

CAUSALITY, MIND, AND FREE WILL

Timothy O'Connor

Indiana University

Whatever the totality of our nature might be, we human beings have bodies that situate us in a physical space. Many of our actions are at least partially constituted by causally connected sequences of events within such bodies. Do these evident facts constrain, on purely conceptual grounds, the account we advance concerning the metaphysical nature of our minds, from which our actions spring?

One familiar affirmative answer to this question holds that these facts suffice to entail that Descartes' picture of the human mind must be mistaken. On Descartes' view, our mind or soul (the only essential part of ourselves) has no spatial location. Yet it directly interacts with but one physical object, the brain of that body with which it is, 'as it were, intermingled,' so as to 'form one unit.' The radical disparity posited between a nonspatial mind, whose intentional and conscious properties are had by no physical object, and a spatial body, all of whose properties are had by no mind, has prompted some to conclude that, pace Descartes, causal interaction between the two is impossible. Jaegwon Kim has recently given a new twist to this old line of thought.⁽¹⁾ In the present essay, I will use Kim's argument as a springboard for motivating my own favored picture of the metaphysics of mind and body and then discussing how an often vilified account of freedom of the will may be realized within it.

I Kim's Argument

Kim contends that the existence of a spatial framework, or something strongly analogous to such a framework, is a necessary condition on causal interaction among objects. He supports this thesis by an analysis of the 'pairing problem,' which invites us to give a

principled way of identifying individual causal relationships in a scenario in which parallel sequences occur. Suppose rifles A and B are fired simultaneously and result in the simultaneous deaths of Andy and Buddy. Kim asks two questions (which he does not clearly distinguish): (1) What makes it true that A's firing killed Andy, and not the other way around? (2) What principles or criteria would lead us to correctly pair individual causes with their effects? Kim answers both these questions in terms of the spatial relations among the rifles and the two individuals. Rifle A's distance from, and orientation in relation to, Andy, is an important general feature of the situation that both made it possible for A to result in Andy's death and gives us a principled reason for supposing these two items to be causally paired, rather than A with Buddy and B with Andy.

I think Kim's treating these two questions as inextricably linked is unwise. He gives no reason to suppose that there could not be a situation in which objects are so distributed that their patterns of causal interactions exhibit a deep symmetry, making it impossible to decide the true causal pairings on empirical grounds. It is instead the first of Kim's questions, concerning the truthmaker for individual causal pairings, that interests me. Kim is implicitly supposing that causality necessarily exhibits a kind of generality. If an object or system A acts on B at time t, this will be due to general characteristics of B: its being the right sort of thing for A to act upon and its being in the right relationship to A at the time. Had it been a different object C that had those characteristics at t (both intrinsic and in relation to A), then A would have acted on it instead. Causality, we might say, is non-haecceitistic: objects do not have a primitive disposition to act on certain other individual objects; they are instead disposed to act on any objects having the right characteristics. At different times, the same object will achieve much the same effect on different tokens of some general type.

I accept this thesis about causal generality. (2) Kim wields it against Cartesian dualism as follows: we can imagine two nonphysical minds with identical intrinsic states at time t₀. Yet one acts on body B₁ and the other acts on body B₂. Why? Given that the minds bear no spatial relationships with these bodies, we must find some other kind of external

relationship that explains the causal selectivity. Kim can think of none and concludes that there probably couldn't be one. So if causation cannot be haecceitistic, selective dualistic interaction (one mind with one body, and vice versa) appears to be impossible. Note that an appeal to God as the one who ordains certain permanent mind-body pairings will not help. For just as God cannot make a round square, likewise God cannot create a haecceitistic model of an essentially non-haecceitistic make.

Kim goes further. He suggests that the same consideration should lead us to suppose that causal interaction even solely among nonphysical minds is probably incoherent. For what kind of relationship could play the role of a structuring environment that is played by space for physical objects? If none is proffered, and we embrace the Eleatic linkage of existence with causal powers, we should suspect the coherence of the very idea of a nonphysical mind altogether.

