My Parts Made Me Do It

Joshua Rasmussen and Andrew Bailey

"…we can always undermine the sense of our own autonomy by reflecting that the chain of explanation... can be pursued till it leads outside our lives." – Thomas Nagel[1]

1. Introduction

Late twentieth century developments in mereology and the logic of identity give us resources for articulating more precisely a puzzle about person's relations to their proper parts. In this article, we aim to accomplish three tasks: (i) unpack the puzzle, (ii) identify possible solutions, and (iii) reveal highly strange and perplexing implications of *every* solution. Our larger aim is to provide a framework for organizing and better analyzing a wide range of theories about how human persons could relate to their proper parts.

We start with a thought experiment. Imagine Black has invented a shrinking machine. His goal is to get inside Jones' head—literally. Black climbs inside a saucer-like craft and shrinks himself and his craft to the size of a molecule. He then navigates into Jones' head as he maneuvers between skin and brain cells. Once he reaches the frontal cortex, he presses a shiny red button. The craft releases energy into calculated areas of Jones' brain. Suddenly, Jones' begins to form an overwhelming desire to kill Smith. He then proceeds to do just that.

Is Jones responsible for acting on his desire to kill Smith? We suspect that many would hesitate to say so. After all, Black's actions *fix* Jones' current psychological states and actions.[2]

[1] Nagel 1986, 136.

[2] While the controller, Black, is an intentional agent, this feature of the story is not essential. We could just as well imagine that Black's device operates in the same controlling way by accident *without* anyone's intentions. Jones would still appear to lack responsibility for his resulting behavior.

We hasten to add that the problem is not determinism per se. We are sympathetic with the position that a person can be morally responsible even if determinism is true,[3] but there is an importantly different threat here to responsibility. The threat we see arises from the fact that the *immediately* determining cause is not even a *state of Jones*. The controller consists of states of someone else, and the influence is immediate and present. Even a compatibilist could think that, in general, if the behavior of object X immediately and deterministically determines the behavior of some *other* object Y, then Y is a mere puppet of X—and so not responsible for its behavior.[4] The real life puzzle arises when we add a simple thought: the atoms inside you are not *you*. How then are you not a mere puppet of their power?

Our statement of the puzzle so far is (too) rough and ready.[5] In what follows, we will unpack the puzzle by showing how certain mereological principles threaten the "ownership" of the states that control an agent. Our goal is not to argue that the problem we raise is unsolvable or that we in fact are not responsible for any behavior. Nor do we aim to show that we must reject a given framework that generates the puzzle. Rather, we bring into focus an important problem that every philosopher from every framework must grapple with—if anyone is to maintain the existence of personal responsibility.

We shall proceed as follow. We will first offer a more careful statement of the conundrum. Second, we will consider a solution stemming from compatibilism about moral responsibility and determinism. Our goal there will be to explain why our puzzle is neutral about such compatibilism. Third, we will examine a promising solution that denies a bottom-up

[3] One of us accepts the compatibility of moral responsibility and determinism, while the other of us does not rule it out.
[4] Cf. McDaniel 2017: 192-193.
[5] For some data on what the folk think about related issues, see Nahmias 2014.

explanatory picture of composite objects. We close by identifying significant implications of denying the bottom-up picture.[6]

2. The Puppet Puzzle Stated

The Puppet Puzzle comprises five plausible but inconsistent theses:

> Composition: I am composed of (proper) parts.
>
> Responsibility: I am responsible for my behavior.
>
> Atomic Priority: If Composition about human persons is true, then there are atoms whose behavior both necessitates and explains my psychological states and behaviors.
>
> Non-Identity: There are no atoms, the ps, such that I am the ps.
>
> Off the Hook: If there are things distinct from [not identical to] me whose behavior both necessitates and explains my psychological states and behaviors, then I am not responsible for my behavior.

We see that the five theses are inconsistent by the following deductions:

A. Therefore, if Composition is true, then the behavior of things distinct from me both necessitates and explains my psychological states and behavior. (from Atomic Priority, Non-Identity)
B. Therefore, if Composition about human persons is true, then I am not responsible for my behavior. (from Off the Hook, A)
C. Therefore, I am not responsible for my behavior and I am responsible for my

[6] Recent arguments against Composition (typically used to undermine materialism, an extension we do not here endorse) include Barnett 2010, Bogardus 2012, Collins 2011, and Lowe 2010.

behavior (from Responsibility, Composition, B)

The Puppet Puzzle is in the first person. We invite you to consider each of its theses from your own perspective.

We will look closer at Atomic Priority, Non-Identity, and Off the Hook. We will not comment further on Responsibility and Composition, except to note that they are widely held. Many philosophers think that (i) we are morally responsible for what we do at least some of the time and that (ii) we are composed of proper parts. If the result of our inquiry presses them to give up (i) or (ii), that will be significant in its own right. We now turn to the prospect of other ways out of the puzzle.

Atomic Priority: If Composition about human persons is true, then there are atoms whose behavior both necessitates and explains my psychological states and behavior.

