A new puppet puzzle

Andrew M. Bailey[a]* and Joshua Rasmussen[b]

[a]Yale-NUS College, Singapore; [b]Department of Philosophy, Azusa Pacific University, Azusa, USA

We develop a new puzzle concerning a material being's relationship to the smallest parts of the material world. In particular, we investigate how a being could be responsible for anything if its behavior is completely determined by the behavior of those small parts. Many discussions of determinism and responsibility have focused on a determinism that moves from past to future. The determination at issue in our paper is importantly different: it moves from the bottom up.

Keywords: Moral responsibility; mereology; parthood; agency; materialism; priority

1. Introduction

" … 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" (Nagel 1986, 136). This article begins with a similar idea. But instead of reflecting on a chain leading without, we will focus on a chain leading within—all the way into our world's smallest parts. Let us begin with a rough and ready statement of the puzzle we have in mind:

> I and my world are made of very small parts. I haven't the faintest clue what's going on with those parts. Despite the fact that some of them dwell within my borders, they are beyond my ken, and I am not at all morally responsible for their activities. And yet everything I do is a consequence of what *they* do. So everything I do is the consequence of things for which I am not at all responsible. It is hard to resist the conclusion, then, that I'm not responsible for anything I do after all.

Much in the reasoning above has the feel of commonsense.[1] Yet that reasoning implies something that hardly anyone believes—namely, that I am a mere "puppet" of various small elements of the material world and so not responsible for my actions. Something must give.

We devote this article to developing and evaluating this "puppet" puzzle. We will first offer a more careful statement of the conundrum in the form of a valid deduction from theses that one could find independently plausible. Then, we will examine a promising solution that denies a bottom-up explanatory picture of composite objects. We close by identifying significant implications of denying the bottom-up picture. Our presentation invites an uneasy choice: (i) abandon *compositism*—the view that we are composed of proper parts, (ii) deny that we are responsible agents, or (iii) live with the implications of top-down determination.[2]

---

*Corresponding author. Email: wrathius@gmail.com

2.   The puppet puzzle stated

The Puppet Puzzle comprises six plausible but jointly inconsistent theses:

> Atomic Priority: If compositism about human persons is true, then there are atoms whose behavior necessitates and explains my behavior.
>
> Compositism: Compositism about human persons is true.
>
> Epistemic Condition: I am not responsible for facts about which I (non-culpably) know little to nothing.
>
> Ignorance: I (non-culpably) know little to nothing about facts about those atoms whose behavior necessitates and explains my behavior.
>
> Connection: if the A-facts necessitate and explain the B-facts, and I am not responsible for the A-facts, then I am not responsible for the B-facts.
>
> Responsibility: I am responsible for my behavior.

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

> A. Therefore, there are atoms whose behavior necessitates and explains my behavior. (from Atomic Priority, Compositism)
> B. Therefore, I am not responsible for my behavior. (from A, Epistemic Condition, Ignorance)
> C. Therefore, I am not responsible for my behavior and I am responsible for my behavior (B, Responsibility)

Two notes. First, the Puppet Puzzle is in the first person. We invite you to consider each of its theses from your own perspective; we think you'll see that the puzzle generalizes. Second, and as shown above, the six theses comprising the Puppet Puzzle are indeed inconsistent – and thus *cannot be jointly true*. One immediate consequence is that one shouldn't accept all six theses. At least one must be false. We do not present the above deduction, then, as an argument that is to be accepted. Quite the contrary: we are confident that it *cannot* be a sound argument and that one of its constituent theses must be false. But which? This is the heart of the matter, and is the central topic of this article. There is, we'll argue, no easy and obvious solution. Each thesis has some initial plausibility, and rejecting any comes at a significant price. This is precisely what makes the Puppet Puzzle a puzzle.

