Epiphenomenalism – the Do's and the Don'ts

Larry Shapiro and Elliott Sober
University of Wisconsin, Madison

***Abstract*:** When philosophers defend epiphenomenalist doctrines, they often do so by way of *a priori* arguments. Here we suggest an empirical approach that is modeled on August Weismann's experimental arguments against the inheritance of acquired characters. This conception of how epiphenomenalism ought to be developed helps clarify some mistakes in two recent epiphenomenalist positions – Jaegwon Kim's (1993) arguments against mental causation, and the arguments developed by Walsh (2000), Walsh, Lewens, and Ariew (2002), and Matthen and Ariew (2002) that natural selection and drift are not causes of evolution. A manipulationist account of causation (Woodward 2003) leads naturally to an account of how macro- and micro-causation are related and to an understanding of how epiphenomenalism at different levels of organization should be understood.

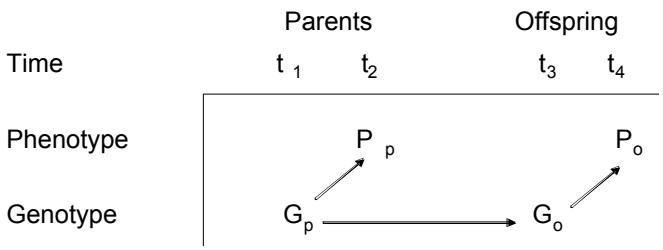
1. The Weismann Model

August Weismann (1889) is widely credited with disproving the Lamarckian theory of the inheritance of acquired characteristics. In one of his famous experiments, Weismann cut off the tails of newborn mice; when the mice grew up and reproduced, their offspring had tails as long as their parents' had prior to surgery. These results remained constant over many generations. Weismann saw the same pattern, and the same evidence against the inheritance of acquired characteristics, in the fact that circumcision over many centuries had not caused boys to be born without foreskins. He also thought that his theory of the continuity of the germ plasm threw further doubt on the inheritance of acquired characteristics, though it is worth asking whether this theory was *evidence* against Lamarckianism (as Gould 2002 argues) or merely *assumed* that Lamarckian inheritance does not occur.

Our interest here is in the logic of Weismann's experiments. Regardless of their general significance for "Lamarckianism," they clearly provided evidence that acquired tailessness in mice parents failed to cause tailessness in mice offspring.[1] How did the experiments manage to perform that function?

[1] Lamarckians argued that Weismann's result concerning mutilations has no bearing on characteristics that are useful to the organism.

Figure 1: The usual representation of Weismann's model of the relationship among parental and offspring genotypes and phenotyp es.



Weismann's theory of the continuity of the germ plasm is now conventionally expressed in the language of genotypes and phenotypes (Figure 1).[2] Parental genes do two jobs. They causally contribute (along with the environment) to parental phenotypes and they are passed along to offspring in reproduction. The Weismannian point is that there is no causal arrow from parental phenotypes to offspring genotypes. The blacksmith whose exertions lead him to acquire big muscles does not thereby alter the genes he gives to his children so that they develop big muscles without needing to exercise. Weismann was not concerned with the transmission of phenotypes by way of teaching and learning and other forms of cultural influence; it is no objection to his position that human children inherit money and accents from their parents. A fuller representation of the Weismannian picture is given in Figure 2.

[2] Although "genotype" and "phenotype" were introduced after Weismann, his distinction between "germ plasm" and "soma" is close enough to allow Figure 1 to be a natural representation of his ideas.

Figure 2: Weismann's theory of the continuity of the germ plasm does not rule out teaching/learning and maternal influence. It does rule out P_p influencing G_o.

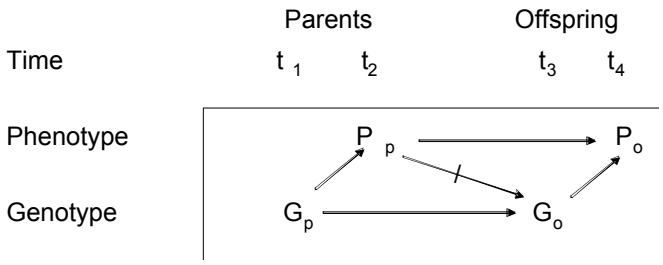
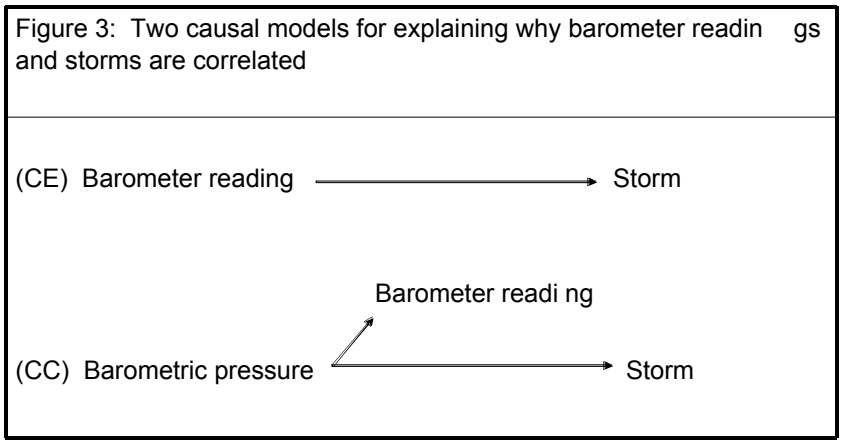


Figure 1 will remind philosophers of epiphenomenalism – the view that mental property instantiations[3] are effects of physical processes but are themselves causally inert. According to epiphenomenalism, having belief X, or desire Y, or feeling Z, are effects, not causes. However, when Figure 1 is fleshed out further to yield Figure 2, we see that Weismann was not arguing that parental phenotypes are causally inert; he wasn't claiming that they have no effects of any kind. Rather, he was advancing the more modest thesis that they have no effect on offspring genotypes. Parental phenotypes are epiphenomenal *with respect to the process of genetic inheritance.* In just the same way, when Salmon (1984, pp. 141-147) argued that a circle of light moving across the ceiling of the Astrodome (a sequence of events that occurs because a spotlight aimed at the ceiling is rotated) is a pseudo-process (Salmon's term), he was not claiming that the moving circle of light has no effects. Rather, he was arguing that the circle of light on the ceiling at one time does not causally influence the shape or color of the circle that appears a moment later. This is the first broad lesson we draw concerning how epiphenomenalism should be formulated. The claim should be that some set of property instantiations fails to affect some other set, not that the first set has no effects at all.

In his experiment, Weismann was using a strategy that is widespread in science. We observe that offspring phenotypically resemble their parents. Is this due to parental phenotypes causing offspring phenotypes, or is the resemblance a "mere correlation"? The latter would be true if the phenotypes of parents and offspring were joint effects of a common cause. An old chestnut from the philosophy of science nicely illustrates the

[3] It is not the property itself (an abstract object that exists outside of space and time) that constitutes a cause, but one or more objects tokening one or more properties; it is property *instantiations* that epiphenomenalists and their critics are discussing. We resist describing the problem in terms of mental events, since this inevitably leads to a problem we feel plays no essential role in questions about epiphenomenalism – namely, the problem of how events should be individuated.

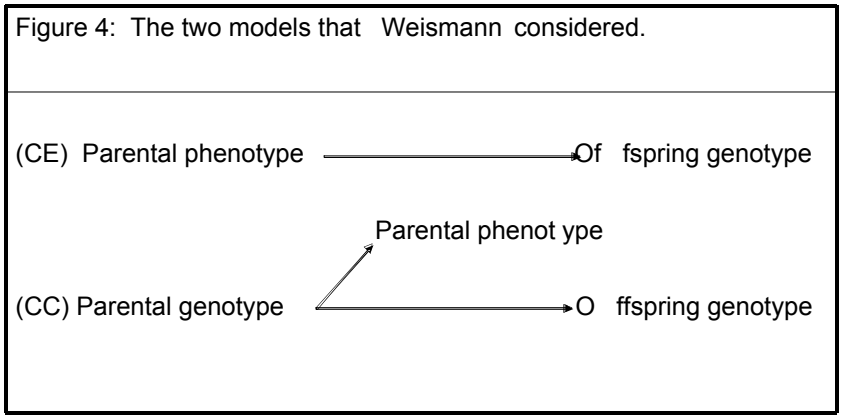
difference. Barometer readings are correlated with storms. Is this because barometer readings and storms are related as cause and effect (the CE hypothesis) or is it because the readings and the storms trace back to a common cause, namely the barometric pressure (the CC hypothesis)? Figure 3 depicts these two hypotheses. An obvious way to decide which of these models is better is to *intervene* on the barometer reading and see if changes in the state of the barometer will be associated with changes in the storm.[4] If the CE model is true, the association should be observed; if the CC model is correct, it should not. Making this intuitive idea precise requires careful specification of the notion of an intervention (Woodward 2003, 2005). One element in the required clarification is that "interventions" can't be ham-fisted; when you change the barometer reading, you shouldn't simultaneously change the barometric pressure.



