

Acta Cogitata

Volume 2

Article 1

Defending Downward Causation Only To Bring It Back Down

Michael Pratt
Grand Valley State University

Follow this and additional works at: <http://commons.emich.edu/ac>

 Part of the [Philosophy Commons](#)

Recommended Citation

Pratt, Michael () "Defending Downward Causation Only To Bring It Back Down," *Acta Cogitata*: Vol. 2 , Article 1.
Available at: <http://commons.emich.edu/ac/vol2/iss1/1>

This Article is brought to you for free and open access by DigitalCommons@EMU. It has been accepted for inclusion in Acta Cogitata by an authorized editor of DigitalCommons@EMU. For more information, please contact lib-ir@emich.edu.

Defending Downward Causation Only To Bring It Back Down

Michael Pratt, Grand Valley State University

Abstract

The so-called bottom-up threat to free will assumes a bottom-up metaphysics: the view that atoms (or whatever the most basic and fundamental element of material substances is) at the microphysical level dictate the behavior and ontology of material substances at higher macrophysical levels of composition. More specifically, bottom-up metaphysics maintain that all macrophysical states of affairs supervene on their constituent microphysical properties, such that any macrophysical change requires a change at the microphysical level. This metaphysical picture implies that human persons—and consequently, human actions—supervene on what their constituent atoms do or are like. In *Objects and Persons*, Trenton Merricks argues that a human person having the capacity for downward (mental-physical) causation is both necessary and sufficient for that person to have a choice about what her atoms do or are like, and moreover, that downward causation blocks the bottom-up threat to free will. In “Can Downward Causation Save Free Will?” Justin Capes argues that Merricks’s response to the bottom-up threat is unsuccessful. In this paper, I will (i) explain the bottom-up threat and Merricks’s response to it, (ii) explain Capes’s criticisms of Merricks’s response, (iii) argue that Capes’s criticisms are unsuccessful in rebutting Merricks’s response, and (iv) suggest an alternative way to object to Merricks’s response.

Defending Downward Causation Only To Bring It Back Down

In *Objects and Persons*, Trenton Merricks argues that a human person having the capacity for downward (mental-physical) causation is both necessary and sufficient for that person to have a choice about what her atoms do or are like, and moreover, that downward causation blocks the bottom-up threat to free will. In “Can Downward Causation Save Free Will?” Justin Capes argues that Merricks’s response to the bottom-up threat is unsuccessful. In this paper, I will (i) explain the bottom-up threat and Merricks’s response to it, (ii) explain Capes’s criticisms of Merricks’s response, (iii) argue that Capes’s criticisms are unsuccessful in rebutting Merricks’s response, and (iv) suggest an alternative way to object to Merricks’s response.¹

The Bottom-Up Threat to Free Will

The so called bottom-up threat to free will assumes a bottom-up metaphysics: the view that atoms (or whatever the most basic and fundamental element of material substances is) at the microphysical level dictate the behavior and ontology of material substances at higher macrophysical levels of composition. More specifically, bottom-up metaphysics maintain that all macrophysical states

¹ Thanks to Andrew Spear for the invaluable comments and feedback. Thanks also to Michael Robinson for the conversation and recommendations that spawned this paper.

of affairs supervene on their constituent microphysical properties, such that any macrophysical change requires a change at the microphysical level. This metaphysical picture implies that human persons—and consequently, human actions—supervene on what their constituent atoms do or are like. The argument for the bottom-up threat to free will (hereafter, the bottom-up threat) takes the following form:

- (1) If some agent *S* has a choice about what actions she performs, then *S* has a choice about what her atoms do or are like.
- (2) *S* does not have a choice about what her atoms do or are like.
- (3) So, *S* does not have a choice regarding what actions she performs.
- (4) *S* having Free will requires that *S* does have a choice about what actions she performs.
- (5) So, *S* does not have free will.

Premise (1) derives from the bottom-up assumption that every action performed by an agent supervenes on what her atoms do or are like. For, given the relationship between macrophysical states (e.g., some agent's choice) and microphysical states (e.g., some agent's constituent atoms), an agent having a choice about what actions she performs necessitates that agent having a choice about what her actions supervene on—namely, her constituent atoms. It is also important to note that above the argument assumes that the following principle of entailment is valid: "If *p* entails *q*, and no one has a choice about the fact that *p*, then no one has a choice about the fact that *q*" (Capes 133).² This inference rule, taken together with premises (1) and (2), generate the bottom-up threat.