To show all this, let's have a closer look at Atomic Priority, Epistemic Condition, Ignorance, and Connection. We will not comment further on Responsibility and Compositism except to note that they are both widely considered extremely plausible. A great many philosophers maintain that we are morally responsible for what we do at least some of the time and that we are composed of proper parts.

Atomic Priority expresses an orthodox view about human persons. The materialist formulation of orthodoxy is familiar enough: humans are built of wholly material parts (brains, bones, lungs, cells, and so on), and the activity of those parts, perhaps in concert with the activity of atoms elsewhere in the material world, jointly explains the activity of the whole person.[3] We'll mostly focus on materialist versions of compositism, but we note that even non-materialist compositists (union dualists, for example) may here face puzzles too.

In a materialist framework, Atomic Priority falls out of the more general view that the activities of larger physical objects are explained and necessitated by the activities of smaller items. 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. Applying this idea to people: we do what we do because of what various material parts do.

We make no assumption about the kinds of things that, given compositism, compose us or our surroundings. They could be cells, strings, fields, or something else besides. Nor do we assume that there is a lowest level of decomposition. For ease of presentation, we'll talk of "atoms". This is shorthand for items at a very low level of decomposition – a level lower than that of the chemist's molecules, let us say – where we completely overlap at least some of those items. We do not assume that such atoms completely overlap us. Some may lie far outside our borders (more on this in a moment). 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 those atoms entail truths about the motion of that rock.

Regarding explanation, "x's behavior explains y's behavior" means this: y is doing what it does because x is doing what it does. We take ourselves to have a pre-philosophical understanding of the notion of "because" as it is used here.[4] More generally, we seem to readily understand what people mean when they say such things as "the kitten meowed because its tail was pulled"; or "the boat is sinking because there are too many people in it"; or "it is true that snow is white because snow is white"; or "you are in pain because certain neurons in your brain are firing a certain way"; and so on. We do not assume that the explanation in question is distinctively causal; nor do we assume that it is not. Further, we are not suggesting that the nature of explanation cannot be analyzed further. Indeed, much recent work in metaphysics is devoted to better characterizing the nature of explanation (see e.g. Audi 2012; Bennett and McLaughlin 2005; Correia 2008; Fine 2012; Rettler 2017; Rosen 2010; Schaffer 2009; Trogdon 2013).[5] For our purposes, we require just one assumption about explanation: where x and y are distinct, if *x because y*, then it is not the case that *y because x*. Explanation in the target sense is asymmetric (Schnieder 2011, 451).

Another note of clarification: 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 threaten responsibility independently of any threat posed by determinism. We will return to this point below.

For all Atomic Priority says, even if compositism is true, there could be *other* factors – in addition to certain atoms' behavior– that *also* explain and necessitate my behavior. Atomic Priority does not entail that "bottom up" explanations *exclude* other kinds of explanations. It is consistent, then, with Atomic Priority that compositism is true, that all my behavior is necessitated and explained by certain atoms' behavior, and that any of the following are also true: (a) my behavior is also necessitated and explained by my various psychological states; (b) my behavior is also necessitated and explained by various biochemical facts; (c) my behavior is also necessitated and explained by various social facts; (d) my behavior is necessitated and explained by facts about the world as a whole.[6]

Atomic Priority, furthermore, is neutral on *where* the relevant atoms may lie – and in particular, whether they must *all* fall within my borders. This is as it should be. That Serena is executing a serve is not just a matter of what's going on within Serena. It is

also a matter of whether she is playing tennis, what the rules of tennis are, and so on. For all Atomic Priority says, some of those atoms whose behavior explains and necessitates Serena's behavior fall outside of her boundaries – overlapping, for example, a ball and two racquets, an audience, and an opponent. What matters is *that there are such atoms*, and that at least some of them are her parts.[7] The key claim Atomic Priority makes is not an *individualistic* one, according to which in every case the whole is necessitated and explained by just *its* parts. Rather, it is a *directional* one, according to which there is always a necessitating explanation from "small to big".[8]

Epistemic Condition expresses one plausible and widely accepted necessary condition for being morally responsible. Responsibility requires (some kind of) control, but it also requires (some kind of) knowledge.[9] Though epistemic conditions on responsibility might outstrip those expressed here, it's plausible that they include at least some minimal knowledge of that for which one is responsible.