Weismann was using this strategy when he intervened on the parents' phenotypeThe question was whether the intervention would be associated with a change in the offspring's genotype. The two models that Weismann compared are given in Figure 4. Despite the structural similarity that unites the barometer problem in Figure 3 with Weismann's problem in Figure 4, there is an obvious difference. We can tell by observation whether changing the barometer reading is associated with a change in whether there is a storm. But Weismann had no way of observing whether cutting off the parents' tails was associated with a change in the offspring's genes. Genes at the time were not observable. Weismann solved this problem by assuming that genetic differences must, in his experiment, manifest themselves phenotypically. More precisely, he assumed that if two organisms are reared in identical environments, they will differ phenotypically if and only if they differ genetically. This assumption – which involves

[4] We do not claim that intervention is the only way to gather data relevant to determining causal relationships. Purely observational studies can do so as well. What we do claim is that X causes Y if and only if a suitably defined intervention on X would be associated with some change in Y (or in the probability of Y).

defining "same gene" functionally in terms of the phenotypes produced – allowed him to look at the offspring's phenotype as a reliable indicator of the underlying genotype.



This functional definition of the gene was standard in early genetics but has long since been superseded. Thinking of genes as sequences of DNA and recognizing that different sequences can have identical phenotypic upshots is now routine. Weismann's experiment was not bullet-proof, at least by modern standards. Nevertheless, the logic of his experiments was sound.

The basic idea behind Weismann's experiment was this: to find out whether X causally contributes to Y, you manipulate the state of X while holding fixed the state of any common cause C that affects both X and Y; you then see whether a change in the state of Y occurs. We find it useful to think of this procedure probabilistically. The question is whether the following inequality is true:

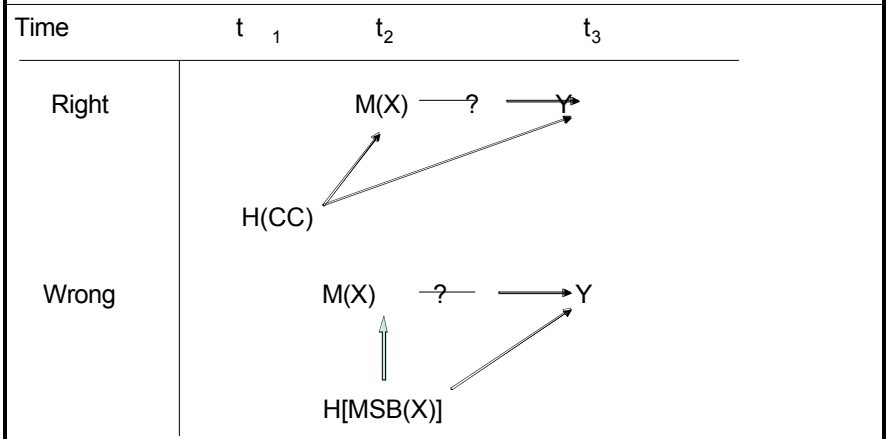
For any genotype G_p that a parental pair might have,
Pr(offspring has a tail_ parents have G_p & parents have tails) >
Pr(offspring has a tail _parents have G_p & parents' tails were cut off when they were born).

Notice that the conditional probabilities compared here both have "parents have genotype G_p" to the right of the conditional probability sign, but one has "parents have tails" on the right while the other has "parents' tails were cut off when they were born." This is what we mean by varying the parents' phenotype while holding fixed their genotype. Weismann's experiments indicated that this inequality is false; the two conditional probabilities are equal.[5]

[5] Here we see a point of agreement between interventionist theories of causation like Woodward's (2003) and probabilistic theories like Eells' (1991).

We have gone into some detail about the logic of Weismann's experiment because we think it is important for philosophers to see clearly what Weismann did *not* do. As we have explained, Weismann manipulated the parental phenotype while holding fixed the parental genotype. He did not manipulate the parental phenotype while holding fixed the micro-supervenience base of that phenotype. What's the difference between these two procedures? We assume that cause precedes effect; causation is a diachronic relation. In contrast, the (mereological) supervenience relationship is synchronic; it relates the macro-state of the organism at time t to its micro-state at time t. The idea is that the micro-state at t determines the macro-state at t; however, the converse is not true, owing to the fact that macro-properties are multiply realizable at the micro-level. The most important lesson we draw from Weismann is that *investigating whether X causes Y involves figuring out whether wiggling X while holding fixed whatever common causes there may be of X and Y will be associated with a change in Y. It is not relevant, or even coherent, to ask what will happen if one wiggles X while holding fixed the micro-supervenience base of X.* Figure 5 depicts the right and the wrong ways to investigate the causal question. The arrows connecting variables at different times represent causality; the arrow connecting simultaneous variables represents the supervenience relation.

Figure 5: The right and the wrong manipulations for determining whether X causally influences Y. M(x) means that x is manipulated by an intervention, H(z) means that z is held fixed, and MSB(X) is the micro - supervenience base of X.



What's wrong about the wrong way to test whether X causally influences Y? Because a supervenience base for X provides a sufficient condition for X, where the entailment has at least the force of nomological necessity, asking this question leads one to attempt to ponder the imponderable – would Y occur if a sufficient condition for X

occurred but X did not?[6] The question makes no more sense when one asks whether the following probabilistic inequality is true:

Pr[Y_X & MSB(X)] > Pr[Y_notX & MSB(X)].

The conditional probability on the right-hand side has an impossible proposition to the right of the conditional probability sign. The standard Kolmogorov definition of conditional probability entails that the right-hand probability is not defined. There is no saying how probable Y would be if a nomologically sufficient condition for X obtained and yet X did not (Eells 1991; Sober 1999b).

We conjecture that the mistake we just described underlies the main intuition that makes epiphenomenalist doctrines attractive. Consider the question of mental causation. How could believing or wanting or feeling cause behavior? Given that any instance of a mental property X has a physical micro-supervenience base MSB(X), it would appear that X has no causal powers in *addition* to those that MSB(X) already possesses. The absence of these additional causal powers is then taken to show that the mental property X is causally inert. We call this *the master argument for epiphenomenalism*, not because there are no other arguments in the epiphenomenalist's armamentarium, but because this is the argument that makes the view attractive in the first place, or so we suspect. The crucial mistake in this line of reasoning is that it requires one to consider a counterfactual situation that is in fact impossible and is in any case irrelevant to the question of whether the mental property X, or any other supervening property, is epiphenomenal with respect to an effect term Y. To see if X has an effect on Y that is additional to whatever effect MSB(X) has on Y, one would have to compare what would happen to Y if both MSB(X) and X were present with what would happen to Y if MSB(X) were present and X were absent. The master argument purports to evaluate this counterfactual and then concludes that the mental property X makes no contribution to Y additional to the effect that MSB(X) has. The conclusion is then drawn that mental properties are causally inert.[7] The principal fallacy is the thought that if X causes Y, then X must have an impact on Y *additional* to the impact on Y that MSB(X) has.

In summary, well-conceived arguments for epiphenomenalism should be both limited and have a certain empirical specificity. They should aim to show that one class of properties does not affect a second class, not that the first has no effects at all. And general theses about supervenience and the nature of causation do not settle whether manipulating X while holding fixed the common causes of X and Y would be associated with a change in Y. Weisman did not rely on general metaphysical considerations to

[6] Those who find this question intelligible may want to address the problem of what would happen if an immoveable object encountered an irresistible force. Notice that it does no good to count counterfactuals true whenever they have impossible antecedents; that leads the master argument to the unsatisfactory conclusion that "X causes Y" and "X does not cause Y" are both true.

[7] Graham (1998) discusses what he terms the master argument for consciousness epiphenomenalism, but his master argument and ours are different. His says: (1) Every physical event has a physical explanation. (2)No event has more than one explanation. (3)Physical causal explanations never invoke consciousness as a cause. (4) Therefore, consciousness is causally inert. Our master argument is not committed to any of the premises in Graham's, and the conclusion of our version isn't just about consciousness.

establish that cutting off the tails of mice does not cause their offspring to be born without tails; what Weisman *did* do should serve as our model for what good arguments for epiphenomenalism should be like. The lesson we draw from the master argument is that one must be very careful in deploying "holding-fixed arguments." To assess whether X causes Y, the common causes of X and Y must be held fixed, but not the micro-supervenience base of X.[8] We now turn to two recent contexts in which philosophers have felt themselves driven to epiphenomenalist conclusions.

