifies it. The same goes for their notion of "multiple supervenience."

The so-called "supervenience argument" (Kim 1998) or "causal exclusion argument" (Block 2003), if true, would end the autonomy of the special sciences by reducing them to physics, and physics would become the only game in town. Fortunately, the "argument" is fraught with problems, from an oversimplified and imprecise notion of supervenience (that typically only involves physical and mental *properties* but not *n*-ary *relations* for $n > 1$, which are the more important and also much more difficult part to tackle) to an inadequate understanding of the place of "causation" in physics and the special sciences. This is why attempting a rebuttal of the "argument" is challenging, because it is not even clear where to start (although see Block 2003, for a response that puts the ball back in Kim's court).

Whereas many responses have focused on the notion of supervenience (although it is not clear how to define it in a precise way; however, see Humberstone 1998, for a start), Ross & Spurrett (R&S) take on a flaw that can be more easily exposed: the underlying notion of "(physical) causation." They convincingly argue that Kim's view of causation does not square with current physical orthodoxy (e.g., physical laws in quantum mechanics and quantum electrodynamics do not involve "causation"). Unless Kim can make clear what he means by "physical cause" in a compelling way, consistent with the best current physical theories, for all practical matters, the case for a general, ultimate notion of causation (to which all other practically used forms have to report) can be put to rest.

However, Kim is right, and R&S agree, that mind-body supervenience alone cannot account for how mental properties are related to physical properties because it merely states the problem but does not solve it (Kim 1998, p. 14), and, although the reductionist phalanx has (temporarily) come to a halt, by simply putting causation back into the special sciences (but, say, without an account of the mind-body supervenience) the nonreductionist counteroffensive has not yet been initiated.

R&S propose dynamical systems theory applied to feedback-driven servosystems as a way to move forward. In particular, they claim that "we have, up to an important standard of generality, explained how mental properties and physically non-problematic properties are related if we produce a broad account of feedback-driven servosystems and the ways in which evolution has built nervous systems that support them" (sect. 5, para. 14). Unfortunately, this way of putting it is no better than simply stating that mental states supervene on feedback-driven servosystems; it is yet another supervenience thesis formulated for a particular supervenience base.[1]

We agree that dynamical systems theory will likely lead the way (at least in the beginning) but not without having to face serious challenges such as the question about how to isolate proper "inner states" that ground functional descriptions of the system and figure in causal explanations of the system's behavior.

To see this, first note that dynamical systems descriptions of the behavior of a given physical system (consisting of sets of differential or difference equations) are directly derived from physical laws, which reflect the ways in which energy can be stored and transferred in the system (Kutz 1998, p. 796). In these dynamical systems, inputs correspond to energy sources, and outputs correspond to physical variables that are to be measured or calculated. It is then possible to obtain a special "I/O form" of these differ-

ential equations "by combining element laws and continuity and compatibility equations in order to eliminate all variables except input and output" (Kutz 1998, p. 808). This is all in line with R&S's view that the behavior of a physical system can be described by dynamic systems theory and that causation is not necessary for that description (at least for some physical theories). The problem with such a description, however, is that the behavior of the entire system (for all initial conditions) is fixed by a description *without inner states* (i.e., non-input/output state variables). Yet, without inner states it is not clear how to warrant "causation talk" other than of the behaviorist kind – that would have thrown the baby out with the bath water. Worse yet, there are infinitely many different sets of equations of finitely many different non-input/output variables that give rise to the same I/O form. Put differently, there are dynamical systems that describe the behavior of a given system perfectly without having a single non-input/output variable correspond to any "natural candidate" (e.g., energy sources or energy sinks) of an "inner state" of the system (for details, see Scheutz 1999a).

The upshot of all of this is that dynamical systems theory *per se* does not, as R&S seem to suggest, provide a straightforward answer to the question whether a given physical system *realizes* a given functional architecture (Scheutz 2001), nor to the question what functional architecture(s) the system realizes (Scheutz 1999b). Kripke, for example, expresses this worry for program descriptions (i.e., that a physical machine can only "approximately" or "imperfectly" realize an infinite function) because "indefinitely many programs extend the actual behavior of the machine" (Kripke 1981, pp. 33–35). The above shows that the same is true for dynamical systems, and *a fortiori* applies to functional explanations built on or derived from them.

Note that this problem with dynamical systems is different from what R&S seem to mean by "multiple supervenience," which they take to be responsible for being able to grant supervenient properties their explanatory relevance: although the infinitely many functional architectures "induced" by the "inner state variables" all realize essentially the "same architecture" (where "same" has to be spelled out in terms of an extension of the notion of "bisimulation" defined for whole trajectories in state space instead of mere state transitions; see also Scheutz 2001), multiple supervenience seems to allow for non-bisimilar functional architectures to supervene on the same physical system – now *that* is spooky.

NOTES

1. It is also not clear what work the qualifier "broad" is supposed to do: it seems perfectly plausible that one could know all facts about feedback-driven servosystems (e.g., in the sense worked out by "control theory") and still not understand at all how these facts pertain to minds (i.e., how control states are related to mental states).

Functionalism, emergence, and collective coordinates: A statistical physics perspective on "What to say to a skeptical metaphysician"

Cosma Rohilla Shalizi

Center for the Study of Complex Systems, University of Michigan, Ann Arbor, MI 48109. **cshalizi@umich.edu** **http://bactra.org/**

Abstract: The positions Ross & Spurrett (R&S) take on issues of information, causality, functionalism, and emergence are actually implicit in the theory and practice of statistical physics, specifically in the way it relates macroscopic collective coordinates to microscopic physics. The reasons for taking macroscopic physical variables like temperature or magnetization to be real apply equally to mental properties like pain.

Foundational questions of the kind Ross & Spurrett (R&S) worry over often don't matter much to scientists, but sometimes they matter a great deal and shape the kind of research we undertake. The answers to R&S's questions about information, causation, functionalism, and emergence matter a great deal to cognitive science. The position they argue against would inhibit not only cognitive science but also my own field of statistical physics. Moreover, we statistical physicists implicitly rely on what is essentially R&S's combination of Salmon (1984) and Dennett (1997). Therefore, I think their answers are basically right, and the skeptical metaphysician is wrong.

Consider a macroscopic physical system. It consists of many particles, each with three degrees of freedom in position, plus three in momentum, and possibly some internal degrees of freedom. (For simplicity, I'll ignore quantum mechanics.) The total number of degrees of freedom is N, and the dynamics of the system are described by equations of motion in this N-dimensional state space.

Thus far, each coordinate belongs to a particular particle. However, we are free to change our coordinate system as long as the transformation is invertible. Each degree of freedom in the new coordinates need not, and generally will not, belong to a single particle. Rather, it can be a *collective* coordinate, a function of the state of many particles, or even (like the center of mass) of all the particles (Forster 1975). The macroscopic variables that appear in physical theories are collective degrees of freedom: temperature, pressure, molecular concentrations, fluid velocity, stress, vorticity, current, and order parameters. To specify the value of one of them is to say that the system is in some particular region of the microscopic state space.

The advantage of such collective coordinates (beyond ease of measurement) is that often a fairly small number of them (m, say) interact with each other so strongly that their dynamics can be described by a deterministic evolution plus a comparatively small noise term. The noise is the effect of the remaining $N - m$ degrees of freedom on the macroscopic variables and often vanishes in the limit of large N. The macroscopic variables are then said to give a "coarse-grained" description of the system. Properly constructed, the coarse-grained variables satisfy Salmon's (1984) criteria for being a "statistical relevance basis" (Shalizi & Moore 2003). They can definitely store and transmit information over time. Moreover, they satisfy the counterfactual criteria for causality proposed by statistics and AI (Pearl 2000). The macroscopic, coarse-grained description is less precise than the microscopic one, but simpler and accurate to within a level specified by the noise. Theories in statistical mechanics start with a model of the interactions among the microscopic degrees of freedom in some system and then calculate its behavior at the coarse-grained level, including the perturbations caused by the ignored degrees of freedom (Chaikin & Lubensky 1995; Forster 1975; Keizer 1987).

Coarse-grainings that allow us to trade off complexity for accuracy are not unique. There are generally multiple levels of more or less detailed descriptions, *all* simultaneously valid for the same physical system. For instance, one can describe a fluid at a "thermodynamic" level, using quantities defined over the whole fluid, and a "hydrodynamic" one, using local currents and densities of those quantities (Keizer 1987). The thermodynamic description is a coarse-graining of the hydrodynamic one, which in turn is a coarse-graining of a more detailed molecular level. Here, one can show that the coarser levels are more predictively efficient (i.e., each bit of macroscopic information delivers more predictive information at the higher levels than the lower ones; Shalizi & Moore 2003). This gives a natural, non-mysterious definition of emergence, and one imagines it would apply nicely to mental phenomena, with (perhaps) an intentional-system level emerging from a symbolic-cognitive level, in turn emerging from a neuronal, connectionist one, and so forth down through the calcium channels to crawling molecular chaos. At each stage, we have collective coordinates of a physical system, capable of storing and transmitting information, subject to noise.

If multiple instantiation is a worry, then most of what we ordinarily consider physical quantities are in trouble. Take electric

current and temperature. A current of 1 ampere can be instantiated by a certain number of electrons per second going one way, just as many hydrogen ions going the other way, and half as many calcium ions going the same way as the hydrogen, even moving "holes, propagating absences of electrons. Similarly, the property "temperature T = 300 kelvins" is instantiated by many different microphysical configurations and properties, involving momenta, spins, charges, hydrogen bonds, gravitational potentials, and so on. Many important macroscopic variables can equally well be defined as coarse-grainings or through *functional* properties relating to other macroscopic variables. An active area of statistical physics exploits the functional definitions of thermodynamic variables, abstracting ordinary thermodynamics into a purely formal structure (Ruelle 1978), and then constructing quantities that satisfy its axioms in various dynamical systems. This "thermodynamic formalism" has proved its worth in understanding chaotic dynamical systems (Beck & Schlögl 1993), hierarchical structures (Badii & Politi 1997), and turbulent flows (Chorin 1994).