> "Non-culpably" is important. Consider reckless Frank, who *culpably* knows little or nothing about the dangers that beset young Dennis. Is Frank thereby not morally responsible for facts about the dangers Dennis faces? It doesn't seem so; ignorance doesn't seem to excuse. But where one is *non*-culpably ignorant of some facts, such ignorance would seem to remove or mitigate moral responsibility. If Frank were to be innocent in his ignorance of those facts involving Dennis, then Frank wouldn't be morally responsible for them.

We might wonder whether one could be responsible for unknown things that are *involved* in ordinary activity. For example, maybe you want to bake some bread, yet you don't know how exactly the dough rises or how the atoms in your fingers will move as you knead. In this case, perhaps you could be responsible for the resulting bread, even when the micro-physical processes are unknown. That is plausible. We find it significantly less plausible, however, that you are responsible for those unknown microphysical activities involved in baking. You are, rather, off the hook when it comes to the processes themselves precisely because you are innocently ignorant of them.[10]

One might insist here that knowing the intimate details of the micro-physical world is not required for moral responsibility. We would agree. But this observation isn't at war with Epistemic Condition. For all Epistemic Condition says, you are responsible for your behavior – baking some bread – even if you know little to nothing about micro-physical processes involving yeast, the development of gluten networks, and so on. To unseat Epistemic Condition, a rather stronger claim is required – that you are morally responsible for those microphysical processes themselves, even if you know little to nothing about them. And that stronger claim is, we think, significantly less convincing.

As we noted above, moral responsibility requires both control and knowledge. It is hard enough to find the relevant *control* for microphysical items about which you innocently know little to nothing. There is no obvious steering wheel. But even if there is, it is even harder to find *knowledge*. There isn't even a rear-view mirror to help you see what's behind everything you are doing.

We stress, finally, that commitment to principles like Epistemic Condition is commonplace and independent of considerations raised in this article. It would be surprising should it turn out that ordinary cases of agency – making bread without knowing the details of some yeasty process, say – showed these principles were false.

Ignorance has two components: a denial of knowledge, and a claim of non-culpability. You probably don't know much about, say, carbon molecules among your smaller parts. You may know even less about the electrons that compose the items in your immediate

vicinity. Sure, you know where they are, and you may know a bit about the powers they enjoy; but really, the goings-on of the smaller parts of the material world rarely occupy your mind. We think this much is fine, since you haven't violated any epistemic or moral duty in remaining ignorant. Your ignorance of such matters is non-culpable. Both components of Ignorance, then, appear quite plausible.

It is instructive to note one dimension of comparative modesty that the Puppet Puzzle enjoys. Cover and Hawthorne present an argument similar in spirit to ours, discussed further below. Their argument requires that "people don't have a choice about any of their microphysical details" (Cover and O'Leary-Hawthorne 1996, 59). The Puppet Puzzle requires nothing so strong. It doesn't even require that people aren't morally responsible for their microphysical details. Instead, our puzzle posits that people are non-culpably ignorant of their microphysical details. So, even if, for example, I enjoy free will over certain microphysical details (perhaps by freely moving hands, the motion of which requires certain microphysical changes), I might nonetheless be ignorant of those details.

Connection is one formulation of a so-called "Transfer Principle"; it says that non-responsibility "transfers" across relations of some sort – in this case, relations of necessitation and explanation (for more on such see Capes 2016).