2. Epiphenomenalism and Mental Causation

Epiphenomenalism has been a central focus in Jaegwon Kim's work over the past few decades (1984, 1993, 1998). More than perhaps any philosopher, Kim has wrestled with and sometimes defended versions of the master argument that we described in the previous section. Viewing Kim's ideas from the perspective of the empirical framework we developed in Part I will help shed further light on the topic of epiphenomenalism, especially in the context of philosophy of mind.

We begin our investigation of Kim's work with discussion of a position he articulated in 1984:

> Mental causation does take place; it is only that it is epiphenomenal causation, that is, a causal relation that is reducible to, or explainable by, the causal processes taking place at a more basic level. And this, according to the present account, is also precisely what happens with macrophysical causal relations. *Epiphenomenal causal relations involving psychological events, therefore, are no less real or substantial than those involving macrophysical events. They are both supervenient causal relations*. It seems to me that this is sufficient to redeem the causal powers we ordinarily attribute to mental events (Kim 1984, p. 107, his italics).

Although Kim says here that the causal relations that hold between instances of macroproperties are "epiphenomenal," this is not meant as a slight. He doesn't mean that they are epiphenomenal in our sense; he is not saying they are causally inert. Kim here appears to be immune to the siren song of the master argument we sketched in Part I, and he also appears to accept that there is something genuine about causal relations between instances of macroproperties. These causal relations can be reduced to more basic ones, but this reduction does not call into question their reality. As Kim insists, macrocausation *does take place*.

---

[8]Another thing that should not be held fixed in evaluating whether X causes Y are causal intermediaries (i.e., any event I such that X causes I and I causes Y). Otherwise, it will be possible to "prove" that effects have proximate but not distal causes. Matters change when the question is not whether X causes Y, but whether X influences Y along one causal pathway (through intermediary I_1) when it is known that X influences Y along a second pathway (through I_2); here one holds fixed I_2, varies X, and sees whether a change in Y occurs; see Hitchcock (2001) for discussion.

Crucial to Kim's sanguinity is a distinction he draws between two kinds of epiphenomenal causal relations. The first, which we will call a *purely epiphenomenal causal relation*, is exemplified by Salmon's circle of light moving across the Astrodome ceiling and the relationship between parent phenotype and offspring genotype (see Figure 2). Kim illustrates the idea of a purely epiphenomenal causal relationship by describing Jonathan Edwards' example of successive images in a mirror. An image in the mirror at one time is purely epiphenomenal with respect to the image that appears there an instant later. Indeed, one might wonder why Kim thinks "there *is* a causal relation between the two" (1984, p. 94). Isn't "epiphenomenal causation" a contradiction in terms? Kim simply means that the events linked by this relation are effects of a common cause. Like the barometer reading and the storm, the one mirror image does not cause the other.

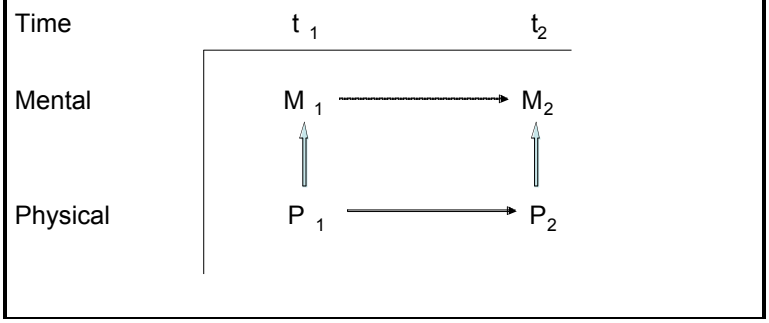
Contrasting with purely epiphenomenal causal relations are what Kim calls *supervenient causal relations*. Figure 6 depicts a supervenient causal relation between two mental states. The vertical arrows represent the supervenience relation that holds between base physical properties and their supervening mental properties. The horizontal arrow between the two physical properties represents causation. The horizontal arrow between the mental properties depicts Kim's supervenient causal relation. It is dashed to indicate that M_1 causes M_2 only because P_1 causes P_2. Supervenient causation, unlike purely epiphenomenal causation, is real causation, but it is a causal relation that depends on causation at a more micro-level. As Kim says,

> in both [kinds of causation] there is an *apparent* causal relation that is explained, or explained away, at a more fundamental level. The difference between the two cases is this: macrocausal relations are *supervenient causal relations* – supervenient upon microcausal relations – whereas cases like Edward's mirror images are not. This can be seen by reflecting on the fact that in a perfectly straightforward sense, mirror images, symptoms of a disease, and so on are causal effects of the underlying process – they are not mereologically supervenient on those processes (1984, p. 103).

This last remark of Kim's is important – if M_1 (purely) epiphenomenally causes M_2, the mereological supervenience base of the former does not cause the mereological supervenience base of the latter, whereas if M_1 superveniently causes M_2, it does.[9] Kim goes on to say that supervenient causal relations are among those that are "real" (though he places the word in scare quotes), in contrast to purely epiphenomenal causal relations, which are not.

[9] We will assume that Kim's concept of supervenient causation is defined by talking about the mereological supervenience bases of M_1 and M_2, not just about any old supervenience bases of those properties. The looser definition allows problems to intrude that are ruled out by the stricter definition.

Figure 6: Kim's relation of supervenient causation connects the mental properties M_1 and M_2 because their mereological supervenience bases, P_1 and P_2, are related as cause to effect.



Kim's distinction is neutral on whether mental properties are causes of behavior. Further argument would be required to show whether the relationship of mental properties to behavior is purely epiphenomenal or an instance of supervenient causation. However, matters change in 1993, when Kim defends an idea he terms "the causal inheritance principle," which states that "if M is instantiated on a given occasion by being realized by P, then the causal powers of *this instance of* M are identical with (perhaps a subset of)[10] the causal powers of P" (p. 355, his italics). Kim thinks this principle has consequences that are "devastating" to nonreductive physicalism. Because M's causal powers are nothing beyond those of its realizer P, M contributes no causal oomph in addition to whatever P contributes. This means, Kim believes, that all the causal work is done by P, and so it is P but not M that figures in causal relations.

We see the master argument lurking behind the scenes here. Consider this claim of Kim's, altered to make it fit the variables we use in Figure 6:

> P_1 appears to have at least as strong a claim as M_1 as a direct cause of P_2 (that is, without M_1 as an intervening link). Is there any reason for invoking M_1 as a cause of P_2 at all? The question is not whether or not P_1 should be considered a cause of P_2; on anyone's account, it should be. Rather, the question is whether M_1 should be given a distinct causal role in this situation. I believe there are some persuasive reasons for refusing to do so (1993, p. 354).

[10] Kim mentions this possibility because he recognizes that not all the causal powers of a realizer need attach to the realized property. For instance, neural cells might realize mental states, but presumably a mental state does not inherit all the causal powers of the mass or color of neural cells.

In asking whether M_1 should be given a *distinct* causal role in the production of P_2, Kim is inviting us to consider whether M_1 has an effect on P_2 *additional* to the effect that P_1 has. The obvious method to use in answering this question involves holding P_1 fixed while wiggling M_1. But this, of course, is just to commit the error that Figure 5 depicts. One cannot manipulate a macroproperty while holding fixed the microproperties on which it supervenes.

Kim's tacit adoption of the master argument also is visible in the following claim:

> If being real means having causal powers, being real and irreducible must mean having causal powers that are irreducible. That is, if mental properties are irreducible to those of physical/biological properties, as the nonreductivist claims, this must mean that they have causal powers that are different from those of physical/biological properties" (1996, p. 230).

Macroproperties, Kim believes, can be real only if they can do things that their physical supervenience bases cannot. But this suggests that to test whether a given macroproperty is real, one must hold fixed its supervenience base while manipulating the macroproperty and see if this makes a difference in some candidate effect term. How else to discover those powers that a macroproperty has *in addition to* those that its supervenience base possesses?