Atomic Priority expresses a common view about human persons. The materialist formulation of this view is familiar enough: humans are built of wholly material parts (brains, bones, lungs, cells, and so on), and the activity of those parts jointly explains the activity of the whole person. We'll mostly focus on materialist versions of Composition, but we note that even certain non-materialist compositists (union dualists, for example) may face a puzzle here.

In a materialist framework, Atomic Priority falls out of the more general view that the activities of physical objects are explained and determined by the activities of their smallest parts. For example, a rock is composed of atoms of various kinds, and the behavior of those atoms both necessitates and explains the behavior of the rock as a whole. Or take a tornado: the

twisting of the tornado as a whole is both necessitated and explained by the joint motions of all the atoms within it. The idea, then, is that material things do what they do because of what their material parts do. Applying this idea to people, the result is that people do what they do because of what their material parts do.

We make no assumption about the kinds of things that compose us. They could be cells, strings, fields, or something else besides. They could even be immaterial souls or soul-like items. Nor do we assume that there is a most fundamental level of decomposition. For ease of presentation, we will talk of our "atoms" as shorthand for things we are composed of at some arbitrary level of decomposition. So understood, the theses of the Puppet Puzzle are compatible with the thesis that we are "gunky" creatures (that all of our parts have proper parts).

Let us be more precise about necessitation and explanation. Necessitation: "the xs' behavior necessitates y's behavior" means this: for every truth F_y about the y's behavior, there is a truth F_x about the xs' behavior, such that necessarily, if F_x then F_y. So, for example, if the motion of a certain rock is necessitated by the motions of certain atoms, then truths about the motions of certain atoms entail truths about the motion of that rock.

Explanation: "x's behavior explains y's behavior" means that y is doing what it does because x is doing what it does. Much recent work is devoted to better characterizing the nature of explanation.[7] For our purposes, we seek to be as neutral as we can about theories of explanation, while we note that different accounts will give rise to different precise articulations of our puzzle. Our only explicit requirement is that explanation is asymmetric: where x and y are distinct, if *x because y*, then it is not the case that *y because x*. For simplicity, we will sometimes use 'fixes' to

[7] See, for example, Audi 2012, Bennett and McLaughlin 2005, Correia 2008, Fine 2012, Rosen 2010, Schaffer 2009, and Trogdon 2013. Wilson 2014 provides a helpful and necessary cautionary note.

abbreviate ‘necessitates and explains’. So, Atomic Priority says that given Composition, the behavior of my atoms fixes my behavior.

We should emphasize that Atomic Priority does not imply determinism—the thesis that the state of the world at any time together with the laws of nature entails all subsequent states. Atomic Priority is only concerned with necessitation and explanation from parts to whole, not determinism from the past to the future. So, Atomic Priority doesn’t imply that the past together with the laws determine the future. The initial theses of the Puppet Puzzle threatens responsibility independently of any threat that may or may not be posed by classical determinism. We will return to this point.

Non-Identity: There are no atoms, the ps, such that I am the ps.

This thesis follows from the general thesis that composition is not identity: a thing is not identical to its proper parts. We leave open whether a person is reducible to or analyzable in terms of an *arrangement* of atoms. Maybe there are some atoms, the ps, which are arranged F-wise. And maybe I exist if and only if (and if so, because) those ps are arranged F-wise. Still, it doesn’t follow that I *am* the ps (whether arranged F-wise or otherwise).

The type of identity at issue here is strict, numerical identity—the sort of identity that obeys Leibniz’s Law: thus if my proper parts have features that I lack, then I am not identical with those parts. With these clarifications in mind, may argue for it like so:

There is *exactly one* of me.

There are *many* of my atoms.

Therefore: I am not my atoms.

Or like so:

I am not a proper part of me.

My atoms are proper parts of me.

Therefore: I am not my atoms.

Despite these arguments, we expect reasonable disagreement over Non-Identity—such is philosophy. Even still, reflection here brings to light a distinction between the thesis that people *have* atoms as parts, on the one hand, and the thesis that people *are* atoms. You, like us, may find it more plausible to suppose that people have atoms as parts (such as parts of the material arrangement or material composite or soul-body composite to which they are identical) than that they literally are atoms.

Similar remarks apply to the relationship between my *behavior* and the plurality of behaviors of my proper parts. Even if the behaviors of my parts jointly compose *my* behavior, they aren't numerically identical to it, since composition is not identity. To further draw out this point, suppose Sam is a composite of exactly two atoms, $Atom_1$ and $Atom_2$. Sam is pictured below:

O O

Figure 1. Sam is expanding in virtue of his two parts moving apart.

Now consider two hypotheses:

$Layer_2$. Sam is expanding.

$Layer_1$. $Atom_1$ and $Atom_2$ are moving apart.

The thing to see is that even while we are supposing that Sam is identical to a composite of $Atom_1$ and $Atom_2$, $Layer_1$ is not identical to $Layer_2$, given Non-Identity.

