

LOCKE'S EXCLUSION ARGUMENT

Walter Ott

1. THE EXCLUSION PROBLEM

Jaegwon Kim's exclusion argument threatens to make mental states and events¹ epiphenomenal.² Take physical event P, say, the stimulation of a certain set of neurons, and mental event M, the volition to move one's hand. Now consider P*, the movement of one's hand. Both P and M compete for the title "cause of P*." But assuming the closure of the physical world and the impossibility of overdetermination, P must win.³ P* has a complete and sufficient physical cause; in our example, this either is or includes P. (It might also include other physical background conditions.) So what work is left for M to do? M seems to be a mere side effect of P, with no causal powers of its own. And even if we allow causal overdetermination, something of the same problem remains: the physical world would still go its own way, even if mental events were removed.

This argument at most threatens those materialist views that take the relation between M and P to be one of supervenience.⁴ An identity theorist has no problem here, since M and P just are the same event, described in different ways.⁵ And a dualist (whether property- or substance-) might be willing to bite one of the many bullets on offer, by, say, denying that P* has a sufficient physical cause.⁶

That John Locke might have offered anything like the exclusion argument is at first sight implausible. Locke steers clear of metaphysical conclusions about the mind. Although friendly to dualism, he thinks that "[a]ll the great Ends of Morality and Religion, are well enough secured, without philosophical Proofs of the Soul's Immateriality" (IV.iii.6, 542).⁷ Locke's agnosticism about the ultimate status of the mind is well known.⁸ What is less known is that he produces an argument to show that some versions of materialism—precisely those that take mental states to depend on, while not being reducible to, material states—make mental states epiphenomenal. His agnosticism is not nearly as thoroughgoing as at first appears. What is more, a careful investigation of the argu-

ment I shall explore both supports and gains support from the proper interpretation of Locke's controversial remarks on the possibility of "superaddition."⁹

2. LOCKE'S TARGET

Taking "God" to mean an immaterial, eternal, thinking being, Locke attempts in IV.x to prove that there is indeed a single object that satisfies this description. In arguing for this first attribute—immateriality—Locke produces at least two importantly different arguments (IV.x.16 and IV.x.19). Both are of interest quite apart from their theological conclusions, since, if successful, they would also show that thinking beings in general—and not just God—cannot be *merely* material.

In his first argument, Locke offers a trilemma: the eternal, thinking material thing must be either (i) a single atom, (ii) all matter, or (iii) "*some certain System of matter duly put together*" (IV.x.16, 627). The first alternative is "absurd" since it privileges a single atom over all others, both in terms of duration and perfection. But every particle of matter "as matter" is everywhere the same.¹⁰ Locke construes the second hypothesis not in Spinozist fashion but as the idea that each atom is itself an eternal thinking thing, something his opponent is unlikely to hold. The real question, then, is about (iii). The key argument is this:

[T]o suppose the eternal thinking Being, to be nothing else but a composition of Particles of Matter, each whereof is incogitative, is to ascribe all the Wisdom and Knowledge of that eternal Being, only to the *juxta*-position of parts; than which, nothing can be more absurd. For unthinking Particles of Matter, however put together can have nothing thereby added to them, but a new relation of Position, which 'tis impossible should give thought and knowledge to them. (IV.x.16, 627)

At first glance, this looks like the compositional fallacy: this atom cannot think; therefore, no collection of atoms can think. But this is not Locke's point. Instead, he argues that if a system of (individually incogitative) particles of matter could be suitably arranged so as to produce thought, it would have to be the case that all mental properties are relations of position, since these relations are the only new properties such an arrangement could produce. And this, Locke thinks, is absurd on its face.

As in his later argument (IV.x.19), Locke's target is the hypothesis that thought arises from the arrangement and "*juxta*-position" of particles of matter, each of which is incogitative. As I shall argue in the penultimate section, this leaves open the possibility that a material thing contains some set of nonmaterial properties in virtue of which it thinks. If this

were the case, the explanation of thought would have nothing, or very little, to do with the arrangement of the material parts; it would depend entirely on this mysterious, extra property. By contrast, the present argument is designed to rule out one version of what we might usefully call the (or, more accurately, *a*) supervenience hypothesis: "(SH): thought arises in a body by virtue of the arrangement and juxtaposition of its (individually incogitative) material parts."