To summarize, everybody agrees that things like temperature and current are physical quantities, but that they are multiply-instantiated, coarse-grained macroscopic constructions. The arguments that say mental properties are at most epiphenomenal thus apply to them, too. Against this, specifying the values of such quantities has considerable predictive power, and one can give self-contained accounts of their dynamics, subject to a certain level of noise. The extra noise and imprecision of the collective coordinates over the microscopic ones is more than offset by the gain in simplicity. They are "real patterns" (Dennett 1997). However, all this is just as true of mental properties, which are also (presumably) emergent, coarse-grained collective degrees of freedom of physical systems. There is just as much reason to treat *pain* as real and causal as to consider *electric current* so. It is not just the special sciences that need functionalism; physics needs it, too, and uses it, although we generally call it reductionism.

Protecting cognitive science from quantum theory

David Wallace

Philosophy Department, Oxford University, Oxford OX1 4JJ, United Kingdom.
david.wallace@magd.ox.ac.uk

Abstract: The relation between micro-objects and macro-objects advocated by Kim is even more problematic than Ross & Spurrett (R&S) argue, for reasons rooted in physics. R&S's own ontological proposals are much more satisfactory from a physicist's viewpoint but may still be problematic. A satisfactory theory of macroscopic ontology must be as independent as possible of the details of microscopic physics.

I find myself in close agreement with Ross & Spurrett (R&S) in the main claims of their paper; I shall confine my comments to some observations about the role which physics plays in their discussion.

R&S rightly criticise Kim's mereological definition of macroproperty for a general term like "water," but the criticism can be sharpened: Even a particular object like a table cannot really be regarded as a simple composite of non-overlapping microscopic parts. It's a tempting idea, to be sure: An extended body is just the mereological sum of its top and bottom halves; therefore, why not subdivide indefinitely until we get to the microconstituents? However, a solid object is a cloud of vastly many overlapping electron and nucleon wave functions: it is not clear even what is *meant* by saying which electron is in which spatial subregion of the object. There are ways around this problem, but they rely on dangerously strong assumptions about the present or future state of physics. (There are interpretations of quantum mechanics, for example, Bohm [1960], in which particles are something like the tiny billiard balls that philosophers treat them as – but do we really want to rest our ontology on contentious claims in quantum mechanics?)

Furthermore, even the paradigmatically "physical" properties of the object are defined not in terms of the microconstituents, but dispositionally – even the mass (!) of a solid object cannot really be defined as the sum of the masses of its atomic constituents. That algorithm gets the answer nearly right in most cases, but a helium nucleus weighs approximately 1% less than its constituents (that's why fusion works); a neutron star weighs approximately 10% less (Arnett 1996) than its constituents (that's why supernovas work). Our actual definition of mass is dispositional: Something has mass m if it behaves thus-and-so on the scales, or creates such-and-such a gravitational field. It is not definitional that mass is additive; it is a physical law, and only an approximate one at that.

This raises the stakes a bit, I think. R&S argue that Kim's account cannot correctly handle the natural kinds of the special sciences. However, it is actually worse: the account (I am claiming) correctly handles *hardly any macroproperty at all.*

This makes the pattern-based view of ontology espoused by Dennett (1991b), and defended by R&S, very attractive. Of course, there must be some sense in which macroscopic objects are built out of microscopic constituents and in which they are supervenient on the properties of the constituents. Dennett, by regarding macro-objects as *patterns in* the micro-ontology, rather than as *mereological sums of* that micro-ontology, provides the sort of account of compositionality that is not hostage to contentious or downright false pictures of physics.

But of course, if such an account is adopted for the whole of macro-ontology, then mental states are real in the same way that tables are real, and the causal power of the mental stands and falls with the causal power of almost everything. This would be close to a *reductio* of Kim's argument: If we are sure of anything about causation, we are sure that macroscopic objects causally influence other macroscopic objects. *Maybe* there is some esoteric notion of "causation" that applies to the ultimate microconstituents of nature only, but that notion can have little to do with "mental causation" as ordinarily understood.

Having supported R&S thus far, I wish to make one cautionary remark about their project. At times, R&S write as though the goal of a pattern ontology is to find, once and for all, the correct notion of substrate; and then define real patterns as patterns in that substrate. (This seems to be the context for their approving citation of Nottale's "fractal space-time" work; target article, sect. 4.4, para. 7) This I find dangerous: It bets our metaphysical structure on the current state of fundamental physics, despite the fact that fundamental physics frequently changes. Are "real patterns" patterns in particle distributions? Then we implicitly bet against an underlying field ontology in which particles themselves are patterns. Are "real patterns" patterns in the distribution of properties over space-time? Then we implicitly bet that space-time is fundamental (*contra* many proposals in quantum gravity) and that its role in fundamental physics is roughly the same as its role in classical physics (*contra* at least some interpretations of quantum mechanics, such as the many-worlds theory; see Wallace 2003). The danger is only heightened if we try to base metaphysics on speculative physics such as Nottale's.

One way around this problem may be to look for a sufficiently abstract characterisation of pattern as to be immune to revisions in microphysics. R&S's proposed information-theoretic approach may well succeed here, although I worry about its appeal to thermodynamic concepts like entropy: thermodynamics itself is an emergent phenomenon; therefore, there is some danger of circularity here. Another, more modest proposal would be to adopt a hierarchical view of pattern ontology: if we accept some stuff into our ontology, we should also accept patterns in that stuff. If the stuff itself turns out to be patterns in substuff, so be it. Thus, particles are patterns in the quantum field; humans are patterns in the particles; stock market crashes are patterns in the people; and so on. Such a metaphysics would be robust against, and relatively

uninterested in, the discovery that the quantum field itself is just a pattern in something deeper.

My intention in this commentary is not to argue that cognitive scientists and philosophers of psychology should add quantum mechanics to the already formidable range of disciplines they are required to learn. In a sense, the reverse is true: Modern physics is so alien, and so changeable, that unless metaphysics is to be postponed until a completed physics is available, then we need an ontology of macroscopic objects that is largely independent of microphysical detail. Surely such an ontology exists: The hard-won generalisations of psychology or economics cannot plausibly be hostage to details of space-time structure at submicroscopic scales. However, it is surprising how many superficially innocuous metaphysical ideas actually fail this test of independence.

Authors' Response

The cognitive and behavioral sciences: Real patterns, real unity, real causes, but no supervenience

Don Ross[a] and David Spurrett[b]

[a]*Department of Philosophy, University of Alabama at Birmingham, Birmingham, AL 35294-1260, and School of Economics, University of Cape Town, Rondebosch 7701, South Africa;* [b]*School of Philosophy and Ethics, University of KwaZulu-Natal, Durban 4041, South Africa.*

dross@commerce.uct.ac.za spurrett@ukzn.ac.za

http://www.commerce.uct.ac.za/economics/staff/personalpages/dross/

http://www.nu.ac.za/undphil/spurrett/

Abstract: Our response amplifies our case for scientific realism and the unity of science and clarifies our commitments to scientific unity, nonreductionism, behaviorism, and our rejection of talk of "emergence." We acknowledge support from commentators for our view of physics and, responding to pressure and suggestions from commentators, deny the generality supervenience and explain what this involves. We close by reflecting on the relationship between philosophy and science.

R1. Introduction

How are the behavioral sciences related to each other and to the rest of the sciences? More specifically, how do sciences other than physics relate to physics, and what is the status of claims about causation in the same systems when multiple causal claims are made by different sciences? In our target article we describe a recent wave of metaphysical work which suggests that sciences besides physics, especially those pursuing functionalist research strategies, are importantly defective compared with physics, that their causal claims are otiose (or, as one commentator [**Boersema**] puts it, "incorrect") unless they can be reduced to physical claims, and that the costs of such reduction are worth paying to establish causal relevance for the sciences in question. We argue against all these suggestions. Physics is importantly different from what the metaphysical challenge assumes, in part by itself being functionalist and in part because there is no reason to suppose that it is the home of some master concept of causation to which other sciences are answerable, and compared to which other causal claims are automatically defective. The costs of imposing intertheoretic reduction on the behavioral sciences would be prohibitively high, but – and partly *because* – physics is not what many metaphysicians (and others) assume, causal claims made by special, including behavioral, sciences are not cornered into choosing between irrelevance and reduction.

Before engaging directly with the set of commentaries, we observe that some aspects of our argument were not challenged by any of the commentators. In particular, none (although see sect. R4) attempts to argue that *reductionism* of the sort at issue is desirable or even less undesirable than we argue. To the extent that our argument relies on defending a view of how things are with *physics,* the commentaries provide nothing but support (see sect. R5).

Part of our answer to the question about the relationships between the behavioral and other sciences concerns *scientific unity.* Some commentators seek clarification of our commitments or subject them to challenge, and we respond below (sect. R2). A number of commentaries light on a commitment to *realism* relied on in our argument but not given full defense in the target article. A brief case for realism to complement the target article follows (sect. R3) the discussion of unity. Although we are wary of the term "*emergence,*" it crops up in the titles of two commentaries and in the text of a third. There are different conceptions of emergence and a related risk of confusion given the range of senses of "reduction" in use in philosophy of science and by scientists. We attempt (sect. R4) to make clear why we prefer to eschew emergence talk and in what senses we are not reductionists. One commentator is concerned that our position is tantamount to behaviorism. We make clear (sect. R6) that it is supposed to be.

A striking feature of the commentaries taken as a group is the widespread and generally critical attention given to our claims about "*multiple supervenience.*" In this case we can neither thank commentators for support nor simply attempt to clarify and refine our explicit position. Rather, we concede that our position as described in the target article is flawed and attempt to replace it with something better (see sect. R7).