Someone might reply that while Sam is not identical to Sam's proper parts, Sam's

behavior is identical to the *behavior* of Sam's proper parts. We do not rule this out. We note, however, that the success of this solution will depend on how one characterizes a behavior. Suppose a behavior is an event, and suppose we adopt Jaegwon Kim's theory that events are ordered triples consisting of a substance, property, and time. Then given Kim's identity conditions of ordered triples, the behavior of distinct things are themselves distinct events. So we'll find no help there.

Moreover, even on a less fine-grained theory of events (such as Davidson's), it is unclear how events involving non-identical things could be identical. You might think that at the very least the constituents of identical events must occupy the same spatial-temporal regions. Yet, the atoms that compose you do not occupy the same spatial-temporal regions as you, assuming atoms sometimes leave your body. Thus, while we could attempt to solve the puppet puzzle by *reducing* a person's behavior to the behavior of their parts, this reduction leads to puzzles of its own.

Off the Hook: If there are things distinct from me whose behavior both necessitates and explains my psychological states and behavior, then I am not responsible for my behavior.

We may understand "responsibility for my behavior" as responsibility for facts about at least certain actions of my body. For example, there is the fact that my fingers are typing keys on a laptop. The thesis, then, is that I am not responsible for any such facts if all such facts are fixed by behavioral facts about things *distinct* from me—i.e., things such that I am not them. For example, when Black controls Jones' brain, Jones' is not responsible for the behavior fixed by

Black's actions.

You might wonder whether the constitution of the controlling device is relevant to whether the device's behavior *just is* Jones' behavior. For example, maybe if Black's machine is made of Jones' atoms, then the actions of that machine are his actions. Yet this thought turns out (perhaps surprisingly) to contradict the premise that composition is not identity. To illustrate, suppose Black's machine is composed of Jones' atoms. Still, the behaviors of the many atoms are not numerically identical to *Jones'* mental behavior. Recall Figure 1: a state of affairs of a larger thing doing something is not identical to a state of affairs of some smaller constituents doing something, precisely because the larger thing is not the smaller things. Previous articulations of the mereological problem of agency have veiled the relevance of this particular distinction. The purpose of the puppet puzzle is to make explicit its relevance to responsibility. A root of the problem is lack of control: Jones lacks control over the motions of Black's device, no matter what atoms happen to compose it.[8]

Here is an instructive solution worth considering: perhaps you can be responsible for your behavior as long as the controllers are genuinely *your* parts. It is plausible, after all, that foreign objects do not become part of you merely by coming inside you. Perhaps, then, the parts of Black's machine are not parts of you. If so, then perhaps *parthood* is the relevant difference between Black's machine and the atoms in Jones' head.

While we think this solution is worth exploring, we note two reasons it is far from complete. First, it translates the mystery about moral responsibility into a mystery about

[8] We suggest that one's responsibility would be entirely eliminated. But even if one's responsibility were merely *mitigated*, our results would still be of great interest. After all, anytime we'd feel inspired to blame or praise someone, we could remember that in *every* case, there are significant mitigating factors: their atoms are making them do it!

parthood. Consider the swarm of atoms inside Jones's head. A portion of them is inside Black's machine. Now suppose only the atoms *outside* Blacks' machine are relevant to Jones' moral responsibility, and suppose that is because none of the atoms in Blacks machine count as parts of Jones. Then we have the puzzle of accounting for what precludes Black's atoms from being part of Jones. Suppose the answer is in terms of *causal integration* with other atoms that are "already" parts of Jones. But then why doesn't Blacks' device, which receives and sends signals to Jones' nervous system, count as causally integrated in the relevant way. If we explain the difference in terms of Jones' ability to *control* the atoms in some morally relevant way, then we're back to the original mystery: what difference between the atoms is morally relevant? In general, any answer in terms of parthood translates the original mystery into another mystery.

Second, and more fundamentally, even if we have the conditions for parthood, we do not thereby have an account of *how* parts that pull your strings can fail to make you a puppet. We can *say* your parts make you responsible, but saying so does nothing to explain how it can be so. We still have this mystery: how can we say the parts of Jones' that are themselves currently determined by Black's machine are more apt to preclude Jones' responsibility than the parts of Jones' that are currently determined by fundamental forces? The parthood solution does not answer this further question.

For our part, we suspect that a root reason the parthood solution is tempting is that we tend to associate persons with the totality of their parts. This association is pragmatic. However, it also temps us to treat composition as *identity*. If composition is identity, then we have a clear and morally relevant account of why Black's device precludes responsibility: it is because Black is not Jones, while Jones' atoms *are* Jones. Yet, for reasons we specified, we think you should resist this temptation, if you can.

To be clear, we have made a simplifying assumption. We have assumed that any responsible behavior we may have *already* performed has no bearing on whether or not some atoms or a device undermines our responsibility now. In other words, we assumed that responsible behavior in the past is irrelevant. This assumption is dubious, however. It is dubious because if you were actually responsible for putting the controlling device in your head, say, and if you knew that doing so would deterministically cause you to punch Peter in the nose, then it seems you would indeed be responsible for punching Peter. So, we shouldn't (automatically) assume that our past behavior is irrelevant to our present responsibility.