It is tempting to cry "anachronism" at this point. Many will balk at my introduction of a notion of such recent vintage as supervenience. There are two things to be said here. First, although the term is indeed recently coined, the corresponding notion need not be. Indeed, it seems a natural way to read Locke's talk of thought's arising from the juxtaposition of bits of matter. Second, the only alternative here seems to be to read Locke as using such language to refer to the identity theory. But as we shall see, if *this* is Locke's target in IV.x.19, his argument simply makes no sense.

What are we to make of this first argument (IV.x.16) against SH? Locke is clearly playing upon the same intuition he attempts to educe in IV.x.10, where he writes, "Matter, *incogitative Matter* and Motion, whatever changes it might produce of Figure and Bulk, *could never produce thought*. . . [The minute parts of matter] knock, impell, and resist one another just as the greater do, and that is all they can do" (IV.x.10, 623–24). Jonathan Bennett rightly claims that this sort of argument is no better—but, one might add, no worse—than any other of the intuition pumps deployed against all forms of materialism, from Gottfried Leibniz's mill to John Searle's Chinese room.¹¹ But the argument in section 16 adds something to this earlier argument. Here, the question is not what matter can do but what kinds of properties it can possess. If "a new relation of Position" is the only property that can accrue to any "*juxta-position*" of matter, then, no matter how matter is arranged, it will never admit of mental properties. Locke's thought here seems to be this: mental properties are monadic, but all properties possessed by arrangements of matter are relations, and hence polyadic; thus, mental properties cannot obtain in virtue of the juxtaposition of material parts. Among the many problems this argument will face is its assumption that mental properties are monadic. One does not need to be an externalist about mental content to believe that at least some mental states, namely, those with intentional contents, are in the typical case relational.

Whatever we make of these issues, however, the argument in IV.x.16, while not a straightforward fallacy, is less than convincing, and not just from a contemporary perspective. The argument seems to be that there

is no way to bridge the gap between thought and relations of position; that there is such a gap is fairly clear, but whether it is metaphysical or merely epistemic can only be decided by further argument.

3. LOCKE'S EXCLUSION ARGUMENT

This further argument is provided three sections later. I shall argue that in IV.x.19, Locke again challenges SH. But the argument does not appeal to the alleged impossibility of thought's arising from relations of position among individual bits of matter. Instead, the problem is that SH entails that there are no voluntary actions, a dire consequence on almost anyone's view.

The immediate point of section 19 is to parry the argument that matter must be coeternal with God, since God cannot create it out of nothing. Locke grants that we cannot conceive how God creates matter but, as usual, argues that this is just one more thing that lies outside of our cognitive grasp.

[I]t is not reasonable to deny the power of an infinite Being, because we cannot comprehend its Operations. . . . We cannot conceive how anything but impulse of Body can move Body; and yet that is not a Reason sufficient to make us deny it possible, against the constant Experience, we have of it in our selves, in all our voluntary Motions, which are produced in us only by the free Action or Thought of our own Minds; and *are not, nor can be the effects of* the impulse or determination of the Motion of blind Matter, in or upon our Bodies; for then it could not be in our power or choice to alter it. For example: my right Hand writes, whilst my left Hand is still: What causes rest in one, and motion in the other? Nothing but my Will, a Thought of my Mind; my Thought only changing, the right Hand rests, and the left Hand moves. This is a matter of fact, which cannot be denied: Explain this, and make it intelligible, and the next step will be to understand Creation. (IV.x.19, 629; my emphasis)

We cannot understand God's creation of matter *ex nihilo*; but this is no reason to declare it impossible since even the most quotidian of events—moving one's hand, for example—is similarly incomprehensible to us. My interest here is not in the aptness of the analogy, or even in Locke's conclusion, but rather in how Locke supports it. From an identity theorist's point of view, there is no special problem in understanding how one's hand moves; just like anything else in the universe, it is a question of bits of matter interacting. Locke resists this. There is indeed a mystery about voluntary action, namely, its source. For whatever that source is, it cannot be the motion of matter, yet impulse is the only cause of motion Locke thinks we can conceive of.