The concerns of the commentators are mostly philosophical, with the second most popular topic being physics rather than the behavioral sciences. While doing our best to engage directly with the points raised by the commentators, in what follows we seek throughout, as in the target article, to connect discussion directly and nontrivially with the behavioral sciences. It is worth bearing in mind that the motivation for the target article and this response to the commentaries is to answer a metaphysical challenge to the effect that the behavioral sciences are ontologically confused and faced with a difficult choice between going ahead as usual, but in so doing abandoning any claim to making genuinely causal explanations or dismantling much of what has been achieved to salvage the capacity to make causal claims, but only while wearing a reductive straightjacket.

We also note that although our project is conservative in the sense that we seek to protect existing sciences, it is not *merely* conservative – the epistemological status, the ontological scope, and the nature of the relationships between the behavioral sciences are subject to serious interrogation and fundamental revision. Therefore, we need to satisfy two different sorts of criteria if our project is to be judged a success. One is to convince philosophers that we have de-

fused the challenge of skeptical metaphysicians. The other, equally important, is to satisfy behavioral scientists that the vision we outline provides a congenial home for their ongoing work. Either by itself just is not good enough.

R2. Unity of science

One way of understanding the motivations behind our article and our reasons for thinking that the metaphysical issues it discusses should be relevant to cognitive scientists is by reference to a concern with *scientific unity.* We presume that it is important for all sciences that the claims they seek to justify be integrated with a wider world picture, because this is what it means for a body of scientific claims not to be *mysterious.* This consideration is especially significant in the cognitive and behavioral sciences, for two reasons. First, the project of understanding mind and behavior is the responsibility of a coalition of disciplines with distinct histories, so unification issues arise *within* the explanatory enterprise, rather than only between it and neighboring domains. Second, cognitive science studies precisely the domain that has been most explicitly taken by folk thought and by a long tradition in philosophy to be explicitly *dis*united from others, by virtue of the conceptually problematic relationship between minds and brains.

Boersema says that he finds "an underlying commitment to unity in [our] rejection (target article, Note 20) of Cartwright's and Dupré's criticisms of . . . unity." We regret having made Boersema, and presumably some other readers, work to find this because the commitment is fundamental to the point of our project. If a cognitive scientist had no concern for scientific unity, then he or she would be right to regard Kim's critique and the issues associated with it as being of little interest. After all, neither Kim nor other metaphysicians we have called scholastic are urging people to *stop doing* cognitive science in favor of doing physics or some other so-called lower-level study. Rather, Kim's claim is that unless mind is understood reductionistically, it cannot be unified with an intuitive conception of the physical. It would follow from this that if cognitive science studies something coextensive with our intuitive concept of mind and if physics studies something coextensive with our intuitive concept of the physical, then cognitive science cannot be unified with other disciplines. Therefore, Kim's is an argument for reductionism addressed to people who are presumed to value unity, either of the common-sense ontology alone or of both it and our scientific ontology. Our criticism of Kim's argument is addressed to the second set of people.

This is not the place for us to try to mount an argument intended to convince the scientist who does not value unity that he or she should. (We attempt a certain limited amount of such persuasion in the article.) Let us briefly indicate why we are prepared to be prescriptive about this. It is not coherent to value scientific explanation while not valuing scientific unity. To disavow concern for unity is, as a matter of logic, to value science exclusively for its facilitation of prediction and control, that is, to appreciate science just for what it shares with engineering. We doubt that most scientists are, or could be, exclusively motivated that way.

Boersema wonders what kind of unity we are worried about and which kind we think reductionists are committed to that we are not. This seems confused. Neither what Boersema identifies as "methodological" unity nor "unity of values" is *at all* relevant to the problem Kim's argument raises for practitioners of special sciences. (For what it's worth, we are skeptical about methodological unity because we think that successful science is generally methodologically opportunistic. We are therefore "meta-skeptical" about methodological unity: We doubt the issue is important.) Questions about "axiological" unity are more interesting but not directly to the present point either. The only sort of unity that matters here is ontological. Are special sciences, particularly cognitive and behavioral sciences, studying *one* domain of processes, relations, and objects (or what have you) that they *share* with other explanatory projects, including physics, or are they not? Reductionism is the historically most common and the conceptually most straightforward way of answering "yes" to this question. Kim's argument is supposed to convince us that other ways will not work. But we argue that reductionism would doom the explanatory significance of the special sciences in the very act of trying to unify them. Fortunately, we also argue reductionism is not nearly as well motivated, either by philosophical arguments or by the practice of science, including physics, as Kim thinks. We *can* have ontological unity without reduction, or so it is among our primary purposes in the target article to argue.

This logic is so fundamental to our case that we need some account of how an astute reader like **Boersema** could have missed it. He makes a revealing comment when he says that special sciences should not contradict physics "because we take physics to tell us about the basic components and constituents of the world." *That* we deny. One of the core arguments of our article is that *if* you think that physics identifies "basic components and constituents" of everything else, then Kim's case is valid, and the cognitive and behavioral sciences cannot be unified with others unless they are reduced. We need, and in the penultimate section of our article provide, an alternative account of why, and in which respects, special sciences are not allowed to run afoul of the generalizations of physics. We will come back to this below (in sects. R6 and R7) when we turn to remarks of other commentators connected more directly to issues from physics and their consequences. For now, note that it is just *because* Boersema apparently shares Kim's hunch that physics supplies generalizations about "the basic constituents" of everything that he also shares Kim's conviction that "good" explanations not cashed out in reductionist terms are not (ultimately) "correct." The point of our discussion in the article about scale-relative informational structures on a single topology is to provide a non-mysterious basis for denying this conviction.

Whereas **Boersema** wonders whether reductionists are committed to unity but then ultimately seems to endorse just the basis for unity that Kim does – and that ushers in all the trouble identified in our article – **Clarke** is overtly skeptical about our commitment to unity (a commitment he recognizes clearly). He offers two motivations for his skepticism. One is a view about what "realism" involves and about what justifies it, that differs from ours. We come back to this below. His second motivation is his belief that Cartwright may be right to promote disunity because it is consistent with what science – all of science together – tells us that the world may be, in Cartwright's phrase, "dappled."

This claim is directly on topic because it seems to address (and forthrightly deny) the kind of ontological unity that makes Kim's argument problematic for cognitive science. If

the world has "gaps" in it, then this may be the basis for providing a response to Kim that is different in kind from ours. We say "may" here because the gaps Cartwright imagines are not necessarily coextensional with the border-zones between disciplinary domains, which are the locus of importance in our attempt to help cognitive scientists locate their own domain on the wider ontological map. Cartwright's gaps are supposed to occur *within* each discipline. They are gaps across which, in some sense, reliable causal powers do not transmit influence. According to Cartwright, there are such gaps within the domains of physics, chemistry, and macroeconomics (the sciences she explicitly studies), and we also should expect to find them within specific cognitive and behavioral sciences (neuropsychology, ethology, etc.).

This thesis certainly denies unity, but in a way orthogonal to what potentially (and actually) perplexes cognitive scientists when they are confronted with reductionistic hunches like Kim's. "Dappledness" is a difficult philosophical idea. To be interesting, it must amount to more than the truism that we do not (and never will) have access to the full network of generalizations that would actually furnish explanations of all events and classes of events. It must be the claim that, as a matter of fact, there is no such overarching network of generalizations to be had. (The Lipton [2002] review of Cartwright cited by **Clarke** is a good source to consult for a reader new to this idea.)

Our article gives no arguments against dappledness, and this set of replies would not be an appropriate place to launch any (although see Spurrett 2001a). We will just note here that we think Cartwright's strong general conclusion well outruns her inductive evidence from the history of science. However, the version of scientific unity we defend does not require a claim that the world is uniform with respect to overarching universal empirical laws that describe all of its different regions (as partitioned along multiple dimensions). It requires only the hypothesis that where there are generalizations about mind and behavior to be had, these will be ontologically related to the generalizations of physics locally governing these regions by informational-constraint relations rather than by reductive identity relations. If Cartwright is so radical as to deny that there are any true generalizations at all, in any sense of "generalization" – a point on which we find her work to be unclear – then she may find our claim and Kim's to be equally uninteresting. However, there does not appear to be any direct disagreement between her and us. Perhaps the issue is merely semantic. We agree that reality has gaps in the sense of singularities. Talking as we are to scientists rather than logicians, we identify "the universe" with what scientists actually study, namely, the portion of reality on our side of the multidimensional boundary of singularities. It is then true by linguistic convention that there are no gaps in the sense of singularities in "the universe" as we define it.

Clarke does useful service in reminding us that we part ways with Philip Kitcher's post-1989 work. Unlike Kitcher, we are not attracted to the Kantian idea that the order we find in nature is projected by us rather than found. Of course, our use of Kitcher's earlier work does not require us to keep traveling with him in the direction of skepticism about ontological unity.

We conclude our discussion of unity by drawing attention to an admirably pithy point made by **Rodin.** "[T]he idea," he says, "that every material entity or process exists (or occurs) in the *same* physical space and time (or space time) sounds commonsensical." Indeed, it does. Although we are not general supporters of "common sense," we are happy to side with it when we see no need not to. This is as much unity as we need. Some scientists will be surprised to see that even this is too much for some philosophers, but we do not think these philosophers have succeeded in generating a burden of argument that a reasonable scientist with a commitment to explanation, and hence to some degree of unity, needs to try to carry.

R3. Realism

In our article we claim that viewing the world as structured into a single working machine is "crucial to any sort of realism worth having." **Clarke** maintains that what we call crucial is an "unwarranted presupposition," insisting that realism amounts only to the view that the world is mind independent.