II A Souler System?

It will be profitable to begin our assessment of the options for dualistic causality with a quick look at this last, fanciful scenario of a monistic system of interacting souls. Kim rather quickly places it outside the bounds of intelligibility, due to the lack of spatial relations that could structure the conditions of selective interaction. But there are other forms of order. One can readily imagine a scenario in which such an alternative to spatial relationships serves to structure the interactions of nonphysical minds, provided they have a suitably rich psychology. Suppose that God, in generating a series of souls, ordains that in their initial state they conform to a mathematically describable array, with each soul carrying the information of its present 'location' in the array as a primitive intentional state. Souls 'move' through the array over time by forming intentions to occupy a specified location. Among the basic laws of this souler system is a dynamical one that governs the actual rearrangements as a function of all such intentions. (Perhaps, analogous to a time-sharing condominium arrangement, souls continually form ordered preferences as to their

subsequent location. The dynamical law might factor in previous success in obtaining highly-ranked preferences, give a certain weight to preferences to remain in one's present 'neighborhood', and so forth. The reader is invited to fill in the details as he wishes.) Causal capacities come in two basic types: the ability to form specific intentions concerning oneself or another and the ability to modify the intentional state of another via one's own intentions. I shall say more about the basic ability to generate intentions later. Let us concentrate now on the effect of such intentions on one's fellows. An example might be this: by intending to communicate to Jaegwon the thought that it would be nice to have a body as humans do, I cause him to register this thought, along with a belief that it is my thought that he is now entertaining. Again, there will be some sort of dynamical law that governs the degree of 'success' in bringing about such states in others: perhaps it will be directly proportional to the recipient's attentiveness and inversely proportional to his present informational load and 'distance' in the array. Perhaps instead of a function dictating continuously diminishing clarity and accuracy in the reception of the thought, these features of the effect will be measurable in discrete quanta of only a few magnitudes.

It will be noticed that I have used spatial metaphors to characterize the ordering that structures the interaction of souls in a space-less world. This should not in itself be objectionable, however. It is analogous to the nonliteral talk of phase space in quantum mechanics. We find it natural and easy to encode information in spatial terms. Propositional logic can be given a spatially encoded formalization, but we can do so without supposing that the logical relationships so represented are actually spatial. One might, though, shape the worry about reliance on spatial metaphor into the following objection:

You have not actually described a framework of objective external relations. Instead, you have merely gestured at an abstract formalism and asserted without argument that there could be a kind of external relation so characterizable that is distinct from spatial relations and holds among nonphysical minds. Your reference to divine decree in

instituting the array seems ineliminable, and in consequence the changing sequence over time that you described smacks of occasionalism, rather than real interaction.

In reply, I grant that a mathematical characterization of an objective ordering does not disclose the qualitative character of the ordering it is meant to describe, in the way that we ordinarily suppose ourselves to directly apprehend suitable instances of spatial relations. The only such external relations among concrete objects that we do seem to apprehend in this way - setting aside the contentious relation of causation itself - are spatial and temporal. Nonetheless, sketching the picture as I did above encourages the thought that there might be other possible instances of such mathematical structures. (Indeed, if one accepts relativity theory as a straightforward guide to the metaphysics of spacetime, one is committed to denying, contrary to appearances, that space and time are objective relations and accepting instead an underlying reality -- the spacetime interval --that is not directly apprehended in experience.)

If necessary, we might mount a further defense of the possibility of soul-soul interactions that exploits the fact that it is easier for us to accept the existence of intrinsic properties that are alien to our world. (Why? Probably because we are committed to the existence of a wider range of such properties together with the fact that we can functionally specify properties in terms of their causal role within a system, whereas external relations merely provide a background, or structured framework, within which properties manifest their dispositional character.) So suppose one remains skeptical of the very possibility of external relations within a system of nonphysical souls. One could recast my description of such a system by eschewing external relations within an array in favor of a primitive sort of intrinsic informational state had by each soul, such that it knows 'where' it is 'in relation to' all the others. (Here, of course, the relational talk is merely metaphorical for a primitive, quantifiable, and intrinsic feature.) Whether a given soul may act upon another will depend in part on these informational states, along with their other intrinsic properties. In such a

scenario, all causal interactions would be a function entirely of intrinsic properties, without reference to any structuring external relations. But I do not see that it can be dismissed on that account. For the necessary role that external relations play in our world's physical transactions -- providing an objective structuring of objects that allows for completely general dispositional tendencies to work selectively from context to context -- is carried out in the envisioned scenario.