This much about the argument, then, is uncontroversial: Locke thinks that some materialist hypotheses are to be rejected because they make voluntary action impossible. And it is simply a fact of experience that there are voluntary actions.

There are at least three quite different ways to read the argument for Locke's conclusion, each with a degree of plausibility. Let us begin with the natural idea that freedom of the will is at stake. After all, Locke says that if it were not "in our power or choice to alter it," an action would not be voluntary. This is not enough to secure Locke's conclusion, of course; one has to supply a number of missing premises. This emerges if we even come close to making the argument formally valid. (As in the initial example in this paper, let M stand for the volition to move one's hand; P for the material state with which M is allegedly identical, or supervenes; and P* for the motion of one's hand.)

1. There are voluntary actions (a fact of experience).
2. Suppose M either supervenes on or is identical with P.
3. P* is voluntary ("in our power") only if we could have chosen otherwise.¹²
4. All physical events have prior physical conditions that suffice to bring them about.
5. If an action is physically determined, it is such that we could not have done otherwise.
6. P* is physically determined by P.
7. P* is such that the agent "performing" it could not have done otherwise.
8. P* is not voluntary.
9. Contradiction (1, 8)

∴ At least some mental states, namely, volitions, are neither identical with nor supervene on physical states.

Unfortunately, this argument requires Locke to appeal to a libertarian conception of free will, where this means a conception that allows for a strong principle of alternative possibilities. For the present argument to work, in other words, Locke must be claiming that a voluntary action is such that the agent, in precisely the same circumstances, could have performed a different action. All that Locke actually says here, however, is that if an action is voluntary, it must be "in our power or choice" to change it. I shall argue for a deflationary reading of this phrase presently. But we can already see that Locke does *not* subscribe to the principle of alternative possibilities as a criterion for an act to

be “voluntary” or “in our power.” This is made clear much earlier in the *Essay* (II.xxi):

A Tennis-ball, whether in motion by the stroke of a Racket, or lying still at rest, is not by any one taken to be a *free Agent*. If we enquire into the Reason, we shall find it is, because we conceive not a Tennis-ball to think, and consequently not to have any Volitions, or preference of Motion to rest, or *vice versâ*; and therefore has not *Liberty*, is not a free Agent; but all its both Motion and Rest, come under our *Idea of Necessary*; and are so call'd. Likewise a Man falling into the Water, (a Bridge breaking under him,) has not herein liberty, is not a free Agent. For though he has Volition, though he prefers his not falling to falling; yet the forbearance of that Motion not being in his Power, the Stop or Cessation of that Motion follows not upon his Volition; and therefore therein he is not *free* (II.xxi.10, 238).

What makes the difference between human and tennis ball is a capacity for preferring¹³ some states of affairs over others. What makes the difference between someone's falling into the water and someone's *jumping* into it is the efficaciousness of her volitions. In other words, it is the causal history of the act that determines its voluntariness. And what shows the man's continued falling to be involuntary is simply the fact that, even if he were to will otherwise, he would continue to fall. This counterfactual is simply a way of pointing to the obvious fact that his volition does not *cause* his continued falling.

This view of the voluntary is borne out by other texts as well. Locke claims that a voluntary action is one that is “consequent to” “an order or command of the mind” (II.xxi.5, 236). So, again, a motion counts as voluntary only if it is caused by an appropriate act of the mind.¹⁴ (Whether being so caused is also *sufficient* for voluntariness is a further, and hotly debated, question, which I shall not attempt to answer).¹⁵ So far, everything Locke has said is consistent with one's *will's* being determined in such a way that one's preference or “command” could not be other than it is. Note that I am not claiming that Locke is a compatibilist but only that, so far as voluntariness goes, nothing he says is at odds with the compatibilist position. Indeed, this seems to be the point of II.xxi.11, 239, where Locke argues that “*Voluntary then is not opposed to Necessary; but to Involuntary.*”