We assume realism in our article but do not directly argue for it. The best argument for realism is the "no miracles" argument, to the effect that the explanatory and predictive successes of various sciences, including cases in which novel phenomena are predicted in advance of empirical testing, would be unacceptably mysterious if we did not hold that there was a real world independent of the content of our thoughts and theories. **Clarke** then is correct insofar as he maintains that a key aspect of realism is commitment to a mind-independent world. However, the "no miracles" argument for realism does not justify brute commitment to just any mind-independent world – the argument takes the successes of various sciences as a premise and leads to the conclusion that a mind-independent world *rather like what the successful sciences say it is like* exists. Therefore, we resist Clarke's suggestion that our commitment to unity is merely a presupposition, nor do we think it is a necessary truth. Our remarks in the preceding section say all that is appropriate here about the positive credentials of the "single working machine" view.

Wallace worries that we are betting our "metaphysical structure on the current state of fundamental physics, despite the fact that fundamental physics frequently changes," and says that he finds this "dangerous." He suggests that a "sufficiently abstract" characterization of patterns, immune to revisions in microphysics, would be preferable. We are not convinced that the danger is as great as Wallace seems to think and enthusiastically welcome such danger as remains.

On the first point, **Wallace** suggests that our approach has the consequence that revisions in fundamental physics will require revisions in the ontologies of all other sciences, raising the alarming prospect that cognitive scientists should be expected to master quantum mechanics to do their work. No such consequence necessarily follows. A pattern is real (sect. 3.2 of the target article) if it is projectible and information-theoretically efficient. A pattern may continue to satisfy *both* criteria even if our views about processes at different (including smaller) scales are revised. Wallace, but also **Shalizi** and **Collier**, gives some of the reasons why some macrostates are relatively insensitive to variations in microstates of the same systems. It is just this sort of stability that makes at least some macropatterns potentially independent of revisions in what we think is going on at smaller scales.

Turning to the second point, what we have just said does not amount to a defense of the view that once some pattern is decided to be real, it is permanently beyond risk of revision. Neither is the question of what patterns are real independent of fundamental physics. Recall that the two criteria for being a real pattern are framed in terms of "physically possible" perspectives. This means that physics does have a distinctive and ineliminable role to play in determining what is real. Our suggestions (sect. 4.4 of the target article) regarding viewing the world as a network of information channels are supposed to be at once physically responsible and sufficiently abstract not to be unstable in the face of just *any* changes in fundamental physics. In particular, despite **Wallace**'s worry, the proposal does not depend on a specific view about a "substrate" to be identified by fundamental physics. However, because we think it is fundamental physics that can tell us what sorts of information can get from one part of the network to another, what its connectedness at various scales is, what the distribution of singularities of various sorts bounding the network is, and so on, what may be regarded as a danger is to us a welcome openness to revision in the light of empirical discoveries, including revision in metaphysics itself. We do not consider sound metaphysics to be *a priori* inquiry.

This answer to **Wallace** also gives an answer to part of **Montero**'s commentary. Montero thinks that the requirement that good explanations must cite true explanans is too demanding and that when it is abandoned, a gap between "scientific explanation and how the world is" gets opened up, a gap that could be further pried open by a "savvy metaphysician such as Kim."

We note that **Montero**'s motivation (also part of **Wallace**'s reason for finding microphysics dangerous) for thinking that explanations do not cite true explanans is the "pessimistic meta-induction" to the effect that because the ontologies of previous scientific theories have been revised, we should expect the same of current theories. This argument is typically used as a weapon by antirealists and presents a challenge to the "no miracles" argument for realism glossed previously. If this argument works at all, it works against a vision of science as primarily concerned to determine the sorts of things there are in the world. Then the fact that scientists used to think there was phlogiston or caloric and now do not (and so forth) is evidence for the induction. However, the vision of science we defended has it that the main business of science is the identification of *structures.* So-called ontic structural realism (e.g., French & Ladyman 2003; Ladyman 2000) does justice to the no miracles argument and eludes the pessimistic meta-induction by confining realist commitments to structures that are preserved through changes to better theories. The "patterns" of Dennett, cited approvingly by Wallace, are, when genuinely "real," such structures.

This does not establish that there is no gap between how we think the world is and how it actually is. We take fallibilism – the admission that any of our current scientific views could be revised in the light of new discoveries – very seriously. Further, we can readily make sense of how new scientific work could confront us with such gaps – we find out that what we thought were two distinct processes are in fact one, we discover that some molecule that we thought did one thing in the brain does a different thing, and so on. Alternatively, as **Collier** explains in his commentary, we can determine that some explanations are in principle unavailable at certain scales of inquiry rather than others. However, we are perplexed by the suggestion that metaphysicians may have distinctive tools *over and above those available to scientists* for "prying" such gaps in any direction at all.

R4. Emergence and reduction

Collier, **Shalizi**, and **Wallace** refer to emergence in their commentaries. We accept and appreciate the points they make regarding the macrofeatures of various sorts of physical system but prefer to avoid using the term "emergence." Our primary reason for this is that the term has been used to refer to a variety of different putative phenomena, some of which we regard as both empirically disconfirmed and spooky.

In the late 19th and early 20th centuries, various emergentist proposals about the relationships between various sciences and physics were articulated. At least some of them explicitly involved commitment to the view that under certain conditions fundamental nonphysical causal powers could be brought into being. At the time many scientists were of the view that such nonphysical causal powers were necessary to account for a variety of phenomena, including chemical bonding, fermentation, and fetal development. That is, some sorts of emergentism clearly involved rejection of the completeness of physics. We take it that empirical work in a wide range of domains, including work on the conservation of known sorts of energy in living and nonliving systems, the laboratory synthesis of various organic molecules, and the quantum mechanical explanation of chemical bonding, has done more than enough to make clear that fundamental nonphysical causal powers are not required in a scientifically responsible picture of the world. Therefore, as made clear in the target article, we see no reason to entertain speculations to the effect that they are. Furthermore, **Collier**, **Shalizi**, and **Wallace** are manifestly not suggesting that we should – so what they mean by "emergence" is not this spooky view, even though use of the term can raise associations with such a view.

More recently, the term "emergent" has also been used to refer to features of various systems that exhibit this or that sort of supposedly unpredictable or otherwise dynamically interesting behavior, including the generation of relatively stable macrostates. (This includes, but is not restricted to, work on so-called emergent computation; e.g., Forrest 1991.) According to many, what is important about these systems is that some of their features cannot be reduced to others. **Collier** makes explicitly clear that what he means by emergence is a matter of a failure of reductive explanation in principle, and his example shares some important characteristics with those offered by **Shalizi** and **Wallace**. But Shalizi thinks, and says at the end of his commentary, that being a functionalist (whether in the behavioral sciences or statistical physics) about emergent features of physical systems *is* being a reductionist. What is going on here?

Clearly, there are at least two senses of "reduction" in play here. In fact, as **Collier** makes clear early in his commentary, there are *three* distinguishable, relevant senses of reduction. One of these is "intertheoretic" reduction, and as explained in the target article (sect. 2.2), this is the sense typically relevant to debates in the philosophy of mind. This

involves one whole theory being shown to be intertranslatable with another. A second involves reducing the "number of fundamental kinds of things" (e.g., by rejecting dualism in favor of materialism), and Collier suggests that this is better referred to as "*ontological deflation.*" One kind of ontological deflationism is physicalism – the view that everything that there is, is physical. Ontological deflation need not involve intertheoretic reduction. The third sort of reductionism, according to Collier, involves the elimination of objects, processes, or properties, as long as this can be accomplished "without any loss of explanatory power in principle." When this is not possible, and Collier and **Shalizi** provide complementary examples of macroscopic features of systems that cannot be eliminated without such loss, then we have a failure of the third sort of reductionism – what Collier and Shalizi effectively take as diagnostic of "emergence."

As we have said, we prefer avoiding talk of emergence and find it sufficient to describe ourselves as (up to the point justified by empirical science) nonreductionists. As made clear in the target article (sect. 1.1), one of the key weapons for the relevant sort of nonreductionism is the multiple realization argument for functionalism. This is an argument against *intertheoretic* reduction. **Shalizi**'s enthusiasm for functionalism and his defense of it by reference to multiple instantiation thus make him an *antireductionist* by our lights, even if an ontological deflationist. This means that the apparent disagreement over reductionism is in the first instance little more than an unfortunate consequence of the fact that, like "emergent," the word does multiple duty.

There is a point we think it apposite to add on the semantics of the word "reduction," that may be of real pragmatic import to cognitive and behavioral scientists when they are addressing the wider public. Our article should have made clear that we do not generally think that philosophers should feel authorized to tell scientists how to talk. However, philosophers draw a useful distinction we do not often find in the nonphilosophical literature between two senses of "realism." We think that this interacts with the multiple meanings of "reduction" discussed previously, in a way that makes cognitive scientists more likely to be tongue-tied in the face of metaphysical critiques like Kim's, knowing that *something* must be wrong with the argument but having trouble articulating what it is.

"Common-sense realism" is the view that the world includes roughly the kinds of objects, events, and processes that it pretheoretically appears to, and that one of the tasks of science is to explain the hidden structures and processes that lie behind this manifest reality. "Scientific realism" is the name for the view that manifest ("folk") ontologies frequently, perhaps usually, fail to partition nature in a way well suited to explanation, and that we should therefore expect such ontologies to be incrementally replaced by alternative schemes developed by the sciences. Common-sense realism comports naturally with ontological reduction because it expects science to discover the hidden microstructures with which the items in the manifest ontology are coreferential or identical. Kim's project is an exercise in common-sense realism, an effort to repair a surd spot in the integration of the folk concepts of mind and causation but without serious regard for what science shows. We think that most scientists are, in working practice, scientific realists just in the sense that they are prepared to junk folk ontologies whenever they find them interfering with explanatory progress. Scientific realists should *not* generally expect reductions (although they may occur here and there) because ontological displacement is incompatible with, is indeed the opposite of, ontological reduction.