We reject the reasoning that leads Kim to endorse, at least tacitly, the master argument. We can grant that a functional property has no causal powers beyond those of its realizer. However, this does not mean that the causal relations into which the functional property enters are unreal. This conclusion requires an additional assumption -- precisely the one that Kim makes -- that functional properties have real causal powers only when they have powers *additional* to the ones possessed by their realizers. But an assumption like this requires that the functionalist accept some kind of magic. Functional properties are generally conceived as second-order properties -- the property of having some property that fills some functional role. Believing that martinis should be shaken, for instance, is the property of having some property (perhaps neural, or silicon-based, or …) that plays the role that defines such a belief. But how could a second-order property – the property of having some first-order property – have powers beyond those of a first-order property?

There is an alternative interpretation of Kim's arguments that we want to consider. It is reminiscent of an argument that Robert Nozick once described, which Dennett (1980) discusses. Suppose that smart Martians wouldn't need to attribute beliefs and desires to us to predict our behavior. Like Laplace's demon, they could simply survey the state of the atoms that make up our bodies, subject that information to a very complex calculation, and thereby predict what we'll do next. It therefore appears that the only reason we have for attributing beliefs and desires to each other is that we can't do what smart Martians can do. Nozick asks whether the possibility of smart Martians' weakens the thesis that beliefs and desires are real states that we occupy. Kim can be interpreted as posing a similar question. If we know that our behaviors have physical

causes, why do we need to say that they also have mental causes? Kim's question does not concern the existence of beliefs and desires, but their causal efficacy. Both questions suggest that Ockham's razor can be applied to the problems posed – leading to eliminativism in the case of Nozick's question and to epiphenomenalism in the case of Kim's.[11] In reply we suggest that Ockham's razor has no application to these problems (Sober 1999b). Ockham's razor is relevant only to the choice between *competing* hypotheses. However, there is no conflict between saying that we have beliefs and desires and saying that we occupy neurophysiological states. And there is no conflict between saying that a behavior has physical causes and saying that it has mental causes. Whether we take Kim to be invoking the master argument for epiphenomenalism or a parsimony argument for the same thesis, his argument fails.[12][13]

Kim continued to reject the possibility of macrocausation on this basis until 1998, when his confidence that macroproperties have causal powers revives. He now maintains that there are "no special problems about the causal powers of functional properties. And if any mental properties turn out to be functional properties, there are no special problems about their causal roles either" (p. 116). This is a surprising reversal. Why does Kim change his mind in this way? Kim is moved by an idea he calls the *functional model of reduction*. According to this model, reduction proceeds first by functionalizing the property to be reduced. Functionalization consists in defining a property in terms of its functional role. Functional reduction is then the identification of an instance of a functional property with the instance of the property that realizes it. To the extent that mental states can be defined by their functional roles, they are candidates for functional reduction. And because the functional property of having a certain belief is the property of having some physical property that plays the proper "belief" functional role, token beliefs, according to Kim, are identical with tokens of whatever physical properties realize them.

Kim sees the functional model of reduction as saving mental causation as well as many other instances of macrocausation. Believing that martinis should be shaken, on this view, can cause other mental or physical states because this belief is identical to some particular physical realizer, and there is no reason to doubt the causal efficacy of physical properties. For example, the belief about martinis can cause you to shake your martini because an instance of this belief is identical with an instance of some brain state

---

[11] Kim (1993, p. 354) apparently endorses this reasoning when he says, "First, there is the good old principle of simplicity: we can make do with P as P*'s cause, so why bother with M?"

[12] Sturgeon (1984) argues that moral facts supervene on physical facts and that this is no impediment to those moral facts' being causally efficacious.

[13] If X causes Y and MSB(X) causes Y, is Y overdetermined, and is that an objection to the two causal claims? We think that overdetermination requires two or more *independent* causes, and X and MSB(X) are not independent. Kim (1996, p. 150) takes the same view. He also says that X's causal efficacy must "piggy back" on that of MSB(X); we hasten to add that this piggyback thesis does not conflict with the claim that "X causes Y" and "MSB(X) causes Y" are both true. There is no "argument from overdetermination" that forces one to choose. However, we do disagree with Kim when he says that "for any single event, there can be no more than a single sufficient cause, or causal explanation, unless it is a case of causal overdetermination." Consider a deterministic causal chain from X to Y to Z. X is a sufficient cause of Z, and so is Y, but Z isn't overdetermined.

that causes you to shake your martini. "Reduction," Kim summarizes, "is essentially functionalization, and if the mental is reduced to the physical, we should expect no special problem about its causal powers" (1998, p. 116). The functional model of reduction, Kim believes, renders macrocausation consistent with the causal inheritance principle. In accordance with that principle, a functional property has no causal powers beyond those of its first-order realizer, but it does have those causal powers. Functionalization does not challenge the efficacy of the functional property because the functional property is identical to the first-order property that realizes it. M_1 can cause M_2 because an instance of M_1 *is* an instance of P_1, and P_1 causes P_2, which in turn is identical to M_2. Problem solved.[14]

This answer to the master argument may appear similar to the one we have recommended. We certainly agree that if the properties M_1 and P_1 are identical, then they have identical causal powers. However, our views and Kim's differ in important ways. In the first place, Kim intends his solution to work only for macroproperties that can be functionalized. Kim believes that qualia cannot be functionalized and, concludes that they cannot be causes (1998, p. 116). In contrast, from the empirical standpoint we have advocated, whether a macroproperty is functionalizable makes no difference to whether it has causal powers. Intervention provides a means by which to test which causal powers a macroproperty has. To see whether a quale causally influences a behavior, one needs to hold fixed any common causes they have; however, one should not hold fixed the micro-supervenience base of the quale. This point follows from straightforward considerations of scientific method and holds independently of whether the macroproperty under investigation is functional.

More significantly, the possibility of multiple realization creates difficulties for Kim's solution that do not attach to our own. Functional properties are ordinarily conceived as multiply realizable. A second-order functional property may have many types of first-order realizers. But Kim's functional model of reduction requires that instances of functional properties be identified with instances of realizing properties. Consequently, as Kim concedes, the reduction of a mental state M "consists in identifying M with its realizer P_i relative to the species or structure under consideration…Thus M is P_1 in species 1, P_2 in species 2, and so on … In fact, each instance of M is an instance of P_1, or of P_2, or…, where the Ps are M's realizers" (1998, pp. 110-111). We see a difficulty with this view that bears on the issue of epiphenomenalism.

Suppose, as Kim maintains, that M is P_1 in human beings and P_2 in Martians, where P_1 and P_2 are different physical properties. If so, they cannot both be identical to M. This suggests that Kim's idea is that P_1 is type identical with the property of *being the human belief about martinis* and P_2 is type identical with the property of *being the Martian belief about martinis*. If this interpretation of Kim's proposal is correct, then he apparently has given up on M. He notes in an earlier paper that as a consequence of

[14] We agree with Kim's thought, mentioned in footnote 15, that M_1 (or an instance thereof) will often not have all the causal powers of its micro-supervenience base P_1; if so, M_1 and P_1 should not be thought of as identical.

multiple realizability, "[e]ach mental kind is sundered into as many kinds as there are physical realization bases for it, and psychology as a science with disciplinary unity turns out to be an impossible project" (1992, 18).[15] M, it appears, dissolves into a potentially infinite collection of M_i, each of which is identical with some P_i. Accordingly, we might wonder whether the words 'belief about martinis' in the predicates 'human belief about martinis' and 'Martian belief about martinis' denotes a single property. Perhaps the predicates are misleading insofar as they seem to suggest that there is a property that human beings and Martians share when each believes that martinis should be shaken. A less misleading way to formulate Kim's vision of psychology as inherently disunified might use predicates that do not invite this kind of misconception – fused predicates like the 'human-belief-that-martinis-should-be-shaken' and the 'Martian-belief-that-martinis-should-be-shaken'.

If this is the conclusion to which Kim's response to the master argument leads, we cannot resist noting an irony. In his effort to avoid epiphenomenalism, he embraces the thesis that macroproperties like M are epiphenomenal. It is false, according to Kim, that you shake the martini because you have the property of believing that martinis should be shaken. The truth is that you exhibit this behavior because you have the *human* belief that martinis should be shaken.[16] In addition, Kim's view isn't just that M is epiphenomenal with respect to one class of properties (e.g., behaviors); Kim embraces the idea that M is epiphenomenal in *all* respects. This goes contrary to our contention that epiphenomenalism should be a thesis about the causal inertness of a class of properties with respect to some delimited other class. By partitioning M into various M_i, each of which has causal properties only in virtue of its identity to some physical property, Kim abandons the hope that M itself might have causal powers.

Kim is compelled to "sunder" functional properties into collections of realizing properties because he believes that macro-properties will otherwise be epiphenomenal. However, if M can supervene on P_1 without thereby losing its causal powers, M also can supervene on P_2 with no loss of causal powers. Because Kim rejects this possibility, he takes the threat of epiphenomenalism to require that M give way to distinct M_i each of which is type-identical with distinct P_i. But no such move is necessary if, as we have argued, the causal inheritance principle is consistent with the causal efficacy of macro-properties.