While most commentators read Locke as a compatibilist, others, notably Gideon Yaffe, disagree. I do not have the space here to settle this debate. Happily, what counts for my argument in this paper is Locke's position on the voluntary, not on free will, and even then only the necessity of the causal condition. When Locke says in IV.x.19 that we would not be able to alter an act resulting from matter in motion,

he means simply that that act is not the result of an appropriate mental state, namely, a volition, just as the motion of the falling man is involuntary because it does not result from an act of his will. If this so, Locke simply does not accept step 3, nor, as a result, the move from steps 7 to 8. Thus, the initial, "free will" version of the argument of IV.x.19 will not work. It relies on a conception of what it is for an act to be voluntary that Locke does not have. But there is further evidence that Locke cannot have any deep issues about freedom of the will in mind here. Recall that IV.x.19 takes as its datum the claim that there are voluntary actions. Now, Locke insists in II.xxi that the voluntariness of an action is necessary, but not sufficient, for that action to count as having proceeded from a free being. Indeed, this is the upshot of the famous example of the locked room (II.xxi.10, 238). "Voluntary" and "free" are not interchangeable terms.

It is nevertheless odd that Locke here (IV.x.19) speaks of "voluntary Motions, which are produced in us only by the *free* Action or Thought of our own Minds" (my emphasis). Earlier in the *Essay* (II.xxi.25, 247), he has already argued that the question whether a *volition* (as opposed to an agent) is free is nonsensical. What is more, Locke is equally clear that what determines the will to choose this or that action is the feeling of uneasiness (II.xxi.29, 249). So, acts of will themselves always have a cause; to put the matter in scholastic terms, there is always a cause that reduces the power of the will to action. I think it unlikely that Locke is reversing himself on these issues in book IV. Given all this, my deflationary reading of "in our power" seems plausible.

If this counts as progress in understanding the argument, though, it is of a curious sort. For if Locke is *not* appealing to libertarian intuitions about freedom of the will here, he is in deadly danger of begging the question. Let us move to the next natural reading of the argument, one that takes it to be directed at the identity theory. (Note that those who are wary of reading Locke as entertaining any supervenience hypothesis *must* read the argument in this way, given the failure of our previous reconstruction.)

On this construal, the argument would have to go something like this:

1. There are voluntary actions (a fact of experience).
2. M is physical state P.
3. P* is voluntary ("in our power") only if it is caused by M.
4. But given 1, P is the cause of P*, not M. Thus,
5. P* is not voluntary.

6. Contradiction (1, 4)

∴ At least some mental states, namely, volitions, are not identical with physical states.

This gets part of our story right. Premise 3 captures precisely the view of voluntariness I have attributed to Locke. Unfortunately, the argument is obviously unsound. M is, of course, the cause of P*, since M just is P. The only way I can imagine rescuing the argument would be to try to say that M is not, in fact, the cause of P*, since it is only qua P that M causes anything. But the point of identity theory is to deny that such “qua” talk imports anything metaphysical.

Locke seems to be back on the ropes. If the argument appeals to considerations of free will, it relies on dubious libertarian intuitions, intuitions Locke does not seem to share. Revised so as to avoid this appeal, it seems to be broken-backed from the start.

But if instead Locke’s target is SH, just as it was in IV.x.16, the argument snaps into place. Recall that SH, in Locke’s terms, proposes that matter in motion can be said to think in virtue of the “*juxta*-position of [its] parts.” Any mental state, then, will obtain in virtue of the arrangement and motion of bits of matter. Does this version of Locke’s argument, construed so as to exploit premise 3 of the preceding argument, fare any better than its competitors?

Let us see just what the argument would look like.

1. There are voluntary actions (assumption).
2. Suppose M supervenes on P.
3. P* is voluntary (“in our power”) only if it is caused by M.
4. Given 2, P, not M, is the cause of P*. Thus,
5. P* is not voluntary.
6. Contradiction (1, 5)

∴ At least some mental states, namely, volitions, do not supervene on physical states.¹⁶

Locke’s point, then, is that SH entails that none of our actions is voluntary. (It is, then, misleading to think of such events as *our* actions at all.) Recall the causal condition for voluntariness: to be voluntary, an action must be caused by a volition. But if volitions depend on matter-in-motion, then it is really the matter, after all, that is doing the causing. Voluntary actions “are not, nor can be the effects of the impulse or determination of the Motion of blind Matter.”

It is important to see here that Locke is exploiting 3 rather than

- (3⁾ P* is voluntary only if its immediate, proximate cause is a volition.