Fortunately, this assumption is not essential to the reasoning behind the Puppet Puzzle. The Puppet Puzzle depends on the thought that if my atoms' behavior fixes my behavior at some time, then this "bottom-up" determination holds at every time at which I exist. From this assumption it follows that if "bottom-up" determination removes my responsibility at some time, then it removes my responsibility at every time at which I exist. Therefore, our puzzle might well make use of the following more complicated thesis in place of Off the Hook: if for *every* time t at which I exist, there are things distinct from me whose behavior fixes my behavior at t, then I am not responsible for my behavior (ever). The argument concerning the neurosurgeon's device makes plausible this more complicated thesis. We set this complication aside, however, for ease of presentation.

Maybe this thought experiment fails to convince you. Still, it is instructive to have the challenge clearly laid out: if you are not worried that your behavior is necessitated by parts of you, then the challenge is to explain what could be relevantly different between Black's device and any other atoms in your brain. If our puzzle inspires philosophers to articulate a successful, satisfying explanation of the difference, that alone would be significant.

One way to investigate our case further is to consider the related literature on conditions relevant for agency. For example, we can consider Pereboom's first few manipulation cases, where the agent's reasoning and wanting is *immediately* determined (and fixed) by the behavior of another agent.[9] Immediate determination by other agents may appear to pose a greater, or different, threat to responsibility than is "across time" determinism. We will develop this idea further in the next section when we discuss compatibilism.

We add one further observation: the Puppet Puzzle exposes a unique and unhappy mixture of morality and composition. While there are a number of arguments about the compatibility of agency and materialism,[10] those arguments are importantly different from the Puppet Puzzle. Three examples will suffice. First, Cover and Hawthorne investigate whether a particular "agent-causal" theory is compatible with materialism. Their inquiry is valuable, yet their argument cannot appeal to those who aren't already attracted to—or inclined to accept—agent-causal theories. Second, the arguments on offer in Cover and Hawthorne, Merricks, and

[9] Pereboom 2001, 112–26.

[10] Cover and Hawthorne 1996, Merricks 2001, 155-161, Turner 2009, and O'Connor 2000 provide a representative sample. See also Capes 2010. Malcolm 1968 offers an important precursor to these arguments, although two differences are worth flagging. First, Malcolm's discussion is couched in terms of *mechanism* and *purpose*, features that make no appearance in the Puppet Puzzle. Second, Malcolm attempts to show that a mechanistic (roughly, non-teleological and materialist) metaphysics of mind is *self-refuting*. This is an interesting idea and has close connections to yet other arguments advanced by Alvin Plantinga and C.S. Lewis; see Plantinga 2011, Chapter 11 and Lewis 1996, Chapter 13. But we here pursue this more modest goal: revealing serious difficulties for compositists who affirm moral responsibility even where such difficulties fall short of self-refutation.

Another argument bearing superficial resemblance to the Puppet Puzzle is Jaegwon Kim's Causal Exclusion Argument, which purports to show that certain principles in the philosophy of mind rule out mental causation. See Kim 2005. There is indeed a resemblance: like the Puppet Puzzle, the Causal Exclusion Argument challenges the compatibility of a "bottom-up" metaphysics of mind with what seem to be facts about our mental lives (that we are sometimes morally responsible or that mental causation is real). But the resemblance is not penetrating. For the Puppet Puzzle makes no use of the principles that drive Kim's argument. It does not deploy a causal exclusion principle, for example. And so denying one or more of those principles, as one does when answering Kim's argument, is of no obvious help in resolving the Puppet Puzzle. For helpful explorations of the Causal Exclusion Argument in relation to human agency, see List and Menzies forthcoming and Wilson and Bernstein 2016.

Turner all deploy controversial "transfer principles". The Puppet Puzzle does no such thing (although, as we'll see later, we can reformulate the Puppet Puzzle using a transfer principle that is not subject to the usual complaints). Third, Merricks and Turner independently argue that free will is incompatible with certain brands of materialism. They, too, raise valuable considerations that are certainly relevant to our inquiry. Still, our question about Composition more generally has independent ramifications, regardless of whether it turns out that, say, moral responsibility does not require freedom. All of these authors, furthermore, target materialist assumptions; our Puppet Puzzle, by contrast, focuses on the more general thesis of Composition. We conclude, then, that while specific debates about agency are certainly relevant to our puzzle, the Puppet Puzzle has a distinctive position in the contemporary literature and invites wide interest.

3. A Compatibilist's Way Out

The Puppet Puzzle is reminiscent of arguments for the incompatibility of determinism and free will / moral responsibility (especially the latter). One might wonder, accordingly, whether the dialectical resources used to defend compatibilism about moral responsibility and determinism might enable an effective reply to the Puppet Puzzle. We suspect not. Instead, we are inclined to think that, while some issues are structurally similar, the issues at play in the Puppet Puzzle are orthogonal to issues at play in disputes over compatibilism about moral responsibility and determinism.