One strategy compatibilists use to resist transfer principles cannot succeed against Connection. This point is worth developing at some length, for it will show that where other similar arguments falter, elements in the Puppet Puzzle may yet succeed. John Martin Fischer has argued against transfer principles like the following (themselves premises in arguments for incompatibilism about moral responsibility and determinism):

> Transfer: If no one is even partly morally responsible for the fact that p, and no one is even partly morally responsible for the fact that (if p, then q), then no one is even partly morally responsible for the fact that q.

Here is one of Fischer's counterexamples to Transfer:

> [Green] … walks along a beach and, noting that there is a child drowning, dives into the water and rescues the child. Though Green has had a device implanted in his brain, the device does not play any role in Green's decision to save the child (and his subsequent action). That is, the device monitors Green's brain activity but does not actually intervene in it. Let us suppose that this is because the scientists can see that Green is about to decide to save the child and to act accordingly. But let's also suppose that the scientists would have intervened to bring about a decision to save the child if Green had shown an inclination to decide to refrain from saving the child. That is, were Green inclined to decide on his own not to save the child, the scientists would ensure electronically that he decide to save the child and also that he act to carry out this decision (Fischer 1986, 41).

Fischer notes:

> … Green is not morally responsible for the fact that the scientists are ready to intervene, and he is not responsible for the fact that, if they are so ready, he will save the child. But he does seem to be morally responsible for saving the child (Fischer 1986, 60–61).

So, it seems, the same "Frankfurt-style" cases that famously undid the Principle of Alternative Possibilities also tell against Transfer (and thus arguments for the incompatibility of responsibility and determinism). Can Frankfurt-style cases also be used to unseat Connection?

No. In Frankfurt-style cases, a prior condition guarantees a target result without bringing about that result. The target result is necessitated but not explained by the prior condition. This is their odd charm: the target result occurs *independently* and not because of

the presence of the prior condition. Green, for example does not save the child because of the presence of the scientists. Green's good work is not explained by the presence of the scientists. But the presence of the scientists nonetheless guarantees that Green saves the child. A successful counterexample to Connection would be one in which (i) certain facts (A-facts) necessitate and explain certain facts (B-facts), and in which (ii) an agent is *not* responsible for the A-facts but *is* responsible for the B-facts. Frankfurt cases, by their very design, fail to exemplify condition (i). The A-facts (that the scientists are present) in Fischer's Green case, for example, may necessitate or suffice for the B-facts (that Green saves the child); but they *do not explain* the B-facts. There is good reason, then, to think that Frankfurt-style cases cannot succeed as counterexamples to Connection.[11]

We add one further observation: the Puppet Puzzle exposes a unique and unhappy mixture of morality and composition. Other philosophers have argued that important aspects of agency are incompatible with materialism.[12] But unlike the Puppet Puzzle, those arguments are entangled in complex controversies about agency and free will. 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 is dialectically limited because it 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 Turner all deploy controversial transfer principles that are subject to Frankfurt-style counterexamples. Ours is not so vulnerable. Third, Merricks and Turner independently argue that free will is incompatible with certain brands of materialism. They, too, raise valuable considerations. But, without further, controversial assumptions about the relationship between free will and moral responsibility, their results do not say whether moral responsibility is compatible with materialism. Our question about compositism more generally is of deep and independent interest, we think, even if it turns out that, say, moral responsibility does not require freedom. All of these authors, furthermore, target materialist assumptions; our Puppet Puzzle, by contrast, targets the slightly more general thesis of compositism (though that thesis is a consequence of many materialist views).[13] We conclude, then, that the Puppet Puzzle has a distinctive position in the contemporary literature and invites wide interest.

3.   Can compatibilism help?

The Puppet Puzzle is reminiscent of arguments for the incompatibility of determinism and either free will or moral responsibility (but 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. Two points are in order.

First: 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.[14] According to compatibilists, the conditions of responsibility can be met, and they can even be met if determinism is true.[15] 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 Epistemic Condition. There is reason to resist this temptation, however. For although compatibilists think moral responsibility is possible, they needn't think that it comes easily, or that it obtains in just any case, or even that it obtains in ordinary actual cases.