On this reading, the problem with the supervenience hypothesis is not that the volitional state M is not *a* cause of P* but that it is not the *proximate* cause. That is, M might cause P, and so, in turn, cause P*; nevertheless, P*'s proximate cause is not M but P. I think this cannot be Locke's argument, for the simple reason that it is hard to see how one could read the hypothesis that the volition M depends upon a material state P as the idea that M *causes* P. (If anything, the reverse is more plausible, though still false.) A further difficulty is that even a nonmaterial volition would not be the proximate cause of the voluntary action. Presumably such volitions, on Locke's view, cause actions only by setting off a complex causal chain, beginning with the activity of the brain.

At this point, we can turn to a further objection to my reading of IV.x.19. One obvious worry is that it does not cover the identity theory. The argument I have reconstructed is simply powerless against a materialism that identifies volitions and physical states. And yet Locke seems to think his argument tells against *any* materialist view. I think Locke has to plead guilty here. In his defense, it might be said that the two versions of materialism we are considering—identity theory and nonreductive physicalism—were not clearly distinguished in Locke's day and that he himself seems not to have observed any such distinction. This is true even though in IV.x we can clearly see him attacking a nonreductive position: the textual evidence strongly suggests that Locke has SH in mind in sections 16 and 19.

To sum up: let us suppose that voluntary actions are the result of matter-in-motion, in the sense specified by SH. It then follows that a voluntary action is produced not by the volition *per se* but by the material event (namely, motion) on which that volition supervenes. The problem with this is that it renders the action that is produced involuntary. For it no longer has the volition as its cause; its cause is the physical base *of* that volition. Locke's talk of the action's proceeding from "mere matter in motion" is his way of pointing to the exclusion of mental causes on the supervenience hypothesis.

4. CONSEQUENCES

If I am right, the argument of IV.x.19 is directed at SH. If all mental states supervened on physical states, then putatively voluntary actions would have those physical states as their causes, with mental states as mere by-products. SH, Locke thinks, renders all actions involuntary, hence not *actions* at all.

This leaves us with a problem. For Locke's official position is agnosticism: whether what thinks in us is material or not is "a Point . . . put out of the reach of our Knowledge" (IV.iii.6, 542). At the same time, as we have seen, Locke argues that God cannot be material in part because *nothing* that is merely matter in motion can think. So, the answer to the question is not put out of our reach after all.

There is still another problem here. For the materialist hypothesis in which Locke claims to find "no contradiction" (IV.iii.6) looks, at first glance, just to be SH. That is, one source of Locke's humility is the idea that it is at least epistemically possible that God has given some bits of matter, "fitly disposed," the power to think. Both problems are raised by the following well-known passage:

We have the *Ideas of Matter and Thinking*, but possibly shall never be able to know, whether any mere material Being thinks, or not; it being impossible for us, by the contemplation of our own *Ideas*, without revelation, to discover, whether Omnipotency has not given to some Systems of Matter fitly disposed, a power to perceive and think, or else joined and fixed to Matter so disposed, a thinking immaterial Substance: It being, in respect of our Notions, not much more remote from our Comprehension to conceive, that GOD can, if he pleases, [i] superadd to Matter a Faculty of Thinking, than that he should [ii] superadd to it another Substance, with a Faculty of Thinking; since we know not wherein Thinking consists, nor to what sort of Substances the Almighty has been pleased to give that Power, which cannot be in any created Being, but merely by the good pleasure and Bounty of the Creator. For I see no contradiction in it, that the first eternal thinking Being should, if he pleased, give to certain Systems of created senseless matter, put together as he thinks fit, some degrees of sense, perception, and thought. (IV.iii.6, 540–41)

On its face, this hardly seems consistent with IV.x.19, yet Locke goes on in IV.iii.6 to refer the reader to that very chapter. So, we cannot wash our hands of the matter simply by suggesting that Locke was careless or inconsistent. This section recommends epistemic humility; IV.x.19 is anything but humble. Is there some way to reconcile these attitudes?