III The Trouble with Cartesian Interactionism

Can the strategy just employed on behalf the coherence of a parliament of souls be adapted to the picture of the Cartesian dualist? It seems not. We should require not just the system of ordered relations among the nonphysical souls, along with the system of spatial relations among the physical objects, but also a cross-grid mapping of the two, identifying in general terms which body will impact which mind, and vice versa, in terms of their locations within their respective systems. And the trouble here is that on the Cartesian picture, we continually have the same pairings of individual souls and bodies, despite constant relational changes on at least the physical space side of the duality. The Cartesian picture of causal interaction seems unrepentantly haecceitistic.

(Is there the barest of possibilities in the following scenario? Suppose a two-dimensional mind-body array, involving external relations on both sides -- not including spatiality -- in which individual minds and bodies never in fact change locations. In consequence, they always act on the same object cross gridwise. Meanwhile, the bodies are acting and acted upon by constantly changing physical objects as their spatial relations change. We may suppose the souls lack potentiality for soul-soul interaction. In principle -- at least by the power of God -- souls and bodies could be reconfigured within the two-dimensional array, coming to act on different objects of the other category. If this scenario were coherent, it would involve the most exquisitely small distance from the objectionable idea of haecceitistic causality. Whether truly possible or not, for it to be true of our world,

we must assume a system of physical relations entirely hidden from ordinary observation and indeed irrelevant to body-body interaction. I judge this sufficiently high a price to motivate the alternative presented in the sequel.)

What the dualist needs, as even Descartes saw but failed to provide, is a metaphysics on which the mind and body constitute a unified natural system. We want a plausible picture on which a particular mind and body are not independent objects that somehow continually find one another in the crowd of similar such objects, but instead constitute a unified single system whose union is grounded independently of particular mental-physical interactions. For note that Kim's argument does not anywhere address self-causality -- a single object or natural system's acting upon itself. The problem of generality does not sensibly arise in this context, apart from the easily satisfied requirement that if a given system has the propensity to act upon itself in a certain manner, a similar propensity should be had by a qualitatively identical system.(3)

IV Mind and Emergence

Here is a way individual souls and bodies might constitute a single natural system. At some specific juncture in the development of the human organism, the body generates the soul, a nonphysical substance. Provided the requisite degree of structural complexity and life-conserving functions of the body are preserved, the soul will likewise persist. Thus, it is completely dependent on the body not just for its coming to be but also for its continuing to be. Given such a baseline, asymmetrical dependency-of-existence relation, it is not arbitrary that these two entities should also interact continuously in more specific ways over time. On this picture, the soul is not entirely an entity in its own right, but is more properly seen as an aspect of the overall, fundamentally biological system that is the human person.

While I do think this emergentist variety of substance dualism is able to overcome Kim's objection to the traditional variety, the kind of causal capacity it attributes to the requisite biological systems is extraordinary. Differentiating details aside, causal agents in

the universe of every sort are taken to act by introducing a qualitative change (or sustaining a persisting qualitative state) within themselves or other entities. The present sort of emergence, by contrast, would involve the generation of fundamentally new substance in the world -- amounting to creation *ex nihilo*. That's a lot to swallow. Note that it's not sufficient, for addressing Kim's problem, that one retreat to supposing a pre-established harmony. For then the apparent dependence of the soul on the body is not real, they do not constitute a single natural system, and the pairing problem is not solved.