To illustrate, suppose that some agent instantiates some mental state that is followed by her raising her hand. On the bottom-up view, that agent's macrophysical states of affairs (her mental state and subsequent hand raising) supervene on their constituent microphysical states. So, our agent's microphysical states entail her mental state and hand raising, and since she does not have a choice about her microphysical states, she thus does not have a choice about her mental state or about her hand raising. Given this example, it seems intuitive why a bottom-up ontology threatens free will, since free will requires that we do have a choice about what actions we perform—a requirement that is plainly impossible given the truth of bottom-up metaphysics. With the bottom-up threat briefly explained, I will now advance to Merricks's response.

Responding to the Bottom-up Threat: supervience and downward causation

Merricks objects to premise (1) by denying the supervience thesis. More specifically, he denies that mental states supervene on physical states and thus, if the occurrence of a mental state or property amounts to performing an action, (1) is false. E.g., if Sally's mental state of deciding to raise her hand qualifies as performing an action, and that mental state does not supervene on a physical state, (1) is false.

Moreover, Merricks rejects (2), maintaining that human agents have non-redundant causal abilities, insofar as "we cause things not also caused by our constituent atoms in virtue of having conscious mental properties that do not supervene on what our constituent atoms do or are like" (Capes 135). Thus, Merricks concludes that human agents have downward causal control over their constituent microphysical states of affairs, and moreover, that downward causation is sufficient for having a choice about what agents' atoms do or are like; thus (2) is false. With Merricks's response to the bottom-up threat cashed out, I will now explain Capes's criticisms of it.

² This inference rule is a version of Peter van Inwagen's 'Rule Beta'. For more on this, see van Inwagen, (1983).

Capes's Criticisms and Cases

Purporting to show that Merricks's response is unsuccessful in dispelling the bottom-up threat—specifically, that it fails to show that (2) is false—Capes distinguishes between what he takes to be two distinct claims:

(M1) "Human persons have downward causal control over their constituent atoms in virtue of having conscious mental properties that do not supervene on physical states or properties . . .

(M2) Downward control of this sort is sufficient for having a choice about what one's atoms do or are like" (Capes 135).

Capes thinks that (M1) is unproblematic, and moreover, concedes that agents possessing downward causal control is probably a necessary condition for free will. It is (M2), rather, that he takes issue with; for agents possessing downward control "is not sufficient for having a choice about anything, much less about what one's constituent atoms do or are like" (135). Capes points out that Merricks offers little justification for accepting (M2), and suggests that perhaps downward causal control is necessary (but not sufficient) for the kind of 'active power' that agent-causal libertarians require.³ At any rate, Capes's criticisms purport to show that downward control is not sufficient for having a choice about what one's microphysical states do or are like. Since I have limited space here, it will suffice to say that Capes thinks downward causal control is not sufficient for free action—the agent must also "be able to determine at will how one's causal powers are employed" (136). For, if one has downward causal control, but does not have a choice regarding how that control is utilized, it seems like one would not have a choice regarding what actions one performs.

Capes offers a series of cases that he takes to be counterexamples to (M2). His initial suggestion is this: possessing downward causal control does not imply that one couldn't be determined to employ that power in various ways.⁴ If this is possible, and assuming that causal determinism is incompatible with free will, then (M2) is false. Capes first considers a case in which some agent S is causally determined to exercise her downward causal control. In such a case, S's free action would be precluded due to the truth of determinism, but S would have nevertheless exercised downward causal control. Thus, Capes claims, downward causal control "is not sufficient for having a choice about what one's atoms do or are like" (137).

Anticipating a question begging complaint, Capes posits the following four cases within possible worlds that are indeterministic:

Case 1:

"Imagine . . . that God causally intervenes in the world at a particular moment and brings it about that Susan decides to raise her hand. Assuming that Susan has no choice about whether God does this or about the fact that her decision is sufficient to bring it about that she raises her hand, it seems to follow that she has no choice about what the atoms that compose her hand do or are like. Nevertheless, she

³ Broadly construed, agent-causal libertarianism is the view that maintains: (i) determinism is false, (ii) we have free will, and (iii) free agents are the direct causes of their actions. Moreover, as Capes points out, agent-causationists tend to think that free agents possess causal powers that enable them to bring about effects in the physical world.