3. Epiphenomenalism and the Causes of Evolution

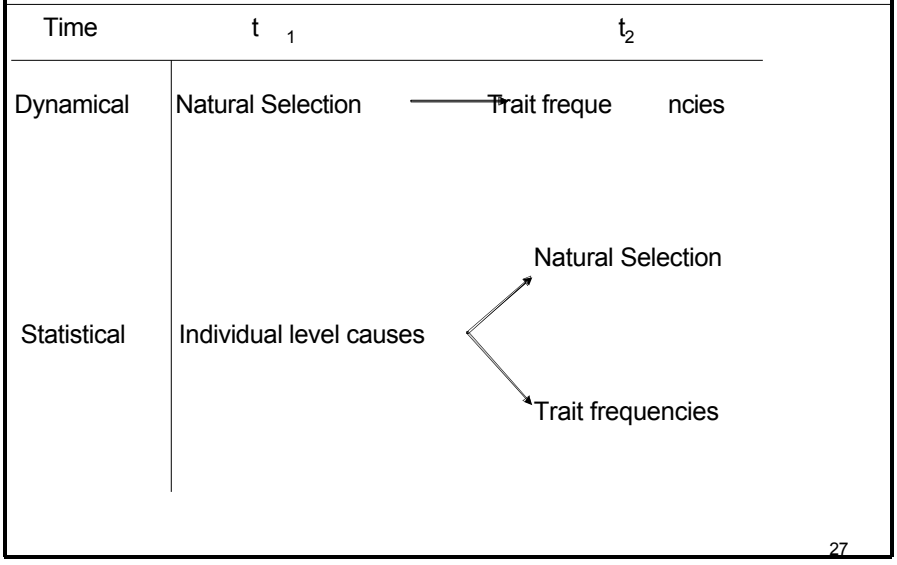
[15] See Shapiro (forthcoming) for discussion of Kim's challenge to a unified psychology.
[16] In fact, if the belief that martinis should be shaken is multiply realized *within* species, and not just *between* them, then the human belief must be causally inert by Kim's argument. And if the belief is multiply realized *within* the lifetime of a single organism, then your believing that martinis should be shaken is also shown to be causally inert.

Just as philosophers of mind sometimes contend that mental properties are epiphenomenal, so philosophers of biology sometimes contend that natural selection is epiphenomenal. This epiphenomenalist position is defended by Walsh (2000), by Walsh, Lewens, and Ariew (2002), and by Mathen and Ariew (2002). These papers develop other theses, and there are differences among them concerning how they defend epiphenomenalism; we nonetheless see a core of agreement, and this is what we have in mind when we talk about the WALM position.[17] The title of Walsh's paper, "Chasing Shadows – Natural Selection and Adaptation," conveys the epiphenomenalist punchline running through all three. Natural selection is like a shadow moving across a wall; it is a pseudo-process. Earlier stages of a shadow cause later stages no more than natural selection causes evolution.

WALM compare two ways of thinking about natural selection, which they label "dynamical" and "statistical." Figure 7 depicts the difference between them. The conventional view, which WALM oppose, is that natural selection, along with drift, mutation, migration, and mating pattern, are possible causes of evolution. These causes impinge on a population, sometimes changing its state while at other times causing the population to remain in the same state.[18] The causes of evolution behave in some ways like Newtonian forces. If two forces promote the evolution of a trait, it will increase in frequency at a faster rate than if just one of them were in place. And a population can be at equilibrium because opposing forces cancel each other; just as a shopping cart can remain stationary because someone is pushing it North while someone else is pushing it just as hard South, so a trait in a population can exhibit the same frequency across generations because, for example, selection is pushing it to increase in frequency while mutation pressure is pushing it to decline. WALM call this the dynamic view; we prefer to call it the causal view, since some philosophers, such as Lloyd (1988), are skeptical about using causal concepts to interpret dynamical models.[19]

[17] Lewens (2004, pp. 22-28) is less friendly to the epiphenomenalist position, and so we do not include this paper as part of the WALM collective.
[18] The state of a population is characterized by specifying the frequencies of different traits in it.
[19] For an elaboration of the causal interpretation, see Chapter 1 of Sober (1984), which is called "Evolutionary Theory as a Theory of Forces."

Figure 7: Two models of how natural selection might be related to change in trait frequencies.



The alternative interpretation, which WALM favor, says that selection is epiphenomenal with respect to changes in trait frequencies. The real causes are at the level of individual organisms.[20] WALM call this the statistical view, but we prefer to call it epiphenomenalist; after all, an explanation can both cite statistics and describe the causes of the event that is being explained.[21] WALM's eiphenomenalism about natural selection is similar to Kim's epiphenomenalism about mental properties. In both cases, a set of macro-properties is said to be causally inert, and a set of micro-properties is said to be doing all the causal work.

We think the master argument for epiphenomenalism is at work in Walsh's (2002) defense of the statistical position and critique of the dynamical interpretation.[22] Walsh claims that "there is no need to invoke a *distinct* force [of natural selection] operating

[20] Walsh (2000, pp.145-149) suggests that Stuart Kauffman (1993) provides important insights into the individual-level causes of adaptation. For Walsh, selection does not cause adaptation; rather, it is the adaptability and homeostatic properties of organisms that make adaptive evolution by natural selection possible. We won't discuss Kaufmann's work here, but will simply note that the question of whether he is right is independent of whether selection causes evolution.

[21] Matthen and Ariew (2002) embrace epiphenomenalism when they say (p. 79) that selection is an "aggregation of individual events … it is not a process within which the earlier events cause the later." However, much of their paper concerns their distinction between two concepts of fitness ("vernacular" and "predictive"); they assert that only the former is causal (p. 56). Later (p. 81) they say that "while predictive fitness values … may thus be considered *probabilistic causes*, they are not causes in the sense appropriate to fundamental processes (their italics)." The reason offered for this last claim is that selection processes involve "discontinuity and irreversibility (p. 82)." Here we have three comments. First, their thesis about two types of causation is incompatible with their epiphenomenalism. Second, drift is often described in terms of continuous diffusion equations that are time-reversible; this raises the question of how Matthen and Ariew can avoid viewing drift as a fundamental cause. Finally, we note that Matthen and Ariew (p. 78) also assert that "the distinction between vernacular fitness … and predictive fitness … is unmotivated."

[22] Matthen and Ariew (2002, p. 82) explicitly disavow the master argument.

over populations," when, at the level of individual organisms, there already are the many causes of individual births and deaths (italics in the original, p. 139; see also p.150). We reply that while it is true that natural selection is not distinct from its supervenience base in a given token selection process, this is not a reason to deny that selection is a cause. In the same way, we regard the temperature, pressure, and volume of the gas in a container as causes even though they supervene on the states of the molecules making up the gas. Walsh demands that selection contribute something to evolution beyond the contributions made by the causal processes the impinge on individual organisms, just as Kim demands that mental properties have powers in addition to those of their supervenience bases. Of course selection cannot do this, but that is no argument against its causal efficacy. To assess whether X causes Y, you shouldn't try to hold fixed the micro-supervenience base of X while wiggling X.[23]

Walsh, Lewens, and Ariew (2002, p. 463) talk about the kinetic theory of gases and grant that temperature, pressure, and volume are explanatory. However, their point is the same as the one made by Walsh (2000) – that these macro-properties provide a kind of noncausal explanation. According to the statistical interpretation, natural selection really does occur and it really does explain. It explains changes in trait frequencies, not by citing their causes, but by a statistically apposite kind of bookkeeping.

WALM need to address the same question that Kim (1998) takes up – if epiphenomenalism is true with respect to mental properties, or with respect to natural selection, why doesn't the argument generalize, leading to the implausible consequence that no macro-properties at all are causally efficacious? If there is a smallest chunk of matter (quarks?), it is properties attaching to those smallest chunks, and no others, that cause things to happen (Sturgeon 1984); and if there is no smallest chunk, there is no causation at all (Block 1997). Walsh, Lewens and Ariew (2002) don't answer this question in their discussion of gases, but Walsh (2000, p. 140) takes it up when he considers metabolism and meiosis. These are genuine causes, he grants, so why, he asks, can't selection also be a cause though it too is an "aggregate of distinct causal processes"? Here is Walsh's answer:

> … selection appears to be composed of its component processes in a quite distinctive way. The mere aggregate of individual survival and reproduction is not sufficient for selection. Selection only occurs as a consequence of the *difference* between individual-level processes. One could eliminate natural selection entirely by increasing the propensity of occurrences of one of its component processes. One could not, by contrast, eliminate the process of photosynthesis entirely by increasing the propensity of one of *its* component processes to occur. Perhaps by altering the propensity of a component process of photosynthesis one could affect the final outcome, perhaps even cause the discontinuation of the process, but that would not be the same as bringing it about that no such process occurs *at all* (italics in original).