Compatibilists and incompatibilists alike agree that subjects in at least some cases (cases of direct manipulation, for example), are not morally responsible.

The case we've made for the constituent theses of the Puppet Puzzle (Epistemic Condition, for example) could, furthermore, be accepted by compatibilists and incompatibilists alike.

Second: as argued above, standard replies to transfer principles simply do not apply to Connection. One standard compatibilist reply to arguments for incompatibilism, then, does not solve the Puppet Puzzle. Two other standard compatibilist theories – "multiple past compatibilism" and "local miracle compatibilism" – target premises in incompatibilist arguments according to which the past and laws are fixed. But the Puppet Puzzle deploys no such premises and so it immune to those theories too.[16]

4. The prospect of top-down determination

The Puppet Puzzle is indeed a puzzle. Here is one way to think about its challenge: the conjunction of several independently motivated theses imply something deeply counterintuitive, that we are not responsible for any of our behavior. We can put our result another way: if we are responsible for some of our behavior, as the vast majority of people imagine, then one or more of the independently motivated theses is false. Those unwilling or unable to go without belief in moral responsibility are thus committed to rejecting one or more of these theses. We suspect that the most promising responses will involve rejecting Atomic Priority, a strategy we'll assess in this section.

Atomic Priority, recall, says this: if compositism 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 if this more general principle is also false:

> Bottom-Up (BU): For any composite x, there are some ys, such that x completely overlaps the ys, and the behavior of the ys necessitates and explains the behavior of x.[17]

BU says that necessitation and explanation proceeds from the bottom up. We could deny that. And then we'd be free to deny that our actions are necessitated and explained by the behavior of various atoms. There are two variations to consider:

> Weak independence: Sometimes what we do is independent of (not necessitated and explained by) the behavior of any atoms.
>
> Strong independence: Sometimes what we do is independent of (not necessitated and explained by) any atoms' behavior; and indeed, what we do in fact necessitates and explains what the relevant atoms do.

Both variations are interesting; let us focus, though, on what they have in common. They both deny Atomic Priority by denying BU.[18]

There are some puzzling consequences of denying BU, however. We briefly draw attention to three. First, if BU is false, then many philosophers are wrong in their basic picture of the material world, a picture according to which explanation proceeds from the bottom up. And it's not just that some philosophers turn out to be wrong. More importantly, a very natural view turns out to be false. For it is quite natural to think that the features of composite wholes supervene and depend on the properties of various smaller items among which are their parts.

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 does a composite whole have properties not necessitated and explained by those of various atoms? Similarly, the Special Top-Down Question asks: under what conditions do things join together to form a complex that enjoys top-down determination?[19] 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, or any others. 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 explain the switch from bottom-up explanation to top-down or the lack of cross-level explanation?

The problem here is that the difference between bottom-up and top-down determination seems to be a deep, categorical difference, and it is hard to see how a mere change in arrangement or spin or mass or any other physical property could explain a flip from bottom-up to top-down determination. Similarly, the gulf from a whole that enjoys bottom-up explanation to one that does not appears wide indeed. The Special Top-Down Question and the Special Independence Question may seem to be unanswerable even in principle.

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 (something composed of everything) is itself the most fundamental thing (Schaffer 2010). If Priority Monism is true, then our behavior is still determined 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. Thus, denying BU doesn't immediately get us out of the Puppet Puzzle. What is needed is top-down determination that stops with us. But why should top-down determination be so courteous?

It may be helpful to explicitly connect top-down determination with two other suspicious metaphysical phenomena:

> Non-redundant causal powers: Some wholes have causal powers not had by their parts, whether individually or in concert.
> Top-down causation: Some wholes cause changes in their parts, and those changes do not simply amount to some parts causing changes in other parts.