I suggest instead that those of us with dualist predilections try to live with a weaker form of dualism, on which token mental events are ontologically *sui generis*, distinct from any complex token physical state, without there being any substance distinct from the body which is the direct bearer of those events. This is a substance monism on which human persons are fundamental biological entities that also have emergent mental states. In the present section, I will sketch in formal terms the notion of emergence I have in mind.(4) In the remainder of the paper, I will address one reason some traditional dualists are dissatisfied with the weaker form of emergence: its implications for freedom of the will.

The informed reader is admonished that what we're after from an account of emergence in the present context is quite different from other, epistemologically-rooted conceptions of emergence employed in some contemporary theories of mind in philosophy and cognitive science. Our notion is ontological. We shall say that a state of an object is emergent if it instantiates one or more simple, or nonstructural, properties and is a causal consequence of the object's exhibiting some general type of complex configuration (whose complexity will probably be a feature of both its intrinsic and functional structure). By calling a property 'nonstructural', I mean that its instantiation does not even partly consist in the instantiation of a plurality of more basic properties. By calling the emergent state a 'causal consequence' of the object's complex configuration, I mean this: in addition to having a locally determinative influence in the manner characterized by physical science, fundamental particles or systems also naturally tend (in any context) toward the generation of such an emergent state. But

their doing so is not discernible in contexts not exhibiting the requisite macro-complexity, because each such tending on its own is 'incomplete.' It takes the right threshold degree of complexity for those tendencies, present in each micro-particle, to jointly achieve their characteristic effect, which is the generation of a specific type of holistic state.

So far I have given only a sufficient condition for a state's being emergent. The reason is that the picture becomes more complicated once we consider not just the generation of an emergent state, but the kinematics of an object's having one or more emergent features for a period of time. Think of the above as a baseline case, involving just such an initial generation of an emergent state. then consider that, as a fundamentally new kind of feature, it will confer certain causal capacities on the object that go beyond even the summation of capacities directly conferred by the object's microstructure. Its effects might include directly determining aspects of the microphysical structure of the object as well as generating other emergent states. In setting forth a general account of how this might go, I am guided not by abstract intuition about how it must go in any possible emergent scenario, but about how it is plausible to suppose it goes with respect to our own mental life, on the supposition that qualitative and intentional features of our mental states are emergent.

On that supposition, it is plausible that there are enduring baseline mental states that partially underwrite more specific and often momentary mental states. Suppose, then, that when a neurophysiological system H comes to have a certain kind of complex configuration P* at time t0, the baseline emergent state E is the direct result at t1. (P*, of course, will have to be of a sufficiently general type as to persist through constant and over time dramatic change.) P* will also partly determine the underlying physical state of H at time t1. Let P0 be the remaining aspect of H's intrinsic state at t0, and P@ be the summation of those factors in H's immediate environment that will bear upon the physical state of H at t1. Letting "—>" represent the causal relation, we have

P* at t0 —> E at t1

and

$$\underline{P^*} + \underline{P0} + \underline{P@} \text{ at } \underline{t0} \longrightarrow \underline{P^*} + \underline{P1} \text{ at } \underline{t1}$$

(the latter conjunction being the total intrinsic physical state of H at time t1, with P1 being the remainder beyond P*). Now E at t1 will help to determine in part the physical state of H at the subsequent moment, t2, but not its continuing to exhibit P*, of course, as that would involve causal circularity. E, we may suppose, will also help to determine the occurrence at t2 of another emergent state, E2. Diagrammatically, the overall picture is this:

time t0	time t1	time t2	time t3	time t4
			E2	E3
		E	E	E
P*	P*	P*	P*	P*
P0	P1	P2	P3	P4
P@	P@	P@	P@	P@

(For simplicity of representation, I'm treating 'P@' schematically; at each moment it represents the sum total of those immediate environmental factors bearing on the intrinsic state of H at the subsequent moment.)

We are now in position to answer two standardly asked questions about any doctrine of mental causation. First, do the emergent properties of H supervene on its physical properties? By 'supervenience,' I here mean a synchronic relation between families of properties. The family of emergent properties would supervene on the family of physical properties just in case having an emergent property implies, of causal necessity, (1) that an object has some physical properties and (2) that its having any specific set of physical properties suffices to determine which, if any, emergent properties it has.