Compatibilism entails that moral responsibility is *possible*: compatibilists think that it's possible for someone to be determined and responsible; and from that it follows that it's possible

for someone to be morally responsible.[11] According to compatibilists, the conditions of responsibility can be met, and they can even be met if determinism is true.[12] So moral responsibility is an attainable feat. It is tempting to conclude from reflections along these lines that compatibilists have special reason to automatically resist theses like Off the Hook. There is reason to resist this temptation, however. For although compatibilists think moral responsibility is possible, they needn't think that it comes easy. Compatibilists and incompatibilists alike agree that subjects in at least some cases (cases of direct manipulation, for example), are not morally responsible. Such cases, we've argued above, offer support for Off the Hook.

However, let us consider a compatibilism-driven objection to Off the Hook. We put the objection as follows:

1. Compatibilism about moral responsibility and determinism is true. (premise)

2. Therefore: possibly, (i) someone S is morally responsible for φ-ing at time t, and (ii) a past state of the world together with the laws entail (necessitate) that S φs at t (from 1).

3. Necessarily: if (ii) above holds, then (iii) there are things distinct from S whose behavior necessitates and explains S's φ-ing at time t.

4. Therefore: possibly, (i) someone S is morally responsible for φ-ing at time t, and (iii) there are things distinct from S whose behavior necessitates and explain S's φ-ing at time t. (from 2, 3)

5. Therefore, Off the Hook is false (from 4).

[11] <Possibly(p and q)> strictly entails both <possibly(p)> and <possibly(q)>; so what the compatiblist affirms—<possibly(someone is morally responsible and determinism is true)>—strictly entails <possibly(someone is morally responsible)>

[12] See in this connection Fischer's 2006 admonition to reject too-stringent requirements on agency that amount to, in his memorable words, "metaphysical megalomania".

Although we think a compatibilist could find this argument compelling, many compatibilists will have reason to reject line 3. Consider compatibilists who grant that in the neurosurgeon case (where a neurosurgeon directly controls an agent's psychological states) the agent is neither free nor responsible. One way to motivate this judgement is to show how one's *history* could make a difference to responsibility (there may be other routes to this judgment too).[13] For example, some compatibilists suppose that history makes a difference to whether or not the agent's *own* psychological states explain the action.[14] To illustrate, a compatibilist might think that across-time determinism is no threat to responsibility so long as the determining events are *one's own*, so to speak. Moreover, they are one's own, one might suppose, by virtue of being causally integrated into one's character over a sufficiently long period of time.[15] This compatibilist could then suppose that even if determinism is true, there is no guarantee that behaviors of things *distinct* from S *explain* S's φ-ing. After all, citing facts about (say) particle formation during Planck's era doesn't obviously illuminate why S φs at t. Therefore, a common defense of compatibilism actually runs counter to line 3 of the argument against Off the Hook.

It is true that, given determinism, events that are one's own would themselves be ultimately determined by events that aren't one's own. Nevertheless, these outside past events are far removed from one's present actions; they do not directly determine one's actions. Hence, their threat to one's responsibility is less obvious.[16] We suggest that, from the compatibilist

[13] For an opinionated and helpful discussion of the historical/non-historical divide amongst compatibilists, see McKenna 2012a and 2012b and Levy and McKenna 2009, section 4.
[14] Mele 2005, 77–8, for example, suggests that one feature that separates Pereboom's bad manipulation cases (the cases in which responsibility is clearly undermined) and the classical determinism case is the presence of factors beyond the agents control that *directly* produce the agents' reasoning. For further discussion of compatibilist replies to such manipulation cases, see McKenna 2008 and McKenna 2014.
[15] Cf. Frankfurt 2002, 27–8.
[16] See, for example, Fischer's reply to the Zygote Argument for incompatibilism in his 2011.

perspective, a plausible account of why history is relevant to responsibility is that history distances the agent from the external causes of her existence, and that such distancing weakens the explanatory power of external factors on her present actions. An agent's actions are more properly explained by the *agent*'s own temporally proximate psychological states.[17]

The point here is that there is a distinction between determination that directly and immediately links external factors with the agent's actions (such as in the neurosurgeon case) and determinism that flows through a long history of causes and effects. The first kind of ("direct") determination seems to pose a far more potent challenge to responsibility than the second. For this reason, even if responsibility is compatible with "across-time" determinism, there remains a problem with thinking that responsibility is compatible with direct determination (where states of entities besides the agent immediately fix the agent's states and actions). In other words, compatibilism in the first case is compatible with Off the Hook.

We have put the compatibilist's retort as an objection to Off the Hook and suggested that this solution gets its life from the assumption that explanations enjoy transitivity, an assumption one might reject. Interestingly, the failure of transitivity itself suggests an intriguing solution to the Puppet Puzzle. The solution we have in mind targets Atomic Priority on the grounds that explanation from parts to whole is not transitive (and so, the thought goes, determination, the conjunction of necessitation and explanation, is not itself transitive). The metaphysical framework underwriting this objection entails that small parts may fix the features of the items they *immediately* compose without in turn determining the features of the larger things composed by those items. One implication of this strategy is that a popular "bottom-up" metaphysical

[17] See Roskies 2012.

picture is mistaken; we explore the consequences of this implication in section 4, below.