⁴ It's worth mentioning that this only applies to Cases 1-3, since Case 4 is indeterministic. So, we may more generally summarize Capes's cases as follows: possessing downward causal control does not imply that one has a *choice* about how that downward causal power is employed.

apparently exercises downward causal control over her atoms, insofar as her raising her hand is a causal upshot of a conscious mental property” (137).⁵

Case 2:

“Similar circumstances arise in cases of covert manipulation. We can replace God with neuroscientists who have created Susan and who directly manipulate her, thereby causing her to decide to raise (and thus to raise) her hand” (138).⁶

Case 3:

“Take [Case 2]. We can alter the case so that while the neuroscientists do not directly control her actions, they have ‘programmed’ her using genetic and social engineering techniques so that when she is in circumstances of a certain sort she will reason in a way that leads her to decide to raise her hand . . . Part of the experiment, we might imagine, is to find out which behavior results from which forms of neural manipulation. Here it is clear that when Susan finds herself in the relevant circumstances, she has no choice about whether she decides to raise her hand and thus no choice about what the atoms that compose her hand do or are like” (138).⁷

Case 4:

“Patrick is planning to shoot the Queen. Unknown to him, there is a diabolical demon monitoring his thought processes (diabolical demons, like nefarious neurosurgeons, can do such things), and if Patrick seriously considers not shooting the Queen, the demon will intervene and force him to (decide to) shoot her. As it happens, Patrick never seriously considers not shooting the Queen, and decides to shoot, and does shoot, her entirely on his own . . . this is a case in which an agent exercises downward control over his constituent atoms (those that compose his trigger finger) and yet lacks a choice about what those atoms do or are like” (139).⁸

The general import of these cases is this: possessing downward causal control does not imply that one has a choice about how that downward causal power is employed. For, it seems reasonable to think that free will requires having downward causal powers and a choice about how those powers are employed, and thus it is unclear how having downward causal powers alone would be sufficient for free will. With these cases and their general import in mind, I will now advance to my responses.

Responding to the cases

As Capes anticipates, Merricks could respond by questioning the coherency of Cases 1 and 2. For, if God or neuroscientists determine that Susan decides to raise and thus raise her hand, what kind of control could Susan be thought to have? It seems reasonable to question whether Susan has any control at all, since her decision was brought about by direct manipulation. Moreover, one might argue

⁵ It’s worth noting that Case 1 is an instance of *local determinism* in an indeterministic world.

⁶ Case 2 is also an instance of local determinism in an indeterministic world.

⁷ While Case 3 is also an instance of local determinism in an indeterministic world, Capes thinks that it is fundamentally different from Cases 1 and 2. This is due to the fact that the manipulation occurs ‘at a distance’, and thus, it is “less obvious that Susan makes no choice of her own” (Capes 138).

⁸ Anticipating Merricks responding that downward control is sufficient for having a choice only in indeterministic contexts, Capes offers Case 4 (which is an instance of a Frankfurt-type case)—a scenario in which determinism is false and the agent is not caused to act due to forces outside of her control.

that Susan herself did not decide, but rather, that God and the neuroscientists have decided for her.⁹ Moreover, Merricks might respond that local determination renders the indeterminate worlds posited by Cases 1 and 2 irrelevant; for Susan was still determined to decide to do as she did, and thus it is unclear how this does not undermine any control she might have.¹⁰ If these lines of response are effective, then Cases 1 and 2 are unsuccessful in showing that (M2) is false.

Capes takes Case 3 to be more threatening to the truth of (M2), since the manipulation occurs “at a distance” and so it is “less obvious that Susan makes no choice of her own” (138). While covert manipulation is certainly descriptively different from direct manipulation, it’s unclear how programming Susan “using genetic and social engineering techniques” would not also amount to some sort of determinism. For, if some neuroscientist N conducted distant manipulation M which lead to Susan making some decision D to perform some action A (provided the right circumstance), wouldn’t it be the case that N caused M which consequently lead to D which ultimately caused A? Perhaps Capes would reply that since the circumstance that gave rise to Susan performing A was indeterministic, Case 3 still threatens the truth of (M2). While the circumstance which gave rise to Susan performing A is indeterministic, it is unclear how this is pertinent to whether Susan made a decision of her own, and more importantly, had any control in doing so. For, it seems like Susan’s acting freely would have to involve indeterminacy regarding her decision making faculties, not merely the circumstance in which she performs some action. But, in Case 3 it seems like Susan’s decision making faculties were not indeterministic—or at least not to the extent that agent-causalists require.

