Benjamin Newman '01
SHS: Philosophy of Mind
Essay Topic #1

Global Explanatory Reduction

"Everything is the way it is because it got that way."
— D'Arcy Wentworth Thompson

Kim [1993] contends that the causal autonomy of the mental is inconsistent with materialism, arguing that materialistic accounts of psycho-physical relationships (even supervenience) must be reductionistic. Antony and Levine [1997] argue for a reductionism which preserves mental autonomy by making mental properties second-order. Following them, Jackson and Pettit [1988], and Dennett [1991, 1995], I will propose and defend (1) a reassessment of what it means for one science or set of properties to reduce to another and (2) a version of functionalism which is reducible in this sense but maintains the causal autonomy of the mental in a meaningful (if unorthodox) sense. Both will rely heavily on the fact that real world properties are messy, massively relational, and historically contingent.

Down with (Global and Local) Supervenience!

One basic problem with supervenience, alluded to by Kim [1993], is that it comes in too many kinds. In general, a set of properties *A* supervenes on a set of properties *B* if there can be no *A*-difference in the absence of a *B*-difference. Many relationships meet this criterion, and most of these are useless to us. We can eliminate *global supervenience*, in which there can be no mental differences *between worlds* in the absence of physical differences between them. Here, arbitrarily small physical differences, such as the displacement of a single atom, can support arbitrarily large mental differences, such as the global absence of mentality [*Ibid.*, p. 277]. This is not what we want. How about *local* supervenience, in which there can be no mental differences *between different individuals in the same world* in the absence of physical differences between them. To avoid another spiral into triviality, it is necessary to further refine this, and Kim argues convincingly that the necessary refinements are sufficient to guarantee that there will be psycho-physical laws, and thus reduction, undermining the causal autonomy of the mental. I agree with Kim on this point. I disagree, however, with the claim that strictly local supervenience is the kind of supervenience we want.

Antony and Levine [1997] pick up on multiple realization — the idea noted by Kim that mental properties may be disjunctions of otherwise unrelated physical properties — to propose a form of reduction that preserves the autonomy of the mental. They claim that the terms of such disjunctions *are* related, because they share second order properties, such as filling a particular role. They also appeal to the context-dependence of multiple realization. In both of these claims they echo Jackson and Pettit [1988], who claim that mental properties are defined in terms of roles, and that contentful states have their contents only extrinsically. In our terms, both are suggesting an *almost, but not quite, local* form of supervenience, backed by necessarily *ceteris paribus* psycho-physical laws. In this regard, I agree, but would contend that an account remains to be given of what *almost-local supervenience* is, and what *ceteris paribus* means in this case. In what follows, I will briefly sketch such an account.

A New Kind of Reduction

All of the parties here are concerned with what we might call *local reductionism* — a single property *P* reduces to a set of properties *S* if it can be defined in terms of them, and a set of properties *A* reduces to a set *B* if everything in *A* reduces to some subset of *B*. This is local both in that it can treat properties individually and in that it treats *instantiations* of properties individually. This is not the sense of reduction which is most relevant to the sciences. I propose that the kind of reduction we want is *global explanatory reduction*, by which I mean the following: A set of properties *A* reduces to a set *B* if there is an *B*-explanation of the fact that the *A*'s are *worthwhile properties*. This entails that properties which are not worthwhile do not reduce to anything at all — so what are worthwhile properties?

Simply, worthwhile properties are those which are worth discriminating. In practice, they pick out *a posteriori* natural kinds. One way of refining this is to say that worthwhile properties are those for which detectors exist *in fact* — thus, worthwhileness is historically contingent. Alternatively, we can say that worthwhile properties are those which figure in *good* causal explanations — explanations which support counterfactuals, *ceteris paribus*. But, causal explanations of what? Typically, the properties in set *A* will figure in causal explanations of events described in terms of other properties in *A*. This is why the form of reduction described above is global — all of the *A*'s must be explained at once. Unfortunately, it seems that it may also be circular!

An example from Dennett [1991, p. 375] may clear things up. In our world, *red* is a worthwhile property, primarily because *red things look different from non-red ones*. Why is this so? Once upon a time, there were no organisms with color vision. At this time, there were also no particularly colorful

plants — no flowers, and no brightly colored fruits. Color vision in animals, particularly pollinating insects, co-evolved with the use of color by plants. Of course, this is a gross simplification, but this much does follow: If there were no animals for which *red* was a worthwhile property, there would, *ceteris paribus*, be no fruits which turned red when ripe, and *vice versa*. Thus, *red* reduces, not to physics, but to natural history! Dennett suggests [*Ibid.* and 1995] is that this is the reduction we should be looking for — that mental properties *in general* reduce globally to natural history (or, in the case of intelligent robots, to engineering), which reduces through several layers of biology and chemistry before finally bottoming out in physics. *All* of these steps are global explanatory reductions.