These are of a kind. First, they tend to elicit astonishment, suspicion, or puzzlement. Many doubt their very possibility or coherence, and even believers find them difficult to swallow. Helen Steward, for example, wonders " … how an animal could possibly bring about a movement or change in a part of its body in such a way that its causing that movement or change did not simply amount merely to *another* part or parts of its body causing that movement or change (and therefore being governed merely by the laws and principles pertaining to the behavior of those parts)" (Steward 2012, 225). Second, and perhaps more importantly, they each suggest instances of or special cases of the denial of BU. The suspicious phenomena entail that some composite whole has a feature not determined by the behavior of various atoms. They thus have a common price—the cost incurred by denying BU. The difficulties we raise for denying BU are of interest, then, not only to those looking for a way out of the Puppet Puzzle, but also to the proponents of non-redundant causal powers or top-down causation.

Seeing the connection between BU and these other suspicious metaphysical phenomena is helpful for two reasons. First, the threefold costs we identified with BU may give us a deeper explanation of what it is about the suspicious phenomena that is suspicious: the phenomenon is hard for many philosophers to accept because BU is hard to deny. We think the worries about denying BU that we've made explicit may well be lurking implicitly and in the background when, for example, philosophers find top down causation so distasteful. Second, and accordingly, one good way to make progress in understanding the suspicious phenomena would be to investigate the Special Independence and Special Top-Down questions. Thinking clearly about these questions may aid us, not only in evaluating the Puppet Puzzle, but also in evaluating the cogency or possibility of the suspicious phenomena.

So far, the metaphysics. There are *meta*metaphysical upshots too. For if the claims of this article are correct, matters of real human importance (including our status as morally responsible agents) closely track questions about the metaphysics of material objects (including the direction of priority, if any, between parts and wholes). This suggests an important corrective against those who doubt the value or importance of the latter but not the former. Metaphysics is an abstruse and strange pastime indeed, but it matters.

In sum, we have seen that bottom-up determination poses a special problem for composite views about human persons. We do not claim that the problem is unsolvable or that we must therefore abandon compositism.[20] The problem is interesting because it challenges our understanding of the relationship between people and the smallest elements of the material world. Our own sense is that the most promising solution will include the proposal that top-down explanations are possible. But this proposal leads to puzzles of its own, we've argued. We conclude, then, that those of us who wish to uphold the orthodox doctrine that we are composite and morally responsible beings have new work to do; the way forward is by no means obvious or without cost. The Puppet Puzzle is indeed puzzling.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Andrew M. Bailey is an Associate Professor of Humanities at Yale-NUS College and specializes in metaphysics and the intersection of philosophy, politics, and economics.

Joshua Rasmussen is an Associate Professor of Philosophy at Azusa Pacific University and specializes in metaphysics and philosophy of religion.