The first condition on supervenience is evidently satisfied. The slogan used to capture the second condition is: 'No mental difference without a physical difference.' Consider first the status of our baseline emergent feature E, with reference to times t0 and t1 in the

diagram. E is absent at t0 and present immediately thereafter. The underlying physical properties are different, too, but that is not the reason for the difference in emergent properties. For the differentiating factors (P0, P1, and the variable P@) are, by hypothesis, not directly relevant to the occurrence of E. P* alone is so relevant. Yet E is absent at the first time, since P*'s obtaining at t0 causally determines not what will occur at that very time, but immediately thereafter. So at the first instant of its instantiation in H, H will not bear E. This indicates that there might be two objects having identical intrinsic physical properties (including P*) and existing in the same external circumstance, yet one has E and the other lacks it.

But this is only a slight departure, restricted to the first instant at which the 'base' property P* occurs. More interesting divergence between emergent properties in the face of physical similarity can be seen when we turn from the baseline emergent property E to the more specific features E2, E3, and E4. Consider E2 at time t2. You might have the underlying physical properties P* and P2 without having had E2. For E2 is a causal product of the immediately prior state of H at t1 (comprised of P*, P1, and E). This prior state presumably could have been different (such that E2 would not then occur at t2), consistent with the same physical state at t2, given a suitably fortuitous change in the environmental circumstance P@. The possibility of a difference with respect to E2 without an underlying physical difference is clearer when we instead suppose a case in which the causal connections are probabilistic only. For then we can hold fixed the immediately prior state of H, and suppose a scenario in which it causes the same underlying physical state at t2, but, owing to a different chancy outcome, causes the occurrence not of E2, but of some distinct property, E*2, at the emergent level.

So emergent states do not in general supervene on physical states. A second question we might have is whether they are epiphenomenal, at least with respect to the purely physical states of H and its immediate environment. Is the system in its purely physical aspect – is physics more generally -- causally closed? Here, too, the answer is No. P3's obtaining at t3

is in part a product of E and E2's obtaining at t2. Had one or both of these failed to obtain at that previous time, something other than P3 would have occurred subsequently. Consistent with this, it is true in an emergentist scenario that everything that occurs rests on the total potentialities of the physical properties. For the occurrence of any emergent properties are among those potentialities, and so the effects of the emergent features are indirectly a consequence of the physical properties, too. The difference that emergence makes is that what happens transcends the immediate, or local, interactions of the microphysics.

In summation, we have seen that property emergentism allows for a form of dualism that escapes Kim's problem, since the mental-physical interactions it posits occur within the context of a natural unitary system, and hence is a form of self causality, rather than multiple-object transaction. I now turn to one test of its being sufficiently robust: whether it is consistent with the kind of freedom of will to which many dualists subscribe.

V Causal Generality and Free Will

Freedom of the will, in my judgment, involves the exercise of a distinctively personal form of causality, one which differs in certain respects from the mechanistic form of causation operative in impersonal causal forces.(5) In the mechanistic case, objects have specific causal powers, or dispositional tendencies, associated with their fundamental intrinsic properties. The powers might concern a unique outcome or a range of possible effects that is structured by a specific probability measure. Either way, they exercise certain of these causal powers as a matter of course when they are placed in the appropriate circumstances. Such circumstances either stimulate a latent mechanism or remove inhibitors to the activity of a mechanism already in a state of readiness. Strictly speaking, the cause here is the event of the object's having these power-conferring properties in those circumstances.