Consider, again, Off the Hook. While we certainly do not expect all compatibilists to accept the premise, we do expect many compatibilist sympathizers (ourselves among them) will find the premise plausible. For even if the psychological states that fix my behavior count as "mine," according to a compatibilist's criteria, my behavior, *along with those very psychological events*, are directly and immediately fixed by the behavior of things that are not themselves me or even "mine". (The mereologically more fundamental, non-psychological states of atoms or fields are not themselves "mine" on any criteria on offer.) Therefore, even if past non-psychological events are too distant in time to threaten my responsibility (because, say, across-time explanation isn't transitive), there are still states of things distinct from me that directly fix my actions and psychological states. They, like the neurosurgeon's device, directly fix my every move, feeling, and thought. The determining hand is direct and constant and so threatens my responsibility in a way that mere across-time determinism does not. This is the unique problem posed by "part-to-whole" determination.

We emphasize that the problem we are uncovering runs underneath the highly discussed problem of how agency might arise out of, or be understood in terms of, constitutive mental states. A venerable solution to the agency problem involves explaining how certain mental events could be "functionally identical" to the agent—or could *belong* to the agent in a relevant way.[18] In this scenario, an agent may be responsible for her behavior even if that behavior is deterministically explained by mental events which are not—strictly speaking—identical to the agent herself. Let us grant this solution. Still, the question we've raised remains: would moral

[18] Velleman 1992.

responsibility be undermined by factors that directly and constantly fix every mental event (including those that constitute the "surrogate agent")? Consider that the agent's mental events are themselves completely fixed by wholly *non-psychological* events at every turn.[19] One could think these determining events, which are not identical to the agent's behavior or to the behavior of any surrogate agent, threaten moral responsibility in a way that mere determinism across time does not. The threat of this immediate bottom-up determinism is easy to miss while one is preoccupied with the familiar, looming threat of across-time determinism.[20]

In summary, determination from parts to wholes is special because it is direct in a way that determinism from the past to the future need not be.[21] Hence, even those who think that responsibility is compatible with "across-time" determinism may worry that responsibility is not compatible with "part-to-whole" determination. The Puppet Puzzle poses a different kind of problem. We suggest, therefore, that many compatibilists (though we do not say all) will have reason to reject line 3 of the proposed solution to the Puppet Puzzle.

4. The Prospect of Top-Down Determination

Atomic Priority, recall, says this: if Composition about human persons is true, then there are atoms whose behavior necessitates and explains my behavior. Atomic Priority is a conditional; it is false if and only if its antecedent is true and its consequent is false. Its consequent is false only

[19] Again, we are assuming composition is not identity. So even if mental states are reduced to material states, the material states themselves are not identical to their proper parts.

[20] We have emphasized connections between the Puppet Puzzle and the threats determinism has been thought to pose to free will and moral responsibility. There is an important sense, though, in which the Puppet Puzzle points to a puzzle that is more akin to mind-body problems than to problems of free will and moral responsibility. In particular, the Puppet Puzzle exploits the difficulties of extracting something that is at least partly mental (morally responsibility) from something not even partly mental (the behavior of very small parts).

[21] See also Merricks 2001, 158.

if this more general principle is also false:

> Bottom-Up (BU): For any composite x, there are some ys, such that x is composed of the ys, and the behavior of the ys necessitates and explains the behavior of x.

BU basically says that determination always run from the parts of a thing to the whole thing itself: that is to say, a whole object does what it does because its parts do what they do. We could deny that. And then we'd be free to deny that our actions are fixed by our parts' behavior. There are two variations to consider:

> Weak independence: Sometimes what we do is independent of (not fixed by) our parts' behavior.
>
> Strong independence: Sometimes what we do is independent of (not fixed by) our parts' behavior; and indeed, what we do in fact fixes what our parts do and not merely by virtue of some of our parts determining that other of our parts to do what they do.

Both variations are interesting; let us first focus, though on what they have in common. They both deny Atomic Priority by denying BU.[22]

There are some puzzling consequences of denying BU, however. We briefly draw attention to three. First, it is quite natural (and common) to think that the intrinsic properties of wholes supervene and depend on the properties of their parts. Those properties surely depend on (or hold in virtue of) *something*, and what better candidates might there be for this office than

[22] The proposal that BU is false is reminiscent of Merricks' proposal that composite objects enjoy non-redundant causal powers. The two proposals are not equivalent, though they may seem similar in spirit. Both imply that there are composite objects that enjoy features not enjoyed by their parts, whether individually or in concert. Yet, in the case of Merricks' proposal, the features are causal, whereas the falsity of BU does not imply that the features are causal. Merricks' proposal, then, appears to imply the falsity of BU, but not vice versa. See Merricks 2001, especially pp.155-161.

their parts? William Hasker puts the point this way:

> ... whatever nonrelational properties the whole has must consist of properties of, and relations between, its parts; there simply is nothing else of which they could consist. If a property of a whole is not logically grounded in the properties of the parts, then it is "floating mid-air," unattached to any real individual—but this is unintelligible.[23]

Second, if BU is false, then we face new and difficult questions. Let's call these the "Special Independence" and "Special Top-Down" Questions. The Special Independence Question asks under what conditions a whole has properties not fixed by those of its parts. How, for example, would one build such a thing? Similarly, the Special Top-Down Question is the question of under what conditions things join together to form a complex that enjoys top-down determination.[24] To illustrate, imagine an arbitrary collection of atoms. Their behavior is normally, if not always, explanatorily prior to the behavior of whatever arrangements they form: the arrangement does what it does because the atoms do what they do. But if strong or weak independence are possible, then it should in principle be possible for those atoms to form a complex whole whose behavior is independent of or even prior to the behavior of those very atoms. In other words, it should be possible to flip the explanatory order from bottom-up to top-down (strong), or at least to disrupt the expected cross-level explanation (weak). How that can be? More precisely: what conditions could *in principle* explain the switch from purely bottom-up explanation to top-down or the lack of cross-level explanation? While you might think this could

[23] Hasker 1999, 138.

[24] This label is inspired by Peter van Inwagen's "Special Composition Question". See van Inwagen 1990, 20. Our questions here are to be distinguished from what we might call the "General Top-Down Question", which would be the question of what top-down determination *is*.

be answered empirically, you might instead wonder how it could even be possible in principle for a mere change in arrangement or spin or mass or any other physical property to explain a flip from *solely* bottom-up to top-down determination (or two-way determination).

A third consequence of denying BU that may be unique to strong independence is that it opens up the door to Priority Monism—the view that the world (which is something composed of everything) is itself the most fundamental thing.[25] If Priority Monism is true, then our behavior is still fixed by something distinct from us—namely by the behavior of the Whole of Everything. We would then be a puppet of the Whole of Everything and so wouldn't be responsible for our behavior (by Off the Hook). Thus, denying BU doesn't immediately get us out of the Puppet Puzzle. What we need is top-down determination that stops with *us*. But why should top-down determination be so courteous? One's answer here will depend on one's answers to the Special Independence and Special Top-Down Questions.

We have seen that part-to-whole determination inspires important questions. We do not claim that the questions are unanswerable. In fact, developments in the grounding literature provide a rich resource for further investigating these questions. Our observation here is just that the puppet puzzle helps to clarify and organize the potential options and relevant questions when theorizing about a person's relationship to their parts. Our own sense is that the most promising solution to the puppet puzzle will include the proposal that whole-to-part explanations are indeed possible. This proposal leads to hard questions of its own, however, as we pointed out. We suggest, then, that anyone who wrestles with the puppet puzzle will find treasures of insight about the nature of composite, responsibly beings—if there can be any.[26]

[25] See Schaffer 2010.

[26] For the curious: the authors are inclined to think that human persons are responsible, composite beings.

Works Cited

Audi, Paul. 2012. "Grounding: Toward a Theory of the *In-Virtue-Of* Relation" *Journal of Philosophy* 109, 12: 685-711.

Barnett, David. 2010. "You are Simple" in *The Waning of Materialism* (Bealer and Koons, eds.). Oxford: Oxford University Press.

Bogardus, Tomás. 2012. "What Certainty Teaches" *Philosophical Psychology* 25: 227-243.

Bebee, Helen. 2014. "Radical Indeterminism and Top-Down Causation" *Res Philosophica* 91, 3: 537-545.

Bennett, Karen and Brian McLaughlin. 2005. "Supervenience" *The Stanford Encyclopedia of* Philosophy.

Capes, Justin. 2010. "Can 'Downward Causation' Save Free Will?" *Philosophia* 38: 131-142.

Chalmers, David. 2006. "Strong and Weak Emergence" in *The Re-Emergence of Emergence,* P. Clayton and P. Davies, eds. Oxford: Oxford University Press, 2006.

Collins, Robin. 2011. "A Scientific Case for the Soul" in Baker & Goetz (eds.), *The Soul Hypothesis*. Continuum Press.

Correia, Fabrice. 2008. "Ontological Dependence" *Philosophy Compass* 3/5: 1013-1032.

Cover, Jan and John O'Leary-Hawthorne. 1996. "Free Agency and Materialism" in *Faith, Freedom and Rationality,* ed. Jeff Jordan and Daniel Howard-Snyder. Lanham, MD: Rowman & Littlefield.

Kit Fine. 2012. "Guide to Ground" in Hoeltje, Benjamin & Steinberg (eds), *Varieties of Dependence: Ontological Dependence, Grounding, Supervenience, Response-Dependence.* Philosophia Verlag

Fischer, John Martin. 1986. "Introduction: Responsibility and Freedom" in *Moral Responsibility,*

John Martin Fischer (ed.). Ithaca and London: Cornell University Press.