If, as **Shalizi** says, scientists indulge the habit of referring to insistence on monism (i.e., ontological deflationism) as "reductionism," this must surely leave them less than ideally prepared to know how to respond when someone like Kim comes along and tells them that in the interests of ontological parsimony (in his case, of causes), they must reduce mental properties to lower-level ones. Of course, we do not argue that mental properties should be *either* reduced or displaced. We argue that many "higher level" *scientific* kinds – regardless of how the folk take them – are real despite being nonreducible, for the reasons discussed immediately above (i.e., they are "emergent" in **Collier**'s precise sense of that term). However, if scientists' usage helped them to better recognize that reduction is a longshot possibility in most domains and that any complex set of ontological structures is much more likely to either be elaborated and rendered more complex by science or else displaced by it, they would be more likely to see straight off that Kim and other conservative metaphysicians do not begin by sharing their view of the world and then go wrong somewhere or other that is hard to exactly find. The conservative metaphysician's picture of the world is, in a deep and important sense, *anti*scientific from its first assumptions.

R5. Physics

Several commentators – **Collier**, **Ladyman**, **Rodin**, **Shalizi**, and **Wallace** – have added new details and examples to our reflections on physics, which were intended to show that the kind of reductive base for good scientific kinds and properties imagined by the neoscholastic metaphysician does not exist. We of course welcome all this shoring up. Particularly gratifying are Wallace's comment that Kim's account "correctly handles *hardly any macroproperty at all,*" Collier's remark that "it is quite possible for an entity to be physical in every respect but not to be reducible in any way," Ladyman's point that "Kim, or anyone who similarly thinks that the real causal processes are only at the fundamental physical level, would then be faced with claiming that there are no true causes in space and time," and Shalizi's affirmations that although "the answers to R&S's questions about information, causation, functionalism, and emergence matter a great deal to cognitive science," "it is not just the special sciences that need functionalism: physics needs it, too, and uses it." To have this many experts telling us that we have got the way things are with physics right and that Kim and the neoscholastics have got it wrong leads us to think that however a neoscholastic may seek to answer us, he or she is going to have to concede our premise about physics. In that case, we cannot imagine how an argument like Kim's could possibly be airborne again.

Montero explicitly denies this last point, arguing that Kim's argument can be stated and answered without regard to any facts about physics. We will reserve our main comment on why we find this denial of hers to be implausible to our discussion below of issues associated with the topic of supervenience because this is the concept on which Montero depends to try to break the link between physics

and Kim's conclusion. However, we should point out here that Montero seems to misconstrue the way in which our discussion of physics is supposed to be relevant to our rejection of Kim's argument. She notes that causal exclusion threatens mental causes with redundancy based on neurophysiological causes – if the psychological supervenes on the neurophysiological – without any appeal to the level of physical causation required. This is correct. However, our point in discussing physics was *not,* as Montero seems to think, that Kim needs to find overdetermining causes in physics, where we then say they are not to be found. Rather, our point was that Kim's argument requires appeal to a level of "real" causation where functionalism does not apply. Surely, if there were *any* such domain, the level of the physical would have to be one of the places we would find it. It seems implausible that neurophysiological causation could be basic relative to psychological causation if physical causation is not. However, we argue, and as **Collier**, **Ladyman**, **Rodin**, **Shalizi**, and **Wallace** agree, physics does *not* provide a home for Kim's kind of causation. This strongly suggests there is no home for it at all at any of the levels to which Montero suggests attention.

R6. Inner states

Scheutz argues that a "serious challenge" for our proposal is the isolation of proper "inner states," because without such states the only warranted "causation talk" will be behaviorist. We accept the behaviorist conclusion but not the presumption that the challenge is serious. As we noted in the target article (sect. 1.1, note 3), functionalism, although historically a reaction to unduly restrictive behaviorism, can be seen, and we think should be seen, as itself a *form* of behaviorism.

Furthermore, the requirement that "cause" talk should pick out distinctive inner states (mental or otherwise) that are properly regarded as *the* causes of what happens is one that we are at pains to reject generally. It does no justice to the content of science. We argue (especially in sect. 4.4 of the target article) that it finds no home in the practice of physics, and some of the commentators, including **Ladyman**, **Shalizi**, and **Wallace**, give further argument and evidence in favor of this view. Wallace, in particular, emphasizes the ways in which even fundamental physical quantities, such as mass, are properly understood as dispositional: "something has mass *m* if it behaves thus-and-so on the scales." There is, one might object here, nothing particularly impressive about our commitment to behaviorism with respect to physics, given that no one seriously suggests that physical systems have any inner *mental* states.

As we made clear (sect. 3.1 of the target article), we are of the view that mental states are individuated extrinsically by triangulation under equilibrating pressures of various sorts. Recall our hunger example (sect. 3.3 of the target article). This individuation also involves identifying relations of interdependence between multiple factors typically of a variety of kinds. In the behavioral sciences, as in physics, causal claims *are* claims about such relations of complex interdependence (this claim is given fuller defense in Spurrett & Ross, under review). The causal claims which **Montero** asserts are made by neurophysiologists, are, we suggest, also of this form. To answer her demand that we say what counts as identifying a "genuinely causal" pattern, we reiterate (see sect. 4.3 of the target article) that when *any* science identifies real relations of functional interdependence that just is identifying genuine causes in the scientific sense. Given our arguments for distinguishing the scientific from a metaphysical concept of cause, an unqualified demand for a criterion for "genuine" causal properties seems to us to be begging the question.

Returning to **Scheutz**, and given these remarks about causes, we can distinguish two senses of "inner state," only one of them acceptable. On the one hand, a state may be "inner" relative to some functional economy, which is to say that it may be a subsystem with identifiable input and output relations that can, for some purposes at least, be treated as a black box. As committed defenders of multiple realization, it would be remarkable were we to deny that, and we do not. Alternatively, a state may be supposed to be "inner" in the sense of being radically unsuitable for extrinsic individuation. However, how could we be expected to convince ourselves that such states existed? We cannot detect anything that does not make a difference, and what we detect are the differences that are made. We are, that is, unabashed Dennettian behaviorists. Mental states such as beliefs that *p* are real (*if* they are real patterns; this is always an empirical matter). The relevant sort of pattern is a complex of attributed dispositions to be identified by Samuelsonian means: the construction of revealed preferences under specific scarcity conditions. Just as "fitness" is not a property specifiable independently of an ecosystem, and hence is relational rather than intrinsic, so it is with beliefs and other mental states.

R7. (Multiple) supervenience

Ladyman, **Macdonald**, **Marras**, and **Scheutz** question our appeal to Meyering's idea of "multiple supervenience." They doubt that we have done enough – or, indeed, anything – to make the notion plausible, and they furthermore suggest that we don't need it to make our argument against Kim. Ladyman and Macdonald suggest that merely denying that supervenience is generally local is sufficient for our purposes. That supervenience is sometimes or usually global does not imply that any relations of multiple supervenience obtain.

Macdonald and **Marras** argue that Meyering's putative examples of multiple supervenience, as cited in our article, show only that "when a categorical base supports different dispositions, *which* disposition is triggered in a particular case depends on the context, the initial conditions" (Macdonald). This, Macdonald goes on, implies only the "unremarkable conclusion" that "specific causes require specific contexts."

Because none of the critics of multiple supervenience think that their point promises to rescue Kim's argument, cognitive scientists will be right to think that here we have a truly in-house contestation among philosophers. We need to be responsible about not indulging this dispute too deeply in the pages of *BBS.* What we will do here is the following. We will discuss the philosophical issue by closely reviewing only the argument of **Marras** because he gives it the fullest and most rigorous airing. We will concede that his argument is valid and that it therefore forces a modification somewhere in our view. The modification we will offer is likely not the one Marras had in mind, but it will al-

low us to directly connect the arcane philosophical issue with the scientific ones that have occupied us elsewhere in these replies.

Marras reports having trouble understanding just what sort of relation we take multiple supervenience to be. We agree on reflection that our thought on this point was not as well-formed as it should have been. However, part of Marras' trouble stems from the fact that he does not suspect that we might be denying the mereological "stacking" of reality in terms of "levels" or "layers" *altogether.* (As will be discussed below, a similar thing also can be said with respect to the comments of, at least, **Boersema**, **Scheutz**, and **Shalizi**.) The possible interpretations of multiple supervenience Marras offers presuppose a "layer cake" world. Our own *positive* metaphysical theory, sketched briefly in the article but receiving forthcoming book-length treatment in Ross et al. (in preparation), is about denying this presupposition. We again emphasize that, as all the commentators discussed in this section agree, our argument against Kim does not depend on acceptance of our metaphysical theory. However, we think that some cognitive scientists may find it interesting.

In both Meyering's original treatment and our article, multiple supervenience is motivated by attention to the fact that different sciences cross-classify events, objects, and processes relative to one another. **Marras** adds welcome clarity here when he says that:

> We can have cross-classification *either* [1] when we can make distinctions in terms of the higher-level properties that we cannot make in terms of the base properties, *or* [2] when we can make distinctions in terms of the base properties that we cannot make in terms of the higher-level properties, *or* [3] both. Now it is clear that when we are dealing with higher-level *functional,* and, in particular, *mental* properties, it is the *second* of the above options that is the relevant one. . . .

This, he goes on, is just standard supervenience, whereas options (1) and (3) deny supervenience altogether.

Notice that this can all be expressed without invoking mereology, that is, without reference to "higher" and "lower" levels. It can be put in terms of the information-theoretical framework used in our article as follows. All relations between, for example, psychological and physical properties would respect standard supervenience if all information physically available in the enumeration of relations among some particular psychological properties were necessarily available (whether any actual measurement device could extract it) in the enumeration of relations among some physical properties. **Marras**' options (1) and (3) can be similarly reconstructed as the cases where this relation fails. In our article, we deny the generality of the relation and call the result "multiple supervenience"; Marras argues that this is not any kind of supervenience.

We think that **Marras**' argument for this last point is valid and that he, **Ladyman**, **Macdonald**, and **Scheutz** are therefore right that our use of the concept of "multiple supervenience" to express our view is inappropriate and misleading. What we should have done in the article, and will now do here, is deny the generality of supervenience, period.