I think there is, but to see it, we must first face up to the second of our two problems, namely, whether Locke is disallowing in IV.x.19 the very hypothesis whose epistemic possibility he admits in IV.iii. What, exactly, is this hypothesis? Bennett discusses the view, most closely associated with Michael Ayers, that the state of affairs Locke entertains in IV.iii is "(Naturally Thinking Matter, or 'NTM') Some matter thinks because of how it is materially organized, its thought arising naturally out of its material nature."¹⁷ Ayers draws on Locke's discussion of the

superaddition of the properties of motion, vegetation, life, and beauty. In a letter to Edward Stillingfleet, Locke writes,

The idea of matter is an extended solid substance, wherever there is such a substance, there is matter, and the essence of matter, whatever other qualities, not contained in that essence, it shall please God to superadd to it. For example, God creates an extended solid substance, without superadding anything else to it, and so we may consider it at rest: to some parts of it he superadds motion, but it still has the essence of matter: other parts of it he frames into plants, with all the excellencies of vegetation, life, and beauty, which are to be found in a rose or a peach-tree, &c. above the essence of matter in general, but it is still matter: to other parts he adds sense and spontaneous motion, and those other properties that are to be found in an elephant.¹⁸

Just as the beauty of a rose arises from matter arranged in a certain way, so might thought arise. NTM, then, just is the hypothesis of supervenience.¹⁹ Although, in principle, matter-in-motion could resolve itself into a beautiful rose, this is vanishingly improbable. Hence, even vegetation requires the activity of a divine being.

In contrast, commentators from Leibniz to Margaret Wilson have read Lockean superaddition as a *contranatural* process. NTM has it that God could organize matter in such a way that it naturally takes on the qualities of thought. But on the Leibniz's reading, God "give[s] things accidental powers detached from their natures, and thus inaccessible to reason in general." Leibniz thinks Locke has opened "a backdoor through which to re-introduce those too occult qualities, which no mind can understand," "little goblins" who "come forward like gods on the stage" and do whatever the philosopher likes.²⁰

Siding with Wilson and Leibniz against Ayers, Bennett points out that IV.x's conclusion is not merely that it is extraordinarily improbable that thought should arise from matter in motion but that it is *impossible*. If NTM were right, it would be, in principle, possible for bits of matter, knocking about for a long enough span of time, to produce thought spontaneously. But this is not Locke's conclusion. So NTM, as a reading of Lockean superaddition, lacks textual support.²¹ I think we can go further. For I think IV.x.19 actually *rules out* NTM. Whatever Locke has in mind in IV.iii, it cannot be the supervenience hypothesis. And NTM and SH are just the same idea, differently put. On Ayers's reading, in IV.iii Locke is granting that God might so arrange matter that thought naturally appears from it, in the same way that he arranges bits of matter to produce flowers and trees. On my reading, this is precisely the view Locke wishes to rule out. A naturally thinking material being would be incapable of performing voluntary actions.

This does not quite solve our second problem, since, for all I have said so far, it could be the case that Locke is simply inconsistent: IV.iii.6 allows, and IV.x.19 rejects, NTM. To see that this is not so, we need to have a closer look at IV.iii.6. In the quotation above, I flagged two distinct views with lower case roman numerals. The first of these is rightly termed the hypothesis of “thinking matter”; the second is simply the substance-dualist hypothesis Locke thinks is (probably) true. Locke’s formulation of (ii) has, quite naturally, been of less interest than (i). But the precise way in which he formulates (ii) can tell us much about what precisely (i) amounts to. Locke asks us to consider the claims

- (i) that God has given some systems of matter “fitly disposed,” or “put together as he thinks fit,” mental states;
- (ii) that God has taken a purely material being and joined to matter “so disposed” a thinking immaterial substance

It is important to note that the talk of “fitly disposed” systems of matter occurs in Locke’s statement of *both* (i) and (ii), even though, on option (ii), the arrangement or disposition of the parts of matter can have nothing at all to do with the presence of thought. Why, then, does Locke mention it in stating (ii), substance dualism? If God is joining an immaterial substance to a material one, it just does not matter which hunk of matter he selects. I presume that Locke is thinking here that it would be pointless (or perhaps cruel) to attach such an immaterial substance to a stone or a toilet bowl. Thus, hypothesis (i), by the same token, need not be the claim that thought could arise from matter *in virtue of* its arrangement, anymore than (ii) entertains the possibility of an immaterial substance’s arising from a material one in virtue of the arrangement of matter.