Notes

1. For some data on what the folk think about related issues, see Nahmias (2014).

2. Recent arguments against Compositism (typically used to undermine materialism, an extension we do not here endorse) include (Barnett 2010; Bogardus 2012; Collins 2011; Lowe 2010). For discussion see Bailey (2014).
3. The explanatory claim is not an obligatory component of materialism, of course; for argument to that effect, see Bailey (forthcoming a).
4. Such will do for our purposes. For a useful and detailed exploration of the logic of "because", see Schnieder (2011).
5. Wilson (2014) provides a helpful and necessary cautionary note.
6. Thanks to an anonymous referee for pressing this point. We say more about exclusion principles below.
7. Thanks to an anonymous referee for extremely helpful comments about various external explanations for human action and for the helpful tennis example.
8. One might wonder *why* necessitating explanations proceed from small to big in the way prescribed by Atomic Priority. Is there some deeper truth here about the big being grounded in the small or being constituted by the small? These are good questions. We will not answer them here. For our aim is to show that Atomic Priority, no matter *why* it holds, can generate a puzzle about our moral responsibility when conjoined with other plausible principle.
9. This claim goes back at least as far as Aristotle. See *Ethics* 1110a-1111b4. For recent discussion pro and con, see all the essays in Robichaud and Wieland (2017).
10. Thanks to an anonymous referee for this case and helpful criticism along these lines.
11. Similar remarks apply to other more straightforward overdetermination cases used against Transfer, such as those in (Ravizza 1994; Fischer and Ravizza 1998, 151–169; Fischer and Stump 2000; Fischer 2004).
12. (Cover and O'Leary-Hawthorne 1996; Merricks 2001, 155–161; Turner 2009) are 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 with 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 mental causation. But the resemblance is not penetrating. First, the Puppet Puzzle makes no plain use of the principles that drive Kim's argument. It does not deploy a causal exclusion principle, for example. Second, that principle is not, furthermore, lurking in the background in any clandestine way. For we have motivated the constituent theses of the puppet puzzle without it. The problem is not that one kind of explanation – from the bottom up, say – excludes another kind of explanation. Rather, the problem is that a plausible claim about ignorance and moral responsibility together with a transfer principle implies that we aren't responsible, on that bottom-up metaphysics of mind. Philosophers have, we note, affirmed principles like Transfer and Ignorance on grounds wholly independent of causal exclusion; so it is not plausible that exclusion principles ultimately drive Transfer or Ignorance. And so denying an exclusion principle, as one does when answering Kim's argument, is of no obvious help in resolving the Puppet Puzzle. A final difference between Kim's argument and the Puppet Puzzle is this: our puzzle concerns, not *mental causation*, but *moral responsibility*. Even if there is no problem of mental causation (because Kim' exclusion principles are false, for example) there may yet be a problem for moral responsibility. Our puzzle brings *that* problem into sharp relief. For helpful explorations of the Causal Exclusion Argument in relation to human agency, see List and Menzies (forthcoming) and Wilson and Bernstein (2016).
13. A final argument in this neighborhood is in Widerker (2016), which claims that free will rules out a certain kind of physicalism. See Bailey (forthcoming c) for objections to Widerker's argument.

14. <Possibly(p and q)> strictly entails both <possibly(p)> and <possibly(q)>; so what the compatibilist affirms—<possibly(someone is morally responsible and determinism is true)>—*strictly entails* <possibly(someone is morally responsible)>

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

16. For the curious: one of the authors accepts the compatibility of moral responsibility and determinism.

17. BU is reminiscent of, but stronger than, *priority pluralism*, according to which wholes are grounded in their parts. On priority pluralism see Bailey (2011).

18. 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 the 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.

19. These labels are 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 Independence" and "General Top-Down" questions, which ask, respectively, what independence (whether weak or strong) and top-down determination *are*.

20. For the curious: one of the authors affirms compositism about human persons, while the other is agnostic about it. Both agree that composition is not by any means without complications; see Bailey (2016). We elsewhere argue against theses in the neighborhood of Atomic Priority. See (Bailey and Rasmussen forthcoming; Bailey forthcoming b; Rasmussen 2015; Rasmussen 2018).

References

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

Bailey, Andrew M. 2011. "The Incompatibility of Composition as Identity, Priority Pluralism, and Irreflexive Grounding." *Analytic Philosophy* 52: 171–174.

Bailey, Andrew M. 2014. "You Needn't be Simple." *Philosophical Papers* 43: 145–160.

Bailey, Andrew M. 2016. "Composition and the Cases." *Inquiry* 59: 453–470.

Bailey, Andrew M. Forthcoming a. "Material Through and Through." *Philosophical Studies*.

Bailey, Andrew M. Forthcoming b. "Magical Thinking." *Faith and Philosophy*.

Bailey, Andrew M. Forthcoming c. "Freedom in a Physical World." *Philosophical Papers*.

Bailey, Andrew M., and Joshua Rasmussen. Forthcoming. "How to Build a Thought." *Thought*.