According to some of us, there is another species of the causal genus, involving the characteristic activity of purposive free agents. Such agents can represent possible courses of action to themselves and have desires and beliefs concerning those alternatives. Against

that background motivational framework, they themselves directly bring about immediately executive states of intention to act in various ways. This direct causing by agents of states of intention goes like this: As with mechanistic causes, the distinctive capacities of agent causes ('active powers') are grounded in some set of properties. So any agent having the relevant internal properties will have it directly within his power to cause any of a range of states of intention delimited by internal and external circumstances. However, these properties function differently in the associated causal process. Instead of being associated with direct causal functions from circumstances to effects, they (in conjunction with appropriate circumstances) make possible the agent's producing an effect. These choice-enabling properties ground a different type of causal power or capacity -- one that in suitable circumstances is freely exercised by the agent himself. (One might say that agent causation is essentially purposive, whereas mechanistic causality is not. Agent causality is triadic -- it involves an agent's causing an intention for a reason. The dyadic form of mechanistic causality is indifferent to whether the causes be reasons or impersonal states.)

Now given his concern with causal generality, Kim might wonder whether I am trying to have things both ways by embracing the causal powers account of mechanistic causation while also defending agent causation. The causal powers account is resolutely 'anti-singularist,' in the sense discussed above. Yet I have purported to identify as one of its species something (involving agent causation) that seems to imply singularism. But there is not really a problem here, except perhaps with the way some draw the singularist/ anti-singularist distinction. Mechanistic and agent causation both require generality with respect to the grounding of causal powers. A given particular has a given type of causal power because of its intrinsic properties, and properties are universals. Where agent and mechanistic causal capacities diverge is in their exercise. The exercise of mechanistic capacities conforms to tendencies of some measure (the limiting case being deterministic). In the agent causal case this is not necessarily so, at least as a conceptual matter. **(6)** Agents may choose any of the options within the range of their power at a given time without

having some fixed probabilistic tendency to do so. (There are some contingent features of human agents that indicate that the exercise of active power has further causal structure in the way reasons govern it. Reasons move us to act, and some do so much more strongly than others. My current proposal, elaborated elsewhere, is that we think of the agent's states of having reasons to act in various ways as structuring the agent causal capacity, such that the agent's freely choosing an action type will have some objective tendency to occur, one which fluctuates over time. Even if this is accepted, it remains true that, in contrast to mechanistic causation, it remains up to the agent to decide how to act. The tendency-conferring state of having a reason does not itself generate the action; it disposes the agent himself to initiate an action sequence.) Whether or not we add this wrinkle, the important point here is that the basic view does not fall afoul of Kim's stricture on non-haecceitistic causal tendencies, since agent causation is grounded in generalized capacities and is a form of self causality within a unitary system.

Can we make sense of agent causation as an emergent capacity of a fundamentally biological system? Note that such a theorist is committed to the emergence of a very different sort of property altogether. Instead of producing certain effects in the appropriate circumstances itself, of necessity, this property enables the individual that has it in a certain range of circumstances to freely and directly bring about (or not bring about) any of a range of effects. It might be thought that because of this distinctive character, it isn't possible that it could naturally emerge from other properties. Such a property could be instantiated only in a very different kind of substance from material substances, as on the problematic Cartesian view.

This thought does not bear well under scrutiny, however. Given that there is nothing inconsistent about the emergence of an "ordinary" causal property, able to causally influence the environments in which it is instantiated, it is hard to see just why there could not be a variety of emergent property whose novelty consists in enabling its possessor directly to effect changes at will (within a narrowly limited range, and in appropriate

circumstances). If properties are able, as a matter of nomological necessity, to produce an entirely novel type of property, what reason do we have to assert that, when it comes to the property-kind distinction just noted, properties can spawn others of their own kind alone? At least, this would seem to be an empirical, not philosophical or conceptual, matter.

Still, taking the agency theory seriously within an emergentist framework raises several theoretical problems. The most fundamental of these is determining the precise underlying properties on which an agent-causal capacity depends. Put differently, what types of features - either functionally or intrinsically characterized - constitute a physical system's being a free agent in the technical sense? Conversely, what structural transformations in the human nervous system would result in long-standing (or permanent) loss of the agent-causal capacity generally? This is an empirical matter (one answerable only by neurobiological science), and not in the province of philosophical action theory. Yet even a casual acquaintance with how neurobiologists approach their craft is enough to give an appreciation of the enormous difficulty this most basic issue poses. A plausible general conjecture is that such a capacity is bound up with the capacities for action representation and for conscious awareness, in their specifically human (and probably certain other mammalian) manifestations. It is highly plausible that this self-determining capacity strictly requires each of these other abilities, as they appear to follow from the very characterization of active power as structured by motivating reasons and as allowing the free formation of executive states of intention in accordance with an action representation.