Thus, our second problem is solved. Locke can happily reject the possibility of mind-body supervenience while accepting that God could add the property of thought to a body. It is just that *how* God does this is constrained by the nature of voluntary actions: he cannot add thinking to a material being *merely* by fooling about with the arrangement and motion of its parts. He has to add a nonmaterial property to a material being.²²

Are we then forced to conclude that God can only bring about a thinking thing from a material one by a miraculous, contranatural act? That is, must we retreat to the Leibniz/Wilson reading of superaddition? No. For as Lisa Downing has pointed out, Lockean superaddition is only with respect to the *nominal* essence of matter.²³ Our idea of body is manifestly impoverished in many ways: as Locke tells Stillingfleet, “The gravitation of matter towards matter, by ways inconceivable to

me, is . . . a demonstration that God can, if he pleases, put into bodies powers and ways of operation, *above what can be derived from our idea of body, or can be explained by what we know of matter*, but also an unquestionable and everywhere visible instance, that he has done so."²⁴ If property dualism turns out to be true, it need be no more miraculous than gravitation. Both sorts of property are left out of our abstract idea of matter. It would be as rash to conclude that nonphysical properties cannot attach to matter as it would to reject gravity as an "occult quality." Hypothesis (i) cannot be ruled out; nor need it involve the addition of a property inconsistent with the other properties of bodies that we *do* know about.

Thus, the choice between Ayers's deflationary account and Leibniz's inflationary account is a false one. On the third option endorsed here, Locke allows conceptual space for property dualism. He does not allow such space for supervenience theories.

5. CONCLUSION

Locke rejects the supervenience hypothesis for much the same reason as Kim: it turns mental events into mere side effects. Yet this is fully consistent with the claim that what thinks in us might, for all that, be a material substance. We must, therefore, temper our view of Locke's agnosticism and erase the myth that he holds all metaphysical hypotheses about the mind to be epistemically possible. There are some that *can* be ruled out. And among them is the idea that mental states supervene on matter in motion.²⁵

Virginia Tech

NOTES

1. Throughout, I shall speak indifferently of "events," "states," and "properties" since nothing in my argument turns on the distinctions among them.

2. For the exclusion argument, see Jaegwon Kim, *Mind in a Physical World* (Cambridge, MA: MIT Press, 1998).

3. By the closure of the physical world, I mean the claim that every physical event has a physical cause sufficient to bring it about. (There are stronger versions of the closure principle that include the rejection of overdetermination.)

4. As a thesis about the coinstantiation of properties within or across possible worlds, supervenience, whether weak or strong, local or global, is compatible

with a wide range of theories I would not want to count as “materialist.” As Terence Horgan argues in “From Supervenience to Superdupervenience,” *Mind* 102 (1993): 555–86, some forms of emergentism are consistent with supervenience. Yet one hardly wants to count what should plainly be labeled a kind of property dualism as materialism. By “materialism,” then, I shall mean views that hold not just that the mind is a material substance but that its states and properties are also, and exhaustively, material. (I hope the reader will forgive my nearly vacuous description of materialism, since arriving at an adequate formulation of materialism is itself a substantial philosophical problem.) This characterization rules out all forms of property dualism, including neutral monism. On the other side of the spectrum, it is worth noting that type-identity theory is, trivially, a form of supervenience theory: what easier way to account for the modal covariance of mental and physical properties than to identify them?

5. This is true only of type-identity theories; a token-identity theory might succumb to the threat of epiphenomenalism (see the following note).

6. Donald Davidson’s anomalous monism might seem to escape this argument. (See esp. “Mental Events” in Davidson, *Essays on Actions and Events* [Oxford: Oxford University Press, 1980], 207–55.) Davidson insists that his view is compatible with supervenience. And indeed, since Davidson’s theory is a form of token-identity theory, it is trivially so. If any token of a mental state is also a token of a physical state, as Davidson claims, then mental states will co-vary with physical states across and within worlds. So, Davidson seemingly can endorse supervenience without calling into question the causal efficacy of mental states. But Kim (*Mind in a Physical World*, 29–37) persuasively argues that this appearance is deceiving and that Davidson’s view collapses into epiphenomenalism.

7. References are to *An Essay concerning Human Understanding*, ed. P. H. Nidditch (Oxford: Clarendon Press, 1975), and follow the standard format: book.chapter.section, page number.