Isn't it still the case that, in a materialistic world, all properties must reduce locally to physics? At one level, yes, since detectors must be physical devices. However, there might be a property P such that two perfectly paradigmatic P-detectors detect different physical properties. This might be because P is not all-or-nothing, and the cases where the detectors disagree are borderline cases. In this case, P's definition in terms of physical properties may be impossible to formalize (so-called "fuzzy logic" grew out of attempts to formalize such definitions). Alternatively, it might be because P is multiply realized. In this case, the two detectors might resemble each other only in that they play similar causal roles in larger systems, and we might wish to define P in terms of those roles, rather than as an unwieldy disjunction of physical properties.

How do worthwhile properties which can only be reduced globally figure in causal explanations? The causal laws which ground A-explanations are contingent upon the existence of systems of interacting A-property producers (pigments in flowering plants) and A-property detectors (the retinas of insects). *Ceteris paribus* clauses in these laws cover cases where these producers or detectors are mismatched, or where they malfunction. The A-properties end up being defined in terms of the roles they play in these systems. Since it is the *actual existence* of such systems that makes the A-properties, defined in this way, worthwhile, a global explanatory reduction of A is no more and no less than an *account of the existence* of systems whose component parts' interactions are mediated by A-properties. By contrast, a local reduction of A is a redescription of such a system (in the case of multiply realized properties, each realization must be reduced separately) in terms of non-A properties. There are two levels of *ceteris paribus* clauses here, corresponding to two levels of causal explanations supporting two levels of counterfactuals which must be qualified — at the global level, the causal B-explanation of the worthwhileness of the A's, and at the local level, a variety of causal A-explanations.

Reducing the Mental

I have already said, above, that mental properties should be *globally explanatorially reduced* to terms of natural history and/or engineering. Because this form of reduction grounds properties that are defined in terms of roles, my proposal is a form of functionalism, quite similar to that proposed by Dennett [1991]. I will thus leave to Dennett the exposition of the particulars of such a theory, and concentrate on clarifications and questions which are pertinent to the topic of this essay.

To begin with, what properties are these "mental properties" we are reducing? It turns out that there are several groups of properties which can be called mental, most of which can be globally reduced much more easily than they can be locally reduced. First, there are properties of *perceptions.* These are the easiest mental properties to reduce locally, since they are realized by features of an agent's external environment. Perceptual properties are, of course, detected by an agent's sensors. Traveling inwards, we find properties which figure in *cognitive psychological* explanations. These are realized in humans by micro-features of brain states, but are best defined in terms of their roles in information-processing systems — this is why they can, under the right circumstances, be genuinely instantiated by computer states. The producers and detectors of these cognitive properties are subsystems of the brain. Moving outwards again, we find properties which individuate *actions.* These are realized by bodily movements, but are not easily defined in terms of them, being more easily defined in terms of the roles they play in the life of an agent. Finally, we have *folk psychological* properties, which are globally reducible to other mental properties. These are in general realized by the state and situation of the whole agent, although some of them may map to particular cognitive properties.

Here is an example: A tiger is thirsty, so he walks to a nearby pond and drinks his fill. Can we give an account of this simple story in terms of the above kinds of mental properties? The tiger is *dehydrated* — a physical and sensory property which is extremely worthwhile (to the tiger!), and which reduces globally to facts of natural history, such as the fact that animals need water to survive. An internal sensor detects this dehydration, and initiates an instance of *thirst*, a cognitive (and folk) state defined by, among other things, its tendency to cause *water-seeking* actions. The tiger initiates an extended action which has this property, realized by body movements which could also be described as *walking*, or even *walking to the pond.* These movements are only water-seeking, however, because the cognitive states which caused them included *thirst* and *recollections that there is water in a particular place* — otherwise, the tiger's action might not be water-seeking, even if it turned out to be water-finding. Folk psychologically, we would describe the tiger as *desiring* to drink some water, *believing*

that the pond is right over there, and thus *going to the pond to get a drink*. The producer of these properties is the tiger, and the detectors are any human observers (and the tiger itself, but only if it is *self-aware*). These properties are globally reducible (in part) to facts about the high survival value, for humans, of being able to predict the behavior of nearby tigers.

This story emphasizes the degree to which mental properties are *relational* — there are causal relations, referential relations, relations of inclusion (the tiger's total state includes the states of its brain, limbs, etc.), and relations of definition. They are also typically *worthwhile*, not just to psychologists and philosophers, but to the *very same agents* that realize them. Our folk psychological account of the tiger's activities is worthwhile to us *because* the tiger's brain states *qua* cognitive states are worthwhile to the tiger. We can also see why global reduction is preferable to local reduction as an account of which properties are mental properties. Telling the above story in terms of basic physics would require trillions of trillions of trillions of brute facts! It would causally explain an arbitrarily individuated collection of brute physical facts in terms of another such collection. This story could only be obtained or communicated at great cost, but it would be of no conceivable use to anyone. Materialism demands that we verify that such a story exists, but not that we tell it! Instead, we have told a story that is easier to obtain *and* easier to use, both as an explanation and in prediction. True, the story we have told is more weakly supportive of counterfactuals, thanks to the *ceteris paribus* clauses in the laws which ground it, but this is a small price to pay.