If there are agent-causal events, there is no neat and simple way of dividing those events from mechanistic-causal ones. It surely must be allowed that some human behavior, even conscious behavior, is directly brought about by mechanistic-causal factors. (Not all action is free action, and I intend agent causation to account only for the latter.) This is likely to be true of behavior governed by unconscious factors and highly routinized actions. Precisely to what extent, then, is an ordinary human's behavior directly regulated by the agent himself, and to what extent is it controlled by sub-personal processes? Even when I act freely, I am

usually not trying to control directly the precise degree of muscle contraction, limb trajectory, and so forth. This makes it plausible to hold that our memory system stores action sequences that we simply activate through conscious choice. (It also explains the facility of an experienced performer in carrying out complex movements, such as a sequence of dance steps.) (7) It is likely that these choices are at times brought about event-causally, while we simply monitor the result and retain the capacity to agent-causally redirect things as need be.

Finally, one might also worry that free will requires the emergence of a degree of indeterminism far beyond what we have any reason to believe is operative (as a function of quantum indeterminacy) at the complex level of neural structures. My reply is that since an emergent property has, relative to its underlying properties, a unique, nonstructural nature, we have no a priori reason to think it must result in processes exhibiting precisely the same degree of indeterminism as is present in its sustaining lower-level processes. Still, we are not supposing 'something's coming from nothing,' as many have thought: the presence of any emergent, on the view I have sketched, will be determined by more fundamental features of its possessor. What it does allow is a stable set of processes giving rise, at certain critical junctures, to a somewhat different order of affairs via 'top-down' controlling features. It is just this possibility that allows the right sort of emergentist view to overcome the opposite complaint from Cartesian sympathizers that agents with such emergent capacities are 'ontologically superficial' -- not among the truly basic entities whose activities determine the way the world is. (8) While it is true, on my picture, that the presence of agent-causal capacities in select complex entities has always been among the potentialities of the world's primordial building blocks, the way those potentialities are exercised is not so prefigured. The agents themselves determine these outcomes. In consequence, any way of completely characterizing what happens in the world must make reference to these agents and their distinctive capacities. This is as ontologically 'deep' as any entity that is not necessary being could aspire to. (9)

References

- Armstrong, D. (1997) A World of States of Affairs. Cambridge: Cambridge University Press.
- Farrer, A. (1958) The Freedom of the Will. London: Adam & Charles Black.
- Kim, J. (2001) "Dualism, Causality and Being a Person," this volume.
- Kosslyn, S. and Koenig, O. (1992) Wet Mind. New York: Free Press.
- Norman, Donald and Shallice, Tim (2000) "Attention to Action: Willed and Automatic Control of Behavior," in Michael Gazzaniga, ed., Cognitive Neuroscience: A Reader. Oxford: Blackwell.
- O'Connor, T. (2000) Persons and Causes: The Metaphysics of Free Will. New York: Oxford University Press.
- Tooley, M. (1987) Causation: A Realist Approach. Oxford: Clarendon Press.

NOTES

1. "Dualism, Causality and Being a Person," this volume.
2. It is denied by 'singularists' who hold that causation is first and foremost a relation between particular events, and the holding of that relation in a given instance has no implications (strict or probabilistic) for what happens elsewhere or elsewhen. C might cause E even though no other C-type event causes or has a tendency to cause an E-type event. It may happen to be true of our world that all causal transactions fall into patterns of certain types, but this, for the singularist, is at best owing to some contingent feature of the way causation is manifested in our world. There might have been ubiquitous causation in a chaotic, anomic world -- including causal patterns between pairs of objects that are not amenable to general analysis of the sort Kim requires.