This is probably the conclusion opposite to the conservative one **Marras** and the others hoped to encourage. **Scheutz** entertains the possibility that we may intend the radical conclusion and pronounces it "spooky." So it is bound to seem. The point of our positive metaphysical theory is to resolve what looks like a contradiction between denying supervenience, on one hand, and insisting on the primacy of physics, on the other.

By "the primacy of physics" we refer to the institutional fact that special sciences are not allowed to propose empirical relations or measurement values declared impossible by the physical generalizations currently accepted, whereas no symmetric restriction holds in the other direction. (Pointing out this asymmetry is one way to confirm that "multiple supervenience" is an inept description of the relations between physical facts and properties and biological or psychological ones because "multiple supervenience" implies symmetry.) Philosophers generally suppose that this relation must be given an interpretation in terms of "higher" and "lower" levels because they take physics to be describing the *constituents* of everything else. However, as **Collier**, **Ladyman**, **Rodin**, and **Shalizi** among the present commentators seem in *some* remarks to agree, this misdescribes what physical theory does and says. (Collier and Rodin appear to us to be fully consistent in this regard, whereas the others wobble on the point; more on this below. **Wallace**, who otherwise approves of what we say about physics, comments off-hand that, "Of course, there must be some sense in which macroscopic objects are built out of microscopic constituents, and in which they are indeed supervenient on the properties of the constituents." We think not. How things are with walls and bricks is *not,* we think, the model of how things are generally that is suggested by contemporary science.)

We take our naturalism seriously. It is, to be sure, an immensely powerful folk hunch that complex structures are made of "little things" and that processes decompose into the banging together of these little things. However, it is not science. In our view, the weight of evidence conferred on an hypothesis by the fact that it is a folk hunch, however entrenched and widespread, is zero. In the general area inhabited by reductionist intuitions, the fact that is inductively supported by the history of science is that ideas that required denial of the primacy of physics – astrology, creationism, vitalism, and 19th century emergentism about chemistry – all failed. But that is it. The progress of physical theory, or of science in general, has not consisted in a systematic or continuous decomposition of complex entities or processes into little things and their local interactions.

Our positive metaphysical theory aims to do justice to the primacy of physics without resort to mereological intuitions. According to it, no information can flow that does not flow physically. However, denial of mereology amounts to the claim that there is no one scale of measurement on the multidimensional topology that is the universe to be identified with "the" scale of ultimate physical flow. As **Shalizi** says, "the macroscopic variables that appear in physical theories are collective degrees of freedom"; there is no *particular* (small) scale associated with them. "Physics," as a whole, is a body of constraints giving us sets of conditions under which information *cannot* flow and, where it is quantitative, giving lower bounds on the amount of noise in particular channels that is ineliminable. However, physics does *not* tell the practitioners of special sciences what information in general *can* flow or *does* flow. As **Collier** says, if a biologist or psychologist or economist wants to know why a system has stabilized around one attractor rather than another, he or she has to do biology or psychology or economics. The relevant information *is not there* in the speci-

fication of the physical variables. **Marras**' standard supervenience condition, as reformulated in our nonmereological terms above, does not generally hold. As he has persuaded us, we should therefore say that the relations between physical facts and facts identified by special science are not generally supervenience relations, rather than saying they are relations of "multiple supervenience."

Is this "spooky"? It is counterintuitive, perhaps; but to the genuine naturalist that is not an objection. Our suggestion that all information that flows must flow on the surface of a single multidimensional topology at some scale or other – no flying above or tunneling beneath the surface – rules out thoughts that bend spoons, personality dispositions directly controlled by the positions of planets, and interventions by supernatural agents, *given* what science has empirically shown us so far about the shape of the topology. It does not rule out any of this spooky stuff *a priori,* something that is against our working naturalistic rules. Supervenience is a stronger claim than the principles that ban spooky phenomena, and it is stronger than what science licenses.

As **Macdonald**, **Marras**, and **Montero** note, coinstantiation is a weaker relation than supervenience, but it is strong enough to do useful work in our argument against Kim. We can recover coinstantiation in the terms of our metaphysical theory. Using Macdonald's example, if a signal carries the information that some *x* is blue, it automatically carries the information that that *x* is colored. However, because the information that *x* is blue is not the *same* information as the information that *x* is colored, we avoid what in our article (sect. 4.3) we call the "information-transmission exclusion problem." This is just the logical twin, in our information-theoretic framework, of Kim's causal exclusion problem. Thus, it is not surprising that when we reformulate coinstantiation in information-theoretic terms, we mirror the logic of Macdonald's suggested answer to Kim. Thus, specifically as against Kim, we, Marras, and Macdonald seem to be on the same page. Macdonald does not notice this, and so says that he "disagrees" with our diagnosis of where Kim goes wrong because he misunderstands the point of our disuniting the distinct concepts of causation. That is not, in itself, our answer to Kim; it is our basis for transforming the causal exclusion problem into the information-transmission exclusion problem, which we then invoke our metaphysical theory to dissolve. However, as just noted, Macdonald proposes a solution logically identical to ours within the framework of the folk picture of causation that our radical naturalism leads us to eschew. Therefore, we differ with Macdonald on the metaphysical frame but not on the logic of Kim's problem. However, we must note, with Kim and against **Montero**, that merely invoking coinstantiation of property instances does not pay for lunch unless one has an underlying metaphysical account that explains, in general, what distinguishes coinstantiation cases from cases in which one has discovered that a property is redundant and should be eliminated. Kim argues that metaphysics should provide an account of why supervenience holds where it does. In light of what we have conceded in the present section, we will not say that anymore. However, we will say that the metaphysician owes an account of why coinstantiation applies where it does. Our metaphysical theory offers such an account.

We will conclude this section by noting some qualms we have with **Shalizi**'s talk about "course-grained" descriptions that "emerge from" finer-grained lower levels. This is still the traditional picture, the one **Marras** takes for granted. Here is a further reason why we are uncomfortable with the word "emergence": it seems to suggest mereology. When **Collier**, whose remarks about physics are *entirely* compatible with our views, uses "emergence" in his innocuous (to us) sense (see above), the underlying metaphor that makes the word the right one for what Shalizi has in mind has died. However, because the metaphor is vigorously alive elsewhere, we think that Collier should reconsider his semantic preferences.

R8. Conclusion: Philosophy and science

A theme that has run throughout this series of replies, but is most explicitly articulated in section R4, concerns the tension between common-sense and scientific ontologies. Philosophers are far more likely than scientists to engage this tension self-consciously – and this is part of the basis of philosophy's relevance to science – but scientists must implicitly face it, too, when they conceptualize their goals to guide the design of their specific interventions in nature. Do they aim to consolidate our inherited image of the manifest world or replace it with a new, less anthropocentric and more objective one? Our target article is the response of committed scientific realists to a paradigmatic instance of a common-sense realist project. Those metaphysicians we have called "neoscholastics" are common-sense realists. They do not assist the progress of scientific development, and sometimes, inadvertently or deliberately, they threaten to retard it.

In light of this dialectic, it is perhaps useful to close by noting how our commentators fall along a spectrum between common-sense and scientific realism, a spectrum that is initially oriented by putting Kim at one extreme end and us at the other. **Collier** and **Rodin** stand right beside us, as does **Clarke**, despite his disagreement with us about the unity of science. **Ladyman**, **Wallace**, and **Shalizi** then line up in increasing order of distance from us but still on our side of the median point. Shalizi, the nonphilosopher in this group, is a particularly interesting case because, as we note in section R7, he combines our readiness to follow the conceptual revisions of science whither they lead with a willingness to invoke the metaphorical structures of classical intuition. We speculate that Shalizi is probably talking more conservatively than he intends to or would acknowledge a good reason to. If this speculation is correct, this may constitute a salutary instance of the potential relevance of philosophy to scientists.

Continuing our exercise, **Scheutz** comes next, standing perhaps around the median point. His language is that of science, but a number of the intuitions he uses to express it are those of common sense. The others – **Boersema**, **Macdonald**, **Marras**, and **Montero** – stand on the other side of the median, closer to, although in no case all the way out to, Kim. The commentaries do not provide enough evidence for us to try to sort them relative to one another. We should also note that Marras, doing the commentator's job with exemplary professionalism and focusing rigorously on our logic while keeping himself out of the picture, plays his cards especially close to his chest. However, as noted in section R7, his presumption of the intuitive world of levels seems clear.

Our target article is primarily addressed to cognitive sci-

entists; however, as we say in it, among its tasks is to try to urge philosophers over towards our end of the spectrum so they can participate less ambiguously in the project of explaining the world. Our wording here is deliberate. Although there are of course projects other than the scientific one, none of them successfully contributes to the explanation of the world. This attitude of ours is a form of "scientism" we think licensed by the track record of the scientific disciplines and institutions.

Yet is not "common sense" a good thing, too? Who should feel comfortable in deciding not to care about *that*? We will close with an anecdote about the circumstances that motivated us to write our target article. In late 2001, one of us attended the annual meetings of a major national philosophical society. At a seminar, a roomful of philosophers influenced by Kim and other traditional-style analytic metaphysicians unanimously agreed, in the course of discussing mental causation, that baseballs can't break windows. The reason is causal overdetermination: Some specific molecules of the baseball, it was said, interact with some specific molecules of the window. This local interaction is causally sufficient for everything that follows with respect to breaking. If the baseball as a whole is also a causal agent, we have too many causal agents.

We asked ourselves what a cognitive scientist might have made of this had he or she seen from the symposium title that mind was to be discussed and attended in hopes of learning something useful from philosophers. Embarrassed for our discipline in that hypothetical world, we decided to write the article.

We relate this anecdote as a way of showing how unstable a thing "common sense" can be. We prefer science.

ACKNOWLEDGMENTS

David Spurrett acknowledges the support of the National Research Foundation (South Africa) under grant number 2067110, and the Philosophy Department at the University of Alabama at Birmingham.