_____. 2004. "The Transfer of Nonresponsibility" in *Freedom and Determinism*, Michael O'Rourke & David Shier (eds.). Cambridge, MA: MIT Press.

_____. 2006. "The Cards that are Dealt You" *Journal of Ethics* 10: 107-129.

_____. 2011. "The Zygote Argument Remixed" *Analysis* 71: 267-272.

Fischer, John Martin & Mark Ravizza. 1998. *Responsibility and Control: An Essay on Moral Responsibility*. Cambridge: Cambridge University Press.

Frankfurt, Harry. 2002. "Reply to John Martin Fischer" in S. Buss and L. Overton, eds., *Contours of Agency: Essays on Themes from Harry Frankfurt*. Cambridge, Mass.: MIT Press.

Hasker, William. 1999. *The Emergent Self*. Cornell University Press.

Kim, Jaegwon. 2005. *Physicalism, or Something Near Enough*. Princeton: Princeton University Press.

Levy, Neil and Michael McKenna. 2009. "Recent Work on Free Will and Moral Responsibility" *Philosophy Compass* 4: 96-133.

Lewis, C.S. 1996. *Miracles*, rev. ed. New York: Touchstone.

List, Christian and Peter Menzies. Forthcoming. "My brain made me do it: The exclusion argument against free will, and what's wrong with it" in H. Beebee, C. Hitchcock & H. Price (eds.), *Making a Difference*. Oxford University Press.

Lowe, E.J. 2010. "Non-Cartesian Substance Dualism" in *The Waning of Materialism* (Bealer and Koons, eds.). Oxford: Oxford University Press.

Malcolm, Norman. 1968. "The Conceivability of Mechanism" *The Philosophical Review* 77: 45-72.

McDaniel, Kris. 2017. *The Fragmentation of Being*, Oxford University Press.

McKenna, Michael. 2008. "A Hard-Line Reply to Pereboom's Four-Case Manipulation Argument" *Philosophy and Phenomenological Research* 77: 142-159.

_____. 2012a. "Defending Nonhistorical Compatibilism" *Philosophical Issues* 22: 264-280.

_____. 2012b. "Moral Responsibility, Manipulation Arguments, and History: Assessing the Resilience of Nonhistorical Compatibilism" *Journal of Ethics* 16: 145-174.

_____. 2014. "Resisting the Manipulation Argument: A Hard-Liner Takes it on the Chin" *Philosophy and Phenomenological Research* 89: 467-484.

Mele, Alfred. 2005. "Pereboom's Four-Case Argument For Incompatibilism" *Analysis* 65: 75–80.

Merricks, Trenton. 2001. *Objects and Persons*. Oxford University Press.

Nagel, Thomas. 1986. *The View from Nowhere*. Oxford: Oxford University Press.

Nahmias, Eddy. 2014. "Is Free Will an Illusion? Confronting Challenges from Modern Sciences" in Sinnot-Armstrong (ed), *Moral Psychology, vol. 4: Freedom and Responsibility*. Cambridge: MIT Press.

O'Connor, Timothy. 2014. "Free Will and Metaphysics," in *Libertarian Free Will: Contemporary Debates*, edited by David Palmer, Oxford University Press.

Pereboom, Derk. 2001. *Living Without Free Will*. New York: Oxford University Press.

Plantinga, Alvin. 2011. *Where the Conflict Really Lies*. Oxford: Oxford University Press.

Ravizza, Mark. 1994. "Semi-Compatiblism and the Transfer of Nonresponsibility" *Philosophical Studies* 75: 61-93.

Rosen, Gideon. 2010. "Metaphysical Dependence: Grounding and Reduction" in Hale and Hoffmann (eds), *Modality: Metaphysics, Logic, and Epistemology*. Oxford University

Press.

Roskies, Adina. 2012. “Don’t Panic: Self-Authorship Without Obscure Metaphysics” *Philosophical Perspectives* 26: 323-342.

Schaffer, Jonathan. 2009. “On What Grounds What” in Manley, Chalmers, and Wasserman (eds), *Metametaphysics*. Oxford University Press.

_____. 2010. “Monism: The Priority of the Whole” *The Philosophical Review* 119, 1: 31-76.

Steward, Helen. 2012. *A Metaphysics for Freedom*. Oxford University Press.

Trogdon, Kelly. 2013. “An Introduction to Grounding” in Hoeltje, Benjamin & Steinberg (eds), *Varieties of Dependence: Ontological Dependence, Grounding, Supervenience, Response-Dependence.* Philosophia Verlag.

Turner, Jason. 2009. “The Incompatibility of Free Will and Naturalism” *Australasian Journal of Philosophy* 87 (4):565-587 (2009

van Inwagen, Peter. 1990. *Material Beings*. Cornell University Press.

Velleman, David. 1992. “What Happens When Someone Acts?” *Mind* 101: 461-481.

Wilson, Jessica. 2014. “No Work for a Theory of Grounding” *Inquiry* 57: 535-579.

Wilson, Jessica and Sara Bernstein. 2016. “Free Will and Mental Quasation”, *Journal of the American Philosophical Association* 2: 310–331.