[23] Notice that Walsh's framing of the question about natural selection resembles Nozick's "smart Martian problem," which we discussed in the previous section.

If natural selection is epiphenomenal for this reason, it seems that discrimination on the basis of sex or race or religion is too. So much for the claim that these are causes of social inequalities. And the sifting process that makes smaller corn flakes settle to the bottom of the box and larger flakes rise to the top cannot occur when all the flakes have the same size, so it seems that Walsh will have to conclude that sifting is not a causal process. In our view, sifting occurs only when there are differences, but that does not prevent such processes from being causes.[24]

Walsh's argument reminds us of the Leibnizian question of whether relations reduce to nonrelational properties. In some cases they arguably do; x's being taller than y seems to reduce to x's having one height and y's having another, lesser, height. But other relations seem not to reduce; x's kissing y doesn't reduce to x's having one nonrelational property and y's having another. Notice that x can be taller than y even if they never come in causal contact and aren't even affected by a common cause that influences their heights; they could exist at opposite ends of the universe. Kissing, in contrast, precludes such remoteness. Walsh seems to be suggesting that the "relational process" of natural selection is more like x's being taller than y than it is like x's kissing y. WALM repeatedly assert that selection exists when two sets of individuals (call them the x-individuals and the y-individuals) differ in fitness. But this can occur when x has nothing causally to do with y; x and y might just as well be at opposite ends of the universe.

Our reply is that no biologist would treat two individuals as part of the same (token) selection process if they were at opposite ends of the universe (Sober 1980). The fact that x and y differ in fitness does not entail that there is a selection process impinging on both. Sometimes x and y experience the same token selection process because they causally interact; at other times they participate in the same selection process because they are affected by a common token cause. Darwin drew this distinction in a famous passage from the *Origin*:

> I should premise that I use the term Struggle for Existence in a large and metaphorical sense … Two canine animals in time of dearth, may be truly said to struggle with each other which shall get food and live. But a plant on the edge of a desert is said to struggle for life against the drought (Darwin 1859, p. 62)

When two plants in the same population differ in fitness, Walsh sees one process affecting the one and a different process affecting the other; there is no single process affecting them both. Our reply is that we find it entirely natural to point to the drought. We find it even more puzzling how Walsh can think of two dogs fighting over a piece of meat in terms of one process affecting the one and a different process affecting the other. Direct competition between individuals does not decompose into separate, nonrelational processes impinging on each; sometimes selection is like kissing.

[24] The process of seeing an object requires that there be a visual contrast between the object and something else in the visual field; this is why you can't see the proverbial white cat in a snowstorm. Does Walsh have to maintain that visual perception is not a causal process because it depends on the existence of differences?

WALM offer other reasons for thinking that selection is not a cause of evolution. As just noted, they equate natural selection with the existence of variation in trait fitnesses. Since differences in trait fitness don't cause evolution, selection doesn't either. Here they concur with the claim made in Sober (1984) that "fitness is causally inert." Your viability fitness (as measured by your life expectancy) doesn't cause you to live a certain length of time, and your fertility fitness (measured by your expected number of offspring) doesn't cause you to have babies. Attributing causal powers to fitness sounds wrong for the same reason that it sounds wrong to say that someone was caused to fall asleep by drinking a potion that possessed a dormitive virtue. However, we are not so sure that the claim is false; maybe it is true but relatively uninformative. Davidson (1963) says that "the cause of x caused x" is both true and analytic; maybe the same is true of the claim that the component of your fitness that reflects your viability causes you to live a certain number of years.

In any event, even if fitnesses are causally inert, it doesn't follow that natural selection is too. The distinction between the concepts of *selection-of* and *selection-for* explains why (Sober 1984). It is true that

> There is selection *of* trait T in population p if and only if T is fitter than not-T in p.

But it is false that

> There is selection *for* trait T in population p if and only if T is fitter than not-T in p.

"Selection-of" is extensional, while "selection-for" is not. Consider a population in which there is selection for being green; this selection pressure exists because being green camouflages organisms in the green environment they occupy, thus protecting them from predators. Suppose further that there is no selection for being small – body size is selectively irrelevant. And now imagine that all and only the green organisms in the population are small. In this situation, the green organisms are selected, which means that the small ones are too. However, though there is selection for being green, there is no selection for being small. Selection-of is the concept that is tightly connected to variation in fitness; if there is selection of green (small) organisms, then they are on average fitter than those that are not. But the fact that small organisms are fitter than organisms that are not small does not entail that there is selection for being small. Selection for this property and against that one can cause traits to change frequency. Walsh, Lewens, and Ariew (2002, p. 466) equate selection with variation in trait fitness; as a result they focus exclusively on selection-of and neglect the concept of selection-for. But selection-for is where the causal action is (Sober 1984).

WALM's epiphenomenalism also depends on ignoring the fact that "natural selection" sometimes describes a process and at other times describes the outcome or

product of that process (Millstein 2002, Stephens 2004).[25] This is a familiar ambiguity. "Erosion" can refer to the dirt at the bottom of a hillside or to the process that brought it there; "marriage" can refer to the process of getting married or to the outcome state of being married. When WALM talk about natural selection, they usually are discussing the outcome – the fact that various traits were selected. But even if selection-the-product is not a cause of trait evolution, selection-the-process might be. Consider Figure 7. If selection is the outcome that occurs at time t_2, then of course it is not the cause of the trait frequencies that exist at t_2. But it may still be true that the process of selection that occurs at t_1 causally influences the trait frequencies at t_2. The temporal indices in this figure are our invention; they are not found in WALM's discussion. Being explicit about the times at which selection occurs makes it harder to elide the distinction between process and product. WALM are right about selection-the-product, but wrong about selection-the-process.

This distinction is very much needed to understand what is going on when populations are at equilibrium. If a trait's frequency in a population remains unchanged, does it follow that there is no selection? Walsh, Lewens, and Ariew (2002) assert that selection requires change in trait frequencies (p. 464, p. 466).[26] But it is perfectly clear that a selection process and a mutation process can both occur without a change in trait frequencies (Stephens 2004). Suppose selection favors allele *A* over allele *a*, but the mutation rate from *A* to *a* exceeds the rate from *a* to *A*. A population evolving under these two regimes will eventually reach an equilibrium trait frequency. Does selection cease to occur when the population stops evolving? WALM assert that it does. Perhaps they also would deny that there is mutation pressure at equilibrium. But why, then, is the allele frequency in the population not changing? The Newtonian answer is that the two forces cancel each other, with a resultant of zero. Selection (the process) can occur without evolution.

The confusion of process and product pervades WALM's discussion of drift. Walsh, Lewens, and Ariew (2002) say that drift occurs when and only when a population exhibits a trait frequency that deviates from the frequency that would be predicted if selection alone acted. To consider what selection without drift would be like, one needs to consider an infinite population in which the trait frequencies and fitnesses in one generation deterministically predict trait frequencies in the next. Suppose that the current generation contains 50% trait A and 50% trait B and that selection acting alone predicts that there should be 55% A and 45% B in the next generation. Walsh *et al.* claim that if the frequencies in the next generation depart from those figures, then there was drift; but if the observed trait frequencies are precisely 55% and 45%, there was no drift. For WALM, drift is a possible outcome; it is a product, not a process.[27] However, even if drift-the-product at t_2 is not a cause of trait frequencies at t_2 (how could it be, if cause

[25] Endler (1986, p.29) also argues that natural selection is not a cause (or at least not a "force") by focusing on selection-the-product and ignoring selection-the-process: "Natural selection no more 'acts' on organisms than erosion 'acts' on a hillside … It is a *result* of heritable biological differences among individuals, just as erosion is a result of variation in resistance to weathering and running water."

[26] Matthen and Ariew (2002, p. 82) go further; they say "it makes dubious sense to hold evolution is different from natural selection, much less that it is caused by it."