8. See, e.g., II.xxiii.32, 314 and portions of the exchange with Edward Stillingfleet, particularly *The Works of John Locke*, 10 vols. (London: Tegg et al., 1823), 4:457ff.

9. For the controversy, see esp. M. R. Ayers, “Mechanism, Superaddition, and the Proof of God’s Existence in Locke’s *Essay*,” *Philosophical Review* 90 (1981): 210–51, and M. Wilson, “Superadded Properties: A Reply to M. R. Ayers,” *Philosophical Review* 91, no. 2 (1982): 247–52. For a more recent take on the issue, see Matthew Stuart, “Locke on Superaddition and Mechanism,” *British Journal for the History of Philosophy* 6 (1998): 351–79.

10. That is, it is the same in terms of its determinable, not determinate, qualities. If Locke’s preferred hypothesis, corpuscularianism, is true, then even the most fundamental bits of matter can differ in terms of shape. Locke’s point is just that there is nothing special about any of these determinate shapes that should make one bit of matter admit of mental properties while others do not.

11. "God and Matter in Locke: An Exposition of *Essay* 4.10," in *Early Modern Philosophy: Mind, Matter, and Metaphysics*, ed. Christia Mercer and Eileen O'Neill (Oxford: Oxford University Press, 2005), 166–67.

12. Note that I go on to show why Locke could *not*, in fact, have endorsed this premise. My strategy here is simply to present a *prima facie* plausible reading of the argument before going on to undermine it.

13. See Vere Chappell, "Power in Locke's *Essay*," in *The Cambridge Companion to Locke's Essay*, ed. Lex Newman (Cambridge: Cambridge University Press, 2007), 140, on Locke's peculiar use of "preference."

14. The "appropriate" criterion is necessary here if we are to rule out deviant cases, in which a mental state like a thought of the Eiffel Tower causes one's hand to move. It would be odd on anybody's view to describe the motion of one's hand as voluntary in such a case.

15. See Gideon Yaffe, *Liberty Worth the Name* (Princeton, NJ: Princeton University Press, 2000), 99ff., and E. J. Lowe, *Locke* (London: Routledge, 2005), 144ff.

16. Two points to note here: Locke seems clearly to be dealing with event-, rather than substance- or agent causation. It is a further question whether his argument would apply to theories that use other approaches to causation, such as John Searle's intentional causation (*Intentionality* [Cambridge: Cambridge University Press, 1983]). Moreover, Locke's view of causation in the physical world, despite his genuflecting to Newton, is a mechanistic one, according to which bodies act on other bodies only by means of impulse. This, of course, is false, though I do not see how this could vitiate his argument.

17. Bennett ("God and Matter in Locke"), 175; see Ayers's ("Mechanism, Superaddition, and Proof of God's Existence") and *Locke: Epistemology and Ontology*, 2 vols. (London: Routledge, 1991), 2:169ff.

18. In Locke (*Works*), 4:364; cited in Ayers ("Mechanism, Superaddition, and Proof of God's Existence").

19. A difficulty here is created by Locke's mentioning motion among the superadded properties. There is a clear difference between motion and beauty: motion arises when God *pushes* matter around, whereas beauty arises when God *arranges* matter in a given way. For thought to arise in the ways envisaged by NTM, it seems, God must add both motion and a certain arrangement.

20. *Nouveaux Essais*, in C. J. Gerhardt, ed., *Die Philosophischen Schriften von Gottfried Wilhelm Leibniz* (Hildesheim: Georg Olms, 1978), 363.

21. See Bennett ("God and Matter in Locke"), 176–77.

22. For more on Locke's allowance of the epistemic possibility of property dualism, see José Luis Bermúdez, "Locke, Metaphysical Dualism and Property Dualism," *British Journal for the History of Philosophy* 4, no. 2 (1996): 223–45.

23. Lisa Downing, "Locke's Ontology," in *The Cambridge Companion to Locke's Essay*, ed. Lex Newman (Cambridge: Cambridge University Press, 2007).

24. Locke (*Works*), 4:467–68; my emphasis.

25. I am grateful to Michael Jacovides, Antonia LoLordo, the editor of this journal, and an anonymous referee for helpful comments.