Barnett, David. 2010. "You are Simple." In *The Waning of Materialism*, edited by G. Bealer and R. C. Koons, 161–174. Oxford: Oxford University Press.

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

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

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

Capes, Justin. 2016. "Incompatibilism and the Transfer of Non-Responsibility." *Philosophical Studies* 173: 1477–1495.

Collins, Robin. 2011. "A Scientific Case for the Soul." In *The Soul Hypothesis*, edited by M. C. Baker and S. Goetz, 222–246. London: 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*, edited by Jeff Jordan and Daniel Howard-Snyder, 47–72. Lanham, MD: Rowman & Littlefield.

Fine, Kit. 2012. "Guide to Ground." In *Metaphysical Grounding*, edited by F. Correia and B. Schnieder. Cambridge: Cambridge University Press.

Fischer, John Martin. 1986. "Introduction: Responsibility and Freedom." In *Moral Responsibility*, edited by John Martin Fischer. Ithaca and London: Cornell University Press.

Fischer, John Martin. 2004. "The Transfer of Nonresponsibility." In *Freedom and Determinism*, edited by Michael O'Rourke, and David Shier. Cambridge, MA: MIT Press.

Fischer, John Martin. 2006. "The Cards That are Dealt You." *Journal of Ethics* 10: 107–129.

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

Fischer, John Martin, and Eleonore Stump. 2000. "Transfer Principles and Moral Responsibility." *Philosophical Perspectives* 14: 47–56.

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

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 *Making a Difference*, edited by H. Beebee, C. Hitchcock, and H. Price. Oxford: Oxford University Press.

Lowe, E. J. 2010. "Non-Cartesian Substance Dualism." In *The Waning of Materialism*, edited by G. Bealer and R. C. Koons, 439–461. Oxford: Oxford University Press.

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

Merricks, Trenton. 2001. *Objects and Persons*. Oxford: 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 *Moral Psychology, vol. 4: Freedom and Responsibility*, edited by Sinnot Armstrong. Cambridge: MIT Press.

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

Rasmussen, Joshua. 2015. "Building Thoughts From Dust: A Cantorian Puzzle." *Synthese* 192: 393–404.

Rasmussen, Joshua. 2018. "Against Nonreductive Physicalism." In *The Blackwell Companion to Substance Dualism*, edited by Jonathan L. Loose, J. L. Menuge, and J. P. Moreland, 328–339. Wiley-Blackwell.

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

Rettler, Bradley. 2017. "Grounds and 'Grounds'." *Canadian Journal of Philosophy* 47 (5): 631–655.

Robichaud, Philip, and Jan Willem Wieland, eds. 2017. *Responsibility: The Epistemic Condition*. Oxford: Oxford University Press.

Rosen, Gideon. 2010. "Metaphysical Dependence: Grounding and Reduction." In *Modality: Metaphysics, Logic, and Epistemology*, edited by B. Hale and A. Hoffmann, 109–135. Oxford: Oxford University Press.

Schaffer, Jonathan. 2009. "On What Grounds What." In *Metametaphysics*, edited by D. Manley, D. Chalmers, and R. Wasserman, 347–383. Oxford: Oxford University Press.

Schaffer, Jonathan. 2010. "Monism: The Priority of the Whole." *The Philosophical Review* 119 (1): 31–76.

Schnieder, Benjamin. 2011. "A Logic for 'Because'." *The Review of Symbolic Logic* 4 (3): 445–465.

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

Trogdon, Kelly. 2013. "An Introduction to Grounding." In *Varieties of Dependence: Ontological Dependence, Grounding, Supervenience, Response-Dependence*, edited by M. Hoeltje, B. Benjamin, and A. Steinberg. Philosophia Verlag.

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

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

Widerker, David. 2016. "A New Argument Against Libertarian Free Will?" *Analysis* 76: 296–306.

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.