[27] Brandon and Carson (1996, pp.324-325) and Brandon (2005, p. 158, p. 169) view drift in the same way.

must precede effect?), nothing follows as to whether drift-the-process at t_1 affects trait frequencies at t_2. Our view is that drift (the process) occurs in a population whenever the population is finite, just as the process of selection occurs in a population whenever there is variation in fitness; see Beatty 1984, Sober 1984, Millstein 2002, and Stephens 2004 for discussion. Walsh *et al.* argue that drift and selection can't be distinct processes because the same processes that cause selection also cause drift (p. 464). We reply that even if selection and drift had the same causes, it wouldn't follow that they aren't distinct causes of evolution. We view selection and drift as distinct processes whose magnitudes are represented by distinct population parameters (fitnesses on the one hand, effective population size on the other). Changes in each of these parameters will be associated with changes in the probabilities of different outcomes. If you intervene on fitness values while holding fixed population size, this will be associated with a change in the probability of different trait frequencies in the next generation. And the same is true if you intervene on population size and hold fixed the fitnesses.[28]

4. Concluding Comments – The Relation of Macro- to Micro-Causation

The negative arguments of the previous sections can be supplemented with something more positive concerning the relationship of macro- and micro-causation.. Consider Figure 8, which describes a macro-variable X that has n possible macro-states $X_1, X_2, \ldots, X_n$, where each macro-state X_i can be realized by m possible micro-states X_{i1}, $X_{i2}, \ldots X_{im}$. There are nm micro-states depicted here; each micro-state X_{ij} is a possible state of the micro-variable MSB(X), where MSB(X) is the micro-supervenience base of the macro-variable X. The first claim we want to make about the relationship of macro- and micro-causation is this: *if X causes Y, so does MSB(X)*. The reason is clear from the point of view of a manipulationist theory of causation such as Woodward's (2003). If X causes Y, then there are states of the variable X, call them X_i and X_j, such that an intervention on X that changes X's state from X_i to X_j will be associated with a change in the state of Y. If this is true, then there also will be an intervention on the micro-variable B(X), changing it from some state X_{ik} to some state X_{jl} that also must be associated with a change in Y. So macro-causation entails micro-causation.[29,30,31]

---

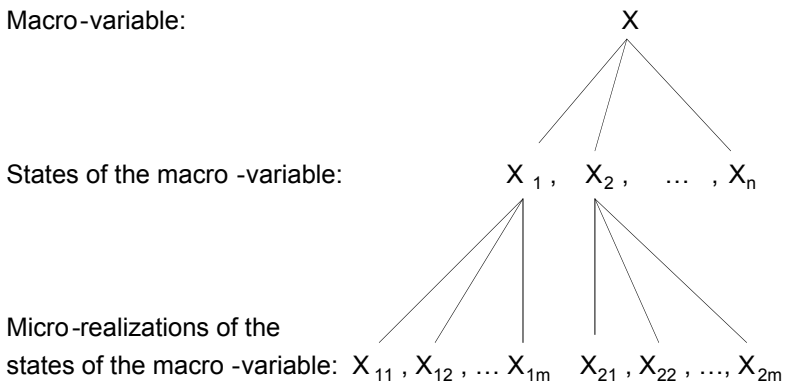
[28] Reisman and Forber (2005) describe an experiment of Dobzhansky and Pavlovsky's in which population size was manipulated in replicate populations and differences in evolutionary outcomes were observed. Reisman and Forber argue that this shows that drift can cause evolution.

[29] The thesis that macro-*causation* entails micro-*causation* does not, by itself, settle the question of whether macro-*explanation* entails micro-*explanation*. There still is room for the familiar anti-reductionist claim, due to Putnam and Fodor, that macro-stories are explanatory while micro-stories are not (or are less so). See Sober (1999a) for criticisms of this anti-reductionist thesis.

[30] This argument shows that the supervenience thesis plus the interventionist account of causation vindicates Kim's causal inheritance principle, at least if the latter is understood as a claim about the relationship of types.

[31] Waters (forthcoming) uses Woodwards' (2003) manipulability account of causation to argue, as we have done, that macro-causation entails micro-causation, and concludes from this that the units of selection controversy rests on a confusion; he contends that if there is group or organismic selection in a lineage, then it also is true that there is genic selection as well. We believe that this conclusion rests on a misunderstanding of what the units of selection controversy is about; for an account that regards group,

Figure 8: The macro-variable X has n macro-states X_1 X_2, …, X_n; each macro-state X_i is multiply realizable by m micro-states X_{i1}, X_{i2}, …, X_{im}. The micro-states X_{ij} are states of the micro-variable MSB(X).



The converse relation, however, is not inevitable. The existence of an intervention on the micro-variable MSB(X) that will be associated with a change in Y does not guarantee that there is an intervention on the macro-variable X that also will be associated with a change in Y. For suppose there are interventions on the micro-variable MSB(X) that change its state from X_{ij} to X_{ik} and that are associated with a change in Y, but that changes from X_i to X_j *on average* are not associated with change in the state of Y. As an example, suppose that for each macro-state X_i, when individuals occupy X_i, they are just as likely to be in micro-state X_{ij} as they are to be in micro-state X_{ik}. And suppose that half the micro-realizations of each X_i confer on Y a probability of 0.3 and half confer on it a probability of 0.9. Then there are changes from one micro-state to another that will be associated with a change in the probability of Y, but an intervention that shifts the system from X_i to X_j will, on average, not affect the probability of Y, since $Pr(Y\_X_i) = Pr(Y\_X_j) = 0.6$. Here we have micro-causation without macro-causation.[32]

---

organismic, and genic selection as separate and *independent* causal processes, see Sober and Wilson (1998).

[32] Yablo (1992) thinks it does make sense to ask whether X at t_1 or MSB(X) at t_1 causes Y at t_2, where MSB(X) is the micro-supervenience base of X. He agrees that you can't vary X while holding fixed MSB(X), but argues that it is relevant to consider the converse manipulation -- what would happen if you varied MSB(X) while holding X fixed? Yablo says that MSB(X) causes Y precisely when the manipulation just described would be associated with a change in Y. We note that this provides no information about whether X is a cause. We suspect that MSB(X) almost always has an impact on Y over and above the impact that X has, provided that $0 < Pr(Y\_X) < 1$ (Sober 1999a, 2005); if this is right, Yablo's criterion entails that MSB(X) almost always causes Y. In any case, we think that Yablo's procedure for deciding whether MSB(X) causes Y is off the mark. If there are manipulations of MSB(X) that do *not* involve holding X fixed, and which are associated with a change in Y (or in the probability of Y), then this shows that MSB(X) is a cause of Y.

The asymmetry obtained here – that macro-causation entails micro-causation, but not conversely -- depends on a particular formulation of the manipulationist criterion for causation. The criterion we have used says that "there exists a possible intervention on X that would be associated with a change in Y (or in the probability of Y)." There is a stronger formulation of the criterion that produces an asymmetry that is opposite from the one for which we have argued. If "X causes Y" is taken to mean that *every* change in the state of X will be associated with a change in the state of Y, then micro-causation entails macro-causation, but not conversely. We think this stronger formulation is implausible; for example, "smoking causes lung cancer" can be true even if there are threshold effects (suppose your probability goes up only after you've smoked more than m cigarettes and that it fails to go up after you've smoked more than n).

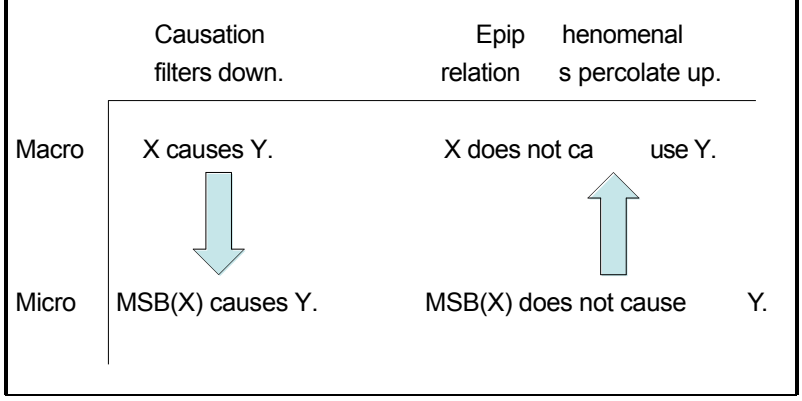
Given our thesis that macro-causation entails micro-causation, but not conversely, it is possible that a given macro-variable X is causally inert though its micro-supervenience base MSB(X) is causally efficacious with respect to some effect term Y. However, this possibility is not the setting within which familiar and unproblematic examples of epiphenomenalism arise. The color of the circle of light on the ceiling of the Astrodome at one time does not influence the color of the circle of light that appears there a moment later, *but the same is true of the micro-supervenience bases of those circles of light*. And although parental eye color does not influence the eye color of offspring, *the same is true of the micro-supervenience bases of parental and offspring eye color*. These examples of epiphenomenalism are not due to a macro-property's being causally inert while its micro-supervenience base is causally efficacious. Rather, they are due to the fact that a property is screened off from an effect term of interest by a common cause; it is entirely incidental whether the screened-off property is macro or micro, or whether the screening off common cause is micro or macro. In the barometer example, the barometer reading (a macro-property) is screened off from the storm by the barometric pressure, which is a macro-property of the volume of air. The micro-supervenience base of the barometer reading also is screened off, and it also is true that the micro-state of the molecules in the volume of air induces the screening-off relation.[33]