The functionalism which I have briefly sketched here accommodates several distinct levels of description within the mental. It is materialistic in that it admits only properties which may be detected by physical devices. (What if the idea of a worthwhile property presupposes mentality? I believe that my redefinition of worthwhileness in terms of systems closes this loophole.) Is it really reductive? This is mainly a matter of definition — I believe that my *global explanatory reduction* is reduction, in one of the senses current in the scientific community (but perhaps not in the philosophical community). If the idea of global explanatory reduction is current in the philosophical community under another name, unrelated to reduction, I would readily concede this point — happily, in fact, for it would make my theory a nonreductive materialism, *contra* Kim. But this is mere wordplay. There remains a serious philosophical issue at stake — does it preserve the causal autonomy of the mental?

What is Autonomy, Anyway?

The *causal autonomy of mental properties* obtains wherever events with mental properties enter into causal relationships with other events (mental or otherwise) *in virtue of their mental properties*. In other words, causal autonomy supports the counterfactual: If event *E* had not had mental property *M*, it would not have caused event *E**, and, *ceteris paribus*, *E** would not have occurred. What does the *ceteris paribus* clause mean here? Local reduction and supervenience guarantee that, to strip *E* of the property *M*, we must change some physical things, so we can be quite sure that other things are *not* equal in the counterfactual world we are imagining. Event *E* was *supposed to have* property *M*, but something went wrong. I will distinguish two possibilities — unsurprisingly, they can be called local and global. Locally, we should ensure that all other things are equal by changing event *E* to the most similar alternative event that lacks *M*, leaving the rest of the world unchanged as much as possible. This might very well prevent *E**. If we altered event *E* by imagining that some external influence interfered with its progress, the failure of *E** to occur is covered by the *ceteris paribus* clause of the law under which *E*'s having *M* caused *E**.

What about removing *M globally*? This corresponds to Kim's counterfactual, *contra* Davidson [1970, cited in Kim 1993], in which all mental properties are removed from the world. Here, it is the internal properties of the event *E* (in other words, relations among its subevents) which remain fixed, while the rest of the world is changed, albeit as little as possible, so that *M is no longer a worthwhile property*. Is the counterfactual claim that *E** does not occur still supported? Yes, provided that *E*'s original causation of *E** in virtue of *M* was mediated by an *M*-detector, for in changing the world so that *M* is no longer a worthwhile property, we have had to eliminate all the *M*-detectors. This is a very strange invocation of *ceteris paribus*, of course, but when we make such sweeping counterfactual assumptions, where else are we to look for "other things" that remain the same?

Thanks to two unorthodox — but necessary — interpretations of *ceteris paribus*, an account of mental properties *via* global explanatory reductionism *does* support the counterfactuals required for them to have causal autonomy. And it is, I claim, materialistic. Thus, it counters Kim's claim that a materialistic account cannot preserve the causal autonomy of the mental. Where is the flaw in Kim's reasoning? Two possibilities come to mind. First, it could be argued that my account is a form of *property dualism*, if global reduction sanctions properties whose local reductions cannot be formalized. But in this case I would claim that property dualism *of this sort* is not really dualism, for nowhere have I introduced nonphysical entities, events, or causes. Second, of course, is Kim's use of an overly strict

sense of *reduction*, one that has already fallen out of favor among scientists. I would agree with Kim's claim that a totally nonreductive account of the psycho-physical relation *is not really an account*, but I choose a less strict sense of *reduction*, so that my account can be reductionistic in this sense while not being so in the strict sense which, as Kim argues, undermines the causal autonomy of the mental.

In Conclusion

The real world is a mess, full of extrinsic, relational properties and catastrophic historical contingencies. These features, too often idealized away in philosophical discussion, give us the leverage we need to cash in the *ceteris paribus* clauses in the laws that bridge the gap between mental and physical properties — laws which include definitions of mental properties in terms of physical ones, causal relations among mental properties, and historical accounts of how mental properties became *worthwhile* in the first place. This allows us to build a new kind of reductionism, which preserves the causal autonomy of the mental.

I have told a story. Have I made an argument? I'm not sure. This essay contains little in the way of genuinely formal proof. I hope that further reflection by myself and critique by others will help to refine the idea of global explanatory reduction — or whatever we want to call it — to the level of rigor to which the philosophical community is accustomed. Then, we can truly see whether or not it gives us reduction with autonomy (or, if you prefer, nonreductive materialism).

References

Antony, L. M., and J. Levine, "Reduction with Autonomy." In [Tomberlin 1997].

Davidson, D., "Mental Events," 1970. (Secondary citation)

Dennett, D. C., *Consciousness Explained*. Back Bay Books, Boston, 1991.

Dennett, D. C., *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. Touchstone, NY, 1995.

Heil, J., and A. Mele ed., *Mental Causation*. Clarendon Press, Oxford, 1993.

Jackson, F., and P. Pettit, "Functionalism and Broad Content." In *Mind* 97, Oxford University Press, 1988.

Kim, J., *Supervenience and Mind*. Cambridge University Press, 1993.

Tomberlin, J. E. ed., *Philosophical Perspectives 11: Mind, Causation, and World*. Blackwell Publishers, Boston, 1997.