References

Letters "a" and "r" appearing before authors' initials refer to target article and response, respectively.

Achinstein, P. (1983) *The nature of explanation.* Oxford University Press. [aDR]

Armstrong, D. (1981) *The nature of mind and other essays.* Cornell University Press. [aDR]

Arnett, D. (1996) *Supernovae and nucleosynthesis: An investigation of the history of matter, from the big bang to the present.* Princeton University Press. [DW]

Badii, R. & Politi, A. (1997) *Complexity: Hierarchical structure and scaling in physics.* Cambridge University Press. [CRS]

Baker, L. R. (1993) Metaphysics and mental causation. In: *Mental causation,* ed. J. Heil & A. Mele. Clarendon Press. [aDR]

Batterman, R. (2000) Multiple realizability and universality. *British Journal for Philosophy of Science* 51:115–45. [aDR]

Beck, C. & Schlögl, F. (1993) *Thermodynamics of chaotic systems: An introduction.* Cambridge University Press. [CRS]

Bickle, J. (1998) *Psychoneural reduction: The new wave.* MIT Press/Bradford Books. [aDR]

Birks, J. B. (1963) *Rutherford at Manchester.* W. A. Benjamin. [aDR]

Block, N. (1980a) Introduction: What is functionalism? In: *Readings in the philosophy of psychology, vol. 1,* ed. N. Block. Methuen. [aDR]

(1980b) Troubles with functionalism? In: *Readings in the philosophy of psychology, vol. 1,* ed. N. Block. Methuen. [aDR]

(2003) Do causal powers drain away? *Philosophy and Phenomenological Research* 67:110–27. [MS]

Block, N. & Fodor, J. (1972) What psychological states are not. *Philosophical Review* 8(2):159–81. [aDR]

Bohm, D. (1952) A suggested interpretation of quantum theory in terms of "hidden" variables. *Physical Review* 85:166–93. [DW]

Brooks, D. R. & Wiley, E. O. (1988) *Evolution as entropy,* 2nd edition. University of Chicago Press. [JC]

Brooks, R. A. (1991) Intelligence without representation. *Artificial Intelligence* 47:141–60. [aDR]

Burge, T. (1993) Mind-body causation and explanatory practice. In: *Mental causation,* ed. J. Heil & A. Mele. Clarendon Press. [aDR]

Campbell, D. T. (1974) "Downward causation" in hierarchically organized biological systems. In: *Studies in the philosophy of biology,* ed. F. J. Ayala & T. Dobzhansky. Macmillan. [JC]

Cartwright, N. (1983) *How the laws of physics lie.* Clarendon Press. [DB, aDR]

(1989) *Nature's capacities and their measurement.* Clarendon Press. [aDR]

(1999) *The dappled world.* Cambridge University Press. [DB, SC, aDR]

Chaikin, P. M. & Lubensky, T. C. (1995) *Principles of condensed matter physics.* Cambridge University Press. [CRS]

Chalmers, D. (1996) *The conscious mind.* Oxford University Press. [aDR]

Chorin, A. J. (1994) *Vorticity and turbulence.* Springer. [CRS]

Churchland, P. (1981) Eliminative materialism and the propositional attitudes. *Journal of Philosophy* 78:67–90. [aDR]

Clapp, L. (2001) Disjunctive properties: Multiple realizations. *Journal of Philosophy* 98:111–36. [aDR]

Clark, A. (1997) *Being there.* MIT Press/Bradford Books. [aDR]

Collier, J. (1986) Entropy in evolution. *Biology and Philosophy* 1:5–24. http://www.nu.ac.za/undphil/collier/papers/entev.pdf [JC]

(1988) Supervenience and reduction in biological hierarchies. In: *Philosophy and Biology: Canadian Journal of Philosophy Supplementary,* ed. M. Matthen & B. Linsky. 14:209–34. http://www.nu.ac.za/undphil/collier/papers/redsup.pdf [JC]

(2002) What is autonomy? In: *Partial Proceedings of CASYS'01: Fifth International Conference on Computing Anticipatory Systems, International Journal of Computing Anticipatory Systems, Liège, Belgium,* ed. D. M. Dubois. CHAOS. http://www.nu.ac.za/undphil/collier/papers/What%20is%20Autonomy.pdf [JC]

Collier, J. & Hooker, C. A. (1999) Complexly organised dynamical systems. *Open Systems and Information Dynamics* 6:111–36. http://www.newcastle.edu.au/centre/casrg/publications/Cods.pdf [JC]

Collier, J. & Muller, S. (1998) The dynamical basis of emergence in natural hierarchies, with Scott Muller. In: *Emergence, complexity, hierarchy and organization: Selected and edited papers from the ECHO III Conference, Acta Polytechnica Scandinavica, MA91,* ed. G. Farre & T. Oksala. Finnish Academy of Technology. http://www.nu.ac.za/undphil/collier/papers/echoiii.pdf [JC]

Dennett, D. (1981) Three kinds of intentional psychology. In: *Reduction, time and reality,* ed. R. Healey. Cambridge University Press. (Reprinted in: Dennett, D. (1987) *The intentional stance.* MIT Press/Bradford Books). [aDR]

(1987) *The intentional stance.* MIT Press/Bradford Books. [aDR]

(1991a) *Consciousness explained.* Little, Brown. [aDR]

(1991b) Real patterns. *Journal of Philosophy* 88:27–51. [aDR, DW]

(1997) *Brainchildren: Essays on designing minds.* MIT Press. [CRS]

(2001a) Are we explaining consciousness yet? *Cognition* 79:221–37. [aDR]

(2001b) The Zombic hunch: Extinction of an intuition? In: *Philosophy at the new millennium,* ed. A. O'Hear. Cambridge University Press. [aDR]

Dupré, J. (1993) *The disorder of things.* Harvard University Press. [DB, aDR]

Elder, C. (2001) Mental causation versus physical causation: No contest. *Philosophy and Phenomenological Research* 62(1):111–27. [aDR]

Fodor, J. (1968) *Psychological explanation.* Random House. [aDR]

(1974) Special sciences, or the disunity of science as a working hypothesis. *Synthese* 28:77–115. [aDR]

(1975) *The language of thought.* Harvard University Press. [aDR]

(1987) *Psychosemantics.* MIT Press. [aDR]

(1994) *The elm and the expert.* MIT Press/Bradford. [aDR]

Forrest, S. ed. (1991) *Emergent computation.* MIT Press/Bradford Books. [rDR]

Forster, D. (1975) *Hydrodynamic fluctuations, broken symmetry, and correlation functions.* Benjamin Cummings. [CRS]

French, S. & Ladyman, J. (2003) The dissolution of objects. *Synthese* 136:73–7. [rDR]

Friedman, M. (1974) Explanation and scientific understanding. *Journal of Philosophy* 71:5–19. [aDR]

(1999) *Reconsidering logical positivism.* Cambridge University Press. [aDR]

Garfinkel, A. (1981) *Forms of explanation.* Yale University Press. [aDR]

Gintis, H. (2000) *Game theory evolving.* Princeton University Press. [aDR]

Glimcher, P. (2003) *Decisions, uncertainty, and the brain.* MIT Press/Bradford Books. [aDR]

Hempel, C. (1965) *The logic of scientific explanation.* Free Press. [aDR]