It is also possible to respond to Case 3 by pointing out that there is no relevant difference between agents in manipulation cases and ordinary agents in deterministic worlds.¹¹ For if Cases 1 and 2 can be diffused by appealing to the fact that determinism undermines any control an agent might have, and moreover, if there is no relevant difference between agents in manipulation cases and ordinary agents in deterministic worlds, then it is unclear how Case 3 is any different from Cases 1 and 2. If there is no relevant difference, then it seems that Case 3 fails to undermine (M2) for the same reasons Cases 1 and 2 do. If either of these responses are adequate, it seems that Case 3 fails to show that (M2) is false.¹²

Aiming to avoid the determinism complaints facing Cases 1-3, Capes offers Case 4, which posits an indeterministic scenario. One could argue, as others have, that if it is really true that Patrick cannot do otherwise in such a context, this is only due to the fact that Patrick was determined to act as he does.¹³ However, I think a more effective response would be to argue that Patrick does have a choice in

⁹ One could also raise a smaller issue with Case 2 by pointing out that direct manipulation of the brain would be presupposing the truth of mind-body supervenience. For, if the mind does not supervene on the body, it is unclear how the neuroscientists’ meddling would amount to Susan coming to have the mental state of ‘deciding to raise her hand’.

¹⁰ This also seems question begging, since Capes construes Merricks as an incompatibilist.

¹¹ This point is a feature of Alfred Mele’s *Zygote argument*. While the import of the entire argument may not be pertinent here, this particular feature of the argument is. For more on this, see Mele, (2006).

¹² A smaller complaint: similarly to Case 2, the indirect manipulation of the brain may presuppose the truth of mind-body supervenience. For, it seems that the efficacy of the neuroscientists’ meddling requires that Susan’s downward causal powers *do* supervene on her constituent atoms (namely, those comprising her brain). However, according to Merricks, conscious mental properties—presumably, those properties including downward causal powers—do not supervene on one’s constituent atoms. So, it seems that Case 3 may also be guilty of question begging, and moreover, it is unclear how the neuroscientists’ meddling would have any effect on Susan’s downward causal powers if they do not supervene on her constituent atoms since, presumably, the meddling could only extend to Susan’s constituent brain atoms.

¹³ See Widerker, (1995).

Case 4.¹⁴ That is, Patrick has the choice to decide to shoot the queen on his own, or to choose to abstain from shooting and (obviously) allow the demon to intervene. As it turns out (as in all Frankfurt-type cases), Patrick did decide to perform the action on his own. It seems that choosing to shoot the queen on his own implies that Patrick did have a choice—namely, whether or not his downward causal powers were in his control rather than the demon’s. Given this response, it is unclear how Case 4 undermines (M2).

If my responses adequately show that Cases 1-4 are ineffective in demonstrating the falsity of (M2), then Capes’s main objection to Merricks’s response to the bottom-up threat fails, or is at least problematic.

Bringing Downward Causation Back Down

I will now consider an alternative objection to Merricks’s response to the bottom-up threat. The essence of my worry is this: are there good reasons for thinking that human persons have downward causal powers? Recall what (M1) and (M2) claim: the use of the phrase ‘in virtue of’ in (M1) functions as a conditional—it indicates sufficient conditions for downward causation. With this in mind, we can reformulate (M1) as ‘human persons have downward causal control over their constituent atoms if they have conscious mental properties that do not supervene on physical states’. (M2) is also a conditional statement. From both (M1) and (M2) we can construct the logical relationship between (a) ‘S has conscious mental properties that do not supervene on physical states’, (b) ‘S has downward causal control over her constituent atoms’, and (c) ‘S has a choice about what her atoms do or are like’. Given that the success of the bottom-up threat rests on (c)’s denial, it seems that (c)’s truth is necessary for (d) ‘S has free will’.

The logical relationships between these claims represents a chain of necessary and sufficient conditions. With this in mind, it seems reasonable to ask what is necessary and sufficient for (a)? Note that (a) makes a claim about the relationship between mental properties and physical properties. Since (a) makes a claim about the mental, it seems like some well defended theory of the mental would be necessary and sufficient for (a). Capes construes Merricks as an agent-causalist, and given that many agent-causalists endorse some sort of dualism, it seems reasonable to construe Merricks as a dualist, at least about human persons.¹⁵ To more closely consider how this theory of mind might play a role in Merricks’ overall ontology, let us consider an argument for free will that derives from Merricks’ view.

- (1) For some agent S, S having free will requires S having a choice about what actions she performs.
- (2) S has a choice about what actions she performs if she has control over or a choice about what her constituent atoms do or are like.
- (3) If S has downward causal control, then S has a choice about or control over what her constituent atoms do or are like.

¹⁴ To see other similar responses to Frankfurt-type cases, see Robinson (2010), Naylor (1984), and van Inwagen (1978).