Just as there are asymmetries between micro- and macro-causation, so too there are asymmetries between micro- and macro-epiphenomenalism. In fact, the main point concerning epiphenomenal relations is just the contrapositive of the conditional that characterizes the relationship between macro- and micro-causation. This is depicted in Figure 9. If no intervention on MSB(X) will be associated with a change in the state of Y, then neither will an intervention on the state of X be associated with a change in Y. A failure of causal relations at the micro-level entails a failure of causal relations at the macro-level. If MSB(X) and Y are nonetheless correlated, this invites the hypothesis that

[33] Mill's (1844) example of day following night and night following day is another example of a pseudo-process, with the motion of the earth and the position of the sun providing the screening-off common cause. The properties of those heavenly bodies are macro-properties, but nothing of substance would change if we talked instead about the properties of the atoms that make up the sun and the earth. Mill's example also illustrates the fact that it is possible to know that the connection of two or more variables constitutes a pseudo-process without actually manipulating any of them; see footnote 3.

they are joint effects of a common cause, and the same holds if X and Y are correlated. Causation filters down and epiphenomenal relations percolate up.

Figure 9. If macro -causation entails micro -causation, then micro-epiphenomenalism entails macro -epiphenomenalism.



Reflection on the nature of pseudo-processes leads us to conclude that the essence of epiphenomenalism is a correlation between two variables that is due solely to their having a common cause. Epiphenomenalism is the right analysis of the mirror image, the shadow, the circle of light moving across the ceiling of the Astrodome, the regular alternation of night and day, and many cases of phenotypic resemblance between parents and offspring. Whether natural selection is screened off from evolution by their common causes, and whether mental states are screened off from behavior by theirs, are questions for science, not armchair philosophy, to determine, the master argument and the smart Martian argument notwithstanding.

Acknowledgements: We are grateful to André Ariew, Kevin Brosnan, Juan Comesaña, Patrick Forber, Malcolm Forster, Clark Glymour, Franz-Peter Griesmaier, Daniel Hausman, Christopher Hitchcock, Tim Lewens, Mohan Matthen, Peter McLaughlin, Mark Moffett, Steve Nadler, Thomas Polger, Stathis Psilos, Carolina Sartorio, Edward Sherline, Christopher Stephens, Denis Walsh, Ken Waters, James Woodward, and Stephen Yablo for useful comments.

References

Beatty, J. (1984): "Chance and Natural Selection." *Philosophy of Science* 51: 183-211.

Block, N. (1997): "Anti-Reductionism Slaps Back." *Philosophical Perspectives* 11: 107-132.

Brandon, R. (2005): "The Difference Between Selection and Drift -- A Reply to Millstein." *Biology and Philosophy* 20: 153-170.

Brandon, R. and Carson, S. (1996): "The Indeterministic Character of Evolutionary Theory -- No 'No Hidden Variables Proof' but no Room for Determinism Either." *Philosophy of Science* 63: 315-337.

Davidson, D. (1963): "Actions, Reasons, and Causes." In *Essays on Actions and Events.* Oxford University Press, 1980, pp. 3-20.

Dennett, D. (1980): "True Believers – the Intentional Stance and Why it Works." *Herbert Spenser Lecture on Scientific Explanation.* Oxford University Press. Expanded version in *The Intentional Stance*. Cambridge: MIT Press, 1989, pp.13-42.

Eells, E. (1991): *Probabilistic Causation.* Cambridge University Press.

Endler, J. (1986): *Natural Selection in the Wild*. Princeton University Press.

Graham, G. (1998): *Philosophy of Mind: An Introduction*, 2nd. ed. Oxford: Blackwell.

Hitchcock, C. (2001): "A Tale of Two Effects." *The Philosophical Review* 110: 361-396.
Kauffman, S. (1993): *The Origins of Order – Self-Organization and Selection in Evolution.* Oxford: Oxford University Press.

Kim, J. (1984): "Epiphenomenal and Supervenient Causation," as reprinted in J. Kim, *Supervenience and Mind*. New York: Cambridge University Press, 1993, pp. 92-108.

Kim, J. (1992): "Multiple Realization and the Metaphysics of Reduction," as reprinted in J. Kim, *Supervenience and Mind*. New York: Cambridge University Press, 1993, pp. 309-335.

Kim, J. (1993): "The Nonreductivist's Troubles with Mental Causation," as reprinted in J. Kim, *Supervenience and Mind*. New York: Cambridge University Press, 1993, pp. 336-357.

Kim, J. (1996): *Philosophy of Mind*. Boulder: Westview Press

Kim, J. (1998): *Mind in a Physical World*. Cambridge: MIT Press.

Lewens, T. (2004): *Organisms and Artifacts.* Cambridge: MIT Press.

Lloyd, E. (1988): *The Structure and Confirmation of Evolutionary Theory*. New York: Greenwood Press.

Matthen, M. and Ariew, A. (2002): "Two Ways of Thinking about Fitness and Natural Selection." *Journal of Philosophy* 119: 55-83.

Mill, J. (1844): *A System of Logic, Ratiocinative and Inductive*. London: Longman, Green, & Co.

Millstein, R. (2002): "Are Random Drift and Natural Selection Conceptually Distinct?" *Biology and Philosophy* 171: 33-53.

Reisman, K. and Forber, P. (2005): "Manipulation and the Causes of Evolution." *Philosophy of Science* //: ///-///.

Salmon, W. (1984): *Scientific Explanation and the Causal Structure of the World.* Princeton: Princeton university Press.

Shapiro, L. (2005): "Can Psychology Be a Unified Science?," *Philosophy of Science* //: ///-///.

Sober, E. (1980): "Holism, Individualism, and the Units of Selection. in P. Asquith and R. Giere (eds.) *Proceedings of the Biennial Meetings of the Philosophy of Science Association* 2: 93-101

Sober, E. (1984): *The Nature of Selection – Evolutionary Theory in Philosophical Focus.* Cambridge: MIT Press; 2nd edition, University of Chicago Press, 2000.

Sober, E. (1999a): "The Multiple Realizability Argument against Reductionism." *Philosophy of Science* 66: 542-564.

Sober, E. (1999b): "Physicalism from a Probabilistic Point of View." *Philosophical Studies* 95: 135-174.

Sober, E. (2005): "Evolutionary Theory and the Reality of Macro-Probabilities." In E. Eells and J. Fetzer (eds.), *Probability in Science.* Open Court Publishers.

Sober, E. and Wilson, D. (1998): *Unto Others – the Evolution and Psychology of Unselfish Behavior.* Cambridge: Harvard University Press.

Stephens, C. (2004): "Selection, Drift, and the 'Forces' of Evolution." *Philosophy of Science* 71: 550-570.

Sturgeon, N. (1984): "Moral Explanations." In D. Copp and D. Zimmerman (eds.), *Morality, Reason, and Truth*. Totowa, NJ: Rowman and Allanheld, pp. 49-78.

Walsh, D.: (2000) "Chasing Shadows – Natural Selection and Adaptation." *Studies in the History and Philosophy of Biology and the Biomedical Sciences* 31: 135-153.

Walsh, D. (2004): "Bookkeeping or Metaphysics? The Units of Selection Debate."

Synthese 138: 337-361.

Walsh, D., Lewens, T. and Ariew, A. (2002): "The Trials of Life – Natural Selection and Random Drift." *Philosophy of Science* 69: 452-473.

Waters, K. (2005): "Why the Genic and Multi-level Selection Theories are Here to Stay." *Philosophy of Science* 72: 311-333.

Weismann, A. (1889): "On Heredity." In *Essays upon Heredity and Kindred Biological Problems.* Oxford: Clarendon Press.

Woodward, J. (2003): *Making Things Happen*. New York: Oxford University Press.

Woodward, J. (2005): "Causation with a Human Face." In H. Price and R. Corry (eds.), *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*. Oxford: Oxford University Press.

Yablo, S. (1992): "Mental Causation," *Philosophical Review* 101: 245-280.