Horgan, T. (1997) Kim on mental causation and causal exclusion. *Philosophical Perspectives* 11:165–84. [aDR]
Hull, D. (1972) Reduction in genetics – biology or philosophy? *Philosophy of Science* 39:491–99. [aDR]
Humberstone, L. (1998) Note on supervenience and definability. *Notre Dame Journal of Formal Logic* 39:243–52. [MS]
Hutchins, E. (1995) *Cognition in the wild.* MIT Press/Bradford Books. [aDR]
Jackson, F. & Pettit, P. (1988) Functionalism and broad content. *Mind* 97:381–400. [aDR]
(1990) Program explanation: A general perspective. *Analysis* 50:107–17. [aDR]
Juarrero, A. (1999) *Dynamics in action.* MIT Press/Bradford Books. [aDR]
Kauffman, S. A. (1993) *The origins of order.* Oxford University Press. [JC]
Keizer, J. (1987) *Statistical thermodynamics of nonequilibrium processes.* Springer. [CRS]
Kim, J. (1993) *Supervenience and mind.* Cambridge University Press. [aDR]
(1998) *Mind in a physical world.* MIT Press/Bradford Books. [DB, JL, AM, aDR, MS]
Kincaid, H. (1997) *Individualism and the unity of science.* Rowman & Littlefield. [aDR]
Kitcher, P. (1976) Explanation, conjunction and unification. *Journal of Philosophy* 73:207–12. [aDR]
(1981) Explanatory unification. *Philosophy of Science* 48:507–31. [DB, aDR]
(1989) Explanatory unification and the causal structure of the world. In: *Scientific explanation,* ed. P. Kitcher & W. Salmon. University of Minnesota Press. [SC, aDR]
(1994) The unity of science and the unity of nature. In: *Kant's epistemology and philosophy of science,* ed. P. Parrini. Kluwer. [SC]
(1999) Unification as a regulative ideal. *Perspectives on Science* 7:337–48. [SC]
Kitcher, P. & Salmon, W., eds. (1989) *Scientific explanation.* University of Minnesota Press. [aDR]
Kripke, S. A. (1981) *Wittgenstein on rules and private language.* Blackwell. [MS]
Kuhn, T. S. (1971) Les Notions de causalité dans le developpement de la physique. *Etudes d'Épistémologie Génétique* 25:7–18. [aDR]
Kutz, M. (1998) Mathematical models of dynamical physical systems. In: M. Kutz, *The mechanical engineer's handbook,* Ch. 27. Wiley. [MS]
Ladyman, J. (2000) What's really wrong with constructive empiricism? Van Fraassen and the metaphysics of modality. *British Journal for the Philosophy of Science* 51:837–56. [rDR]
Leibniz, G. W. (1890) Anti barbarus physicus pro philosophia realis contra renovationes qualitatum scholasticarum et intelligentiarum chimaericarum. *Die philosophischen Schriften von Gottfried Wilhelm Leibniz,* ed. C. I. Gerhardt. Weidmannsche Buchhandlung. [AR]
Lewis, D. (1972) Psychophysical and theoretical identifications. *Australasian Journal of Philosophy* 50:249–58. [aDR]
(1980) Mad pain and Martian pain. In: *Readings in the philosophy of psychology, vol. 1,* ed. N. Block. Methuen. [aDR]
Lipton, P. (2002) The reach of the law. *Philosophical Books* 43:254–60. [SC, rDR]
Loewer, B. (2001) Review of Kim: *Mind in a physical world. Journal of Philosophy* 98:315–24. [aDR]
Macdonald, C. (1989) *Mind-body identity theories.* Routledge. [GM]
Macdonald, C. & Macdonald, G. (1986) Mental causation and explanation of action. *The Philosophical Quarterly* 36:145–58. [GM]
(1995) How to be psychologically relevant. In: *Philosophy of psychology: debates on psychological explanation, vol. 1,* ed. C. Macdonald & G. Macdonald. Blackwell. [GM]
Macdonald, G. (1992) Reduction and evolutionary biology. In: *Reduction, explanation, and realism,* ed. K. Lennon & D. Charles, pp. 69–96. Oxford University Press. [GM]
Marcus, E. (2001) Mental causation: Unnaturalized but not unnatural. *Philosophy and Phenomenological Research* 63(1):57–83. [aDR]
Marras, A. (2000) Critical notice of Kim: *Mind in a physical world. Canadian Journal of Philosophy* 30:137–60. [aDR]
(2002) Kim on reduction. *Erkenntnis* **57**(2):231–57. [aDR]
McClamrock, R. (1995) *Existential cognition.* University of Chicago Press. [aDR]
McGinn, C. (1991) *The problem of consciousness.* Blackwell. [aDR]
Melnyk, A. (2003) *A physicalist manifesto: Thoroughly modern materialism.* Cambridge University Press. [BM]
Menzies, P. (1988) Against causal reductionism. *Mind* 98:551–74. [aDR]
Meyering, T. (2000) Physicalism and downward causation in psychology and the special sciences. *Inquiry* 43:181–202. [AM, aDR]
Millero, F. J. (2001) *The physical chemistry of natural waters.* Wiley-Interscience. [aDR]
Millikan, R. (1999) Historical kinds and the special sciences. *Philosophical Studies* 95:45–65. [GM]
Nagel, E. (1961) *The structure of science.* Harcourt, Brace & World. [aDR]
Needham, P. (2002) The discovery that water is H_2O. *International Studies in the Philosophy of Science* 16(3):205–26. [aDR]
Nottale, L. (1993) *Fractal space-time and microphysics: Towards a theory of scale-relativity.* World Scientific. [aDR]
(2000) Scale relativity, fractal space-time and morphogenesis of structures. In: *Sciences of the interface: Proceedings of International Symposium in Honor of O. Rössler,* ed. H. Diebner, T. Druckrey & P. Weibel. Genista. [aDR]
Oppenheim, P. & Putnam, H. (1958) Unity of science as a working hypothesis. In: *Minnesota studies in the philosophy of science, vol. 2,* ed. H. Feigl, G. Maxwell & M. Scriven. University of Minnesota Press. [aDR]
Papineau, D. (1993) *Philosophical naturalism.* Blackwell. [aDR]
Pearl, J. (2000) *Causality: Models, reasoning and inference.* Cambridge University Press. [CRS]
Pettit, P. (1993) *The common mind.* Oxford University Press. [aDR]
Place, U. T. (1956) Is consciousness a brain process? *British Journal of Psychology* 47:44–50. [aDR]
Ponce, V. (2003) *Rethinking natural kinds.* Doctoral dissertation in philosophy, Duke University. [aDR]
Putnam, H. (1963) Brains and behavior. In: *Analytical philosophy, second series,* ed. R. Butler. Basil Blackwell & Mott. [aDR]
(1967a) Psychological predicates. In: *Art, mind and religion,* ed. W. H. Captain & D. D. Merrill. University of Pittsburgh Press. [aDR]
(1967b) The mental life of some machines. In: *Intentionality, minds and perception,* ed. H.-N. Castañeda. Wayne State University Press. [aDR]
(1975a) *Mind, language and reality: Philosophical papers, vol. 2.* Cambridge University Press. [aDR]
(1975b) Philosophy and our mental life. In: *Mind, language and reality: Philosophical papers, vol. 2.* Cambridge University Press. [aDR]
Pylyshyn, Z. W. (1984) *Computation and cognition: Towards a foundation for cognitive science.* MIT Press. [aDR]
Raynor, H. A. & Epstein, L. H. (2001) Dietary variety, energy regulation, and obesity. *Psychological Bulletin* 127(3):325–41. [aDR]
Redhead, M. (1990) Explanation. In: *Explanation and its limits,* ed. D. Knowles. Cambridge University Press. [aDR]
Reichenbach, H. (1957) *The philosophy of space and time.* Dover. [aDR]
Ross, D. (1991) Hume, resemblance and the foundations of psychology. *History of Philosophy Quarterly* 8:343–456. [aDR]
(1997) Critical notice of Ron McClamrock: *Existential cognition. Canadian Journal of Philosophy* 27:271–84. [aDR]
(2000) Rainforest realism: A Dennettian theory of existence. In: *Dennett's philosophy: A comprehensive assessment,* ed. D. Ross, A. Brook & D. Thompson. MIT Press. [aDR]
(2001) Dennettian behavioural explanations and the roles of the social sciences. In: *Daniel Dennett,* ed. A. Brook & D. Ross. Cambridge University Press. [aDR]
(forthcoming) Chalmers's naturalistic dualism: A case study in the irrelevance of the mind-body problem to the scientific study of consciousness. In: *The mind as scientific object,* ed. C. Ernelling & D. Johnson. Oxford University Press. [aDR]
Ross, D., Ladyman, J., Spurrett, D. & Collier, J. (in preparation) *What's wrong with things.* Oxford University Press. [rDR]
Rowlands, M. (1999) *The body in mind.* Cambridge University Press. [aDR]
Ruelle, D. (1978) *Thermodynamic formalism: The mathematical structures of classical equilibrium statistical mechanics.* Addison-Wesley. [CRS]
Russell, B. (1917) On the notion of cause. In: *Mysticism and logic.* Allen & Unwin. [aDR]
Salmon, W. C. (1984) *Scientific explanation and the causal structure of the world.* Princeton University Press. [DB, aDR, CRS]
(1990) Scientific explanation: Causation *and* unification. *Critica Revista Hispanoamericana de Filosofia* 22:3–21. [aDR]
(1999) *Causality and explanation.* Oxford University Press. [aDR]
Scheutz, M. (1999a) *The missing link: implementation and realization of computations in computer and cognitive science.* Doctoral dissertation, Departments of Cognitive and Computer Science, Indiana University, Bloomington, IN. [MS]
(1999b) When physical systems realize functions. *Minds and Machines* 9:161–96. [MS]
(2001) Causal versus computational complexity. *Minds and Machines* 11:534–66. [MS]
Shadlen, M., Britten, K., Newsome, W. & Movshen, J. (1996) A computational analysis of the relationship between neuronal and behavioural responses to visual motion. *Journal of Neuroscience* 16:1486–510. [aDR]
Shalizi, C. R. & Moore, C. (2003) What is a macrostate? Subjective measurements and objective dynamics. Available at: http://arxiv.org/abs/cond-mat/0303625. (Also submitted to *Studies in the History and Philosophy of Modern Physics*). [CRS]
Shannon, C. & Weaver, W. (1949) *The mathematical theory of communication.* University of Illinois Press. [DB]
Smart, J. J. C. (1959) Sensations and brain processes. *Philosophical Review* 68:141–56. [aDR]

Spurrett, D. (1999) *The completeness of physics.* Doctoral dissertation in Philosophy, University of Natal, Durban, South Africa. [SC, aDR]
(2001a) Cartwright on laws and composition. *International Studies in the Philosophy of Science* 15(3):253–68. [SC, arDR]
(2001b) What physical properties are. *Pacific Philosophical Quarterly* 82(2):201–25. [aDR]
Spurrett, D. & Papineau, D. (1999) A note on the completeness of "physics." *Analysis* 59(1):25–29. [aDR]
Spurrett, D. & Ross, D. (under review) On notions of cause: Russell's thesis after ninety years. *British Journal for the Philosophy of Science.* [rDR]
Stich, S. (1983) *From folk psychology to cognitive science.* MIT Press/Bradford Books. [aDR]
Thalos, M. (2002) The reduction of causal processes. *Synthese* 131:99–128. [aDR]
van Brakel, J. (2000) The nature of chemical substances. In: *Of minds and molecules: New philosophical perspectives on chemistry,* ed. N. Bhushan & S. Rosenfeld. Oxford University Press. [aDR]
van Fraassen, B. (1980) *The scientific image.* Clarendon Press. [aDR]
Van Gulick, R. (1993) Who's in charge here? And who's doing all the work? In: *Mental causation,* ed. J. Heil & A. Mele. Clarendon Press. [aDR]
Vernadsky, V. (1988) Prostranstvo i vremja v zhivoj i nezhivoi prirode. *Philosophskije misli naturalista.* Nauka. [AR]
Wallace, D. (2003) Everett and structure. *Studies in History and Philosophy of Science B: Studies in the History and Philosophy of Modern Physics* 34:87–105. [aDR, DW]
Wilson, R. (1995) *Cartesian psychology and physical minds.* Cambridge University Press. [aDR]
Yablo, S. (1992) Mental causation. *Philosophical Review* 101(2):245–80. [aDR]