The implausibility of this view is apparent when one considers the ultimate positions on the nature of causation in the actual world held by the two contemporary philosophers who have most emphatically argued that causation is a singular relation -- positions which look decidedly anti-singularist to the causal observer. David Armstrong has responded to a serious problem for earlier versions of his position by holding that causation is manifested in our world as a relation among types of states of affairs. (See A World of States of Affairs Cambridge: Cambridge University Press, 1997.) And Michael Tooley's 'speculative proposal' in response to that same problem is to posit unusual features in the mereology of transcendent universals. If it is a law of nature that all things having property P have property Q, then, he says, we might suppose that P "exists only as part of the conjunctive universal, P and Q." (Causation: A Realist Approach Oxford: Clarendon Press, 1987, p.124). It would then follow that any time P is instantiated, Q is as well. Although these moves are critical to salvaging their respective theories of the general nature of causal processes in our world, they insist that these are merely contingent facts about causation, so as to preserve the singularist 'intuition' that there might have been instances of causation in an anomic world -- in a world where events having the effects they do has nothing to do with the kind of events they are. This contingency is mysterious and unmotivated.

3. I note in passing that one might suppose that the problem presently under consideration applies equally to the classical conception of God as a nonphysical mind who causally acts upon the physical universe. But this would be a mistake. The physical universe is not a pre-given object which, as it happens, God encounters, as on the Cartesian picture the soul is an originally independent substance in its own right and is then 'fitted' to a specific body.

Instead, God's acting on the universe in particular ways at particular times is of a piece with His giving being to the universe at that time: the universe is entirely dependent on Him at all times for its very existence and character, and there is no neutral environment akin to space in which He and it and possibly other entities co-exist, which might prompt the question of the general conditions which govern selective interaction among them.

4. I explain this less formally and address objections not considered here in chapter 6 of my Persons and Causes: The Metaphysics of Free Will (New York: Oxford University Press, 2000).
5. I have argued this in several places, most recently in the book cited in the previous note. (Much of the present section of this paper excerpts material from that work.) When I say that agent causation is "purposive," and thereby contrasts with "mechanistic" causation, I don't mean to imply that it is not law-governed in any way. (See the following note.) Nor do I mean to imply that were reasons directly to cause choices, the activity would be nonpurposive in the ordinary sense. Rather, I intend to highlight the fact that agent causation is essentially purposive, whereas mechanistic causality is not. Agent causality is triadic -- it involves an agent's causing an intention for a reason. The dyadic form of mechanistic causality is indifferent to whether the causes be reasons or impersonal states.
6. There are some contingent features of human agents that indicate that the exercise of active power has further causal structure in the way reasons govern it. Reasons move us to act, and some do so much more strongly than others. In ch.5 of my book, I propose that we think of the agent's states of having reasons to act in various ways as structuring the agent causal capacity, such that the agent's freely choosing an action type will have some objective tendency to occur, one which fluctuates over time. Even if this is accepted, it remains true that, in contrast to mechanistic causation, it remains up to the agent to decide how to act. The tendency-conferring state of having a reason does not itself generate the action; it disposes the agent himself to initiate an action sequence.
7. An early philosophical discussion of the implications of this for theories of action is A. Farrer's The Freedom of the Will (London: Adam & Charles Black, 1958). Concrete proposals for how to account for such phenomena within recent cognitive science models may be found in S. Kosslyn and O. Koenig in Wet Mind (New York: Free Press, 1992) and Donald Norman and Tim Shallice, "Attention to Action: Willed and Automatic Control

of Behavior," in Michael Gazzaniga, ed., Cognitive Neuroscience: A Reader (Oxford: Blackwell, 2000), pp. 376-390.

8. This phrase is taken from some recent unpublished work by Peter Unger.

9. A version of this paper was read at the University of North Carolina and Davidson College. For helpful comments, I thank Louise Antony, John Carroll, John Heil, William Lycan, Al Mele, David Robb, Dan Ryder, and Peter Unger.