¹⁵ In rejecting the supervenience of the mental on the physical, Merricks tacitly implies that he endorses, or at least must endorse some theory of mind within his overall ontology. With non-reductive physicalism ruled out for Merricks—since it is typically taken to imply mind-body supervenience, which he flat out rejects—he is left with the following options: (i) physicalism, (ii) eliminativism, (iii) substance dualism, and (iv) property dualism. I would suspect that Merricks would reject (i), given that problems facing physicalist theories of mind typically force physicalists into accepting non-reductive physicalism. Moreover, Merricks explicitly denies that human persons (and hence, mental properties) are eliminative—leaving (iii) and (iv) as Merricks’s only palatable options. Agent-causalists often endorse versions of substance dualism—so perhaps Merricks would be open to endorsing (iii).

- (4) If S has conscious mental properties that do not supervene on physical states, then S has downward causal control.
- (5) S has conscious mental properties that do not supervene on S's constituent atoms (assume).
- (6) So, S has downward causal powers (from 4, 5).
- (7) So, S has a choice about or control over what her constituent atoms do or are like (from 3, 5).
- (8) So, S has a choice about what actions she performs (from 2, 7).
- (9) So, S has free will (from 1, 8).

My objection targets premise (5), for if (5) is false or ill-supported, then the argument cannot go through. Flat out rejecting (5) would certainly evoke an ad hoc criticism, so let us instead consider an argument in support of (5) that Merricks might sympathize with.

- (1) If X and Y are distinct substances, then X cannot supervene on Y.
- (2) The mind (or conscious mental properties) and body (physical states) are distinct substances.
- (3) So, the mind does not supervene on the body.

These three premises constitute a general argument for mind-body dualism. The question, thus, is whether mind-body dualism is tenable. Historically, it has been widely criticized, typically regarding the problem of mind-body interactionism. Briefly explained, the problem of mind-body interactionism questions the coherence of the idea that an unextended, immaterial substance (mental thing) would be able to interact with extended, material substance (physical thing). Jaegwon Kim defends a version of this objection in *Physicalism, or Something Near Enough*. Since I have little space here, I will only briefly cash it out in informal terms.

The real problem facing substance dualism, Kim thinks, is explaining how an immaterial, unextended thing remains intimately connected with a single material, extended thing—such that they constitute a single unit or whole. Why, Kim asks, does one's immaterial, unextended mind not venture into other bodies? What prevents one's mind from causing physical effects to arise in another person's body? I.e., it is unclear why one individual's mind only produces physical effects in that person's body, and moreover, why there seems to be a unique and particular unity between one's mind and one's body. Regarding the general argument for mind-body dualism, Kim's objection seems to question the plausibility of premise (2). For, given the worry raised by Kim, it seems that we may not have good reason for accepting (2), and thus the argument for mind-body dualism is at least problematic. There have been attempts to resolve this issue, but here it will suffice to say that there are, to say the least, serious problems with the argument supporting premise (5), so we may have good reasons for rejecting (5), and thus for rejecting the derived argument for free will.¹⁶

In sum, if avoiding the bottom-up threat requires endorsing mind-body dualism, then more independent argument for dualism's plausibility is needed. For, if having free will (and also blocking the bottom-up threat) requires downward causation, and if having downward causation requires endorsing mind-body dualism, then it seems like problems facing mind-body dualism may render Merricks's response to the bottom-up threat untenable. If this alternative objection is convincing, it seems like the proposal that human persons can have downward causation may be more intractable than it appeared, and moreover, that Merricks's attempt to salvage free will via downward causation may be as unpalatable as the very objection that evoked it.

¹⁶ For responses to the above objection to mind-body dualism, see O'Connor, (2000).

Works Cited

Capes, Justin A. "Can 'Downward Causation' Save Free Will?." *Philosophia* 38.1 (2010): 131-142.

Kim, Jaegwon. *Physicalism, or something near enough*. Princeton University Press, 2007.

Mele, Alfred R. *Free will and luck*. Oxford University Press, 2006.

Merricks, Trenton. "Objects and persons." Oxford University Press, 2001.

Naylor, Margery Bedford. "Frankfurt on the principle of alternate possibilities." *Philosophical Studies* 46.2 (1984): 249-258.

O'Connor, Timothy. "Causality, mind, and free will." *Noûs* 34.s14 (2000): 105-117.

Robinson, Michael. "Modified Frankfurt-type counterexamples and flickers of freedom." *Philosophical studies* 157.2 (2012): 177-194.

Van Inwagen, Peter. "Ability and responsibility." *The Philosophical Review* (1978): 201-224.

Inwagen, Peter van. "An essay on free will." *New York: Oxford* (1983): 68.

Widerker, David. "Libertarianism and Frankfurt's attack on the principle of alternative possibilities." *The Philosophical Review* (1995): 247-261.