Received: 10 July 2019 | Revised: 19 June 2020 | Accepted: 23 June 2020

DOI: 10.1111/mila.12343

SUBMITTED ARTICLE

WILEY

Experiments on causal exclusion

Thomas Blanchard[1] | **Dylan Murray** | **Tania Lombrozo**[2]

[1]University of Cologne, Cologne, Germany

[2]Princeton University, Princeton, New Jersey

Correspondence
Thomas Blanchard, Universität zu Köln, Philosophisches Seminar, Albertus-Magnus-Platz 1, 50923 Köln, Germany,
Email: tblancha@uni-koeln.de

Funding information
John Templeton Foundation

Intuitions play an important role in the debate on the causal status of high-level properties. For instance, Kim has claimed that his "exclusion argument" relies on "a perfectly intuitive … understanding of the causal relation." We report the results of three experiments examining whether laypeople really have the relevant intuitions. We find little support for Kim's view and the principles on which it relies. Instead, we find that laypeople are willing to count both a multiply realized property and its realizers as causes, and regard the systematic overdetermination implied by this view as unproblematic.

KEYWORDS

causation, exclusion problem, experimental philosophy, multiple realization, proportionality

1 | INTRODUCTION

Non-reductive physicalism holds that while everything that exists in spacetime is physical (or constituted by the physical), "higher-level" properties such as mental and biological properties are typically multiply realized by (and hence distinct from) physical or "lower-level" properties. While this view enjoys widespread support among philosophers, its implications for the causal status of higher-level properties are hotly debated.

Kim (1998, 2003) has famously argued that on non-reductive physicalism, the causal powers of lower-level properties *preclude* higher-level properties from being causally efficacious—call this claim *upward exclusion*. To illustrate, suppose that a higher-level property M is multiply realized by physical properties P_1 and P_2—that is, some instances of M are realized by P_1 while others are realized by P_2 (such that M is identical to neither P_1 nor P_2). And suppose that P_1

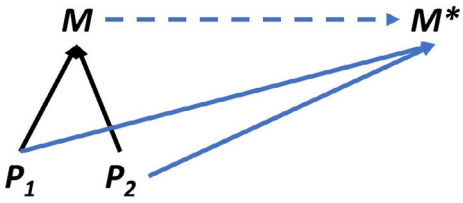


FIGURE 1 Blue arrows represent causal relations and black arrows supervenience relations. (For a colour figure see the online version) [Color figure can be viewed at wileyonlinelibrary.com]

and P_2 in turn individually cause another higher-level property M* (see Figure 1). The intuition behind upward exclusion is that the realizers of M do all the work of causing[1] any of its putative effects (such as M*), so that there is no causal work left for M to do; and admitting M as an additional cause of said effects would entail, implausibly, that effects of higher-level properties are systematically overdetermined. Kim's famous "exclusion argument" attempts to precisify this line of thought. On one of its versions,[2] the argument relies on two premises: The *non-overdetermination principle*, according to which effects of higher-level properties cannot be caused *both* by a property and the realizer of that property,[3] and the *causal inheritance principle* (Kim, 1993, p. 355), according to which any causal power that a higher-level property M has must be inherited from the causal powers of the physical property P that realizes M on a given occasion. The latter principle entails that if M causes M*, P must also be a cause of M*; from the non-overdetermination principle it then follows that M does not cause M*.

In response to Kim, non-reductive physicalists have sought to salvage the causal relevance of higher-level properties in either of two ways. *Conciliationists* (e.g., Bennett, 2003; Jackson & Pettit, 1990; Loewer, 2007) hold that when a higher-level property M has an effect M*, M* is caused *both* by M and its physical realizer, and they claim that the systematic overdetermination implied by this view is unproblematic. Others, such as List and Menzies (2009), offer more radical responses to Kim, arguing for *downward exclusion*—the view that in typical cases of multiple realization it is the lower-level (realizer) properties, and not the higher-level property, that are excluded from playing any causal role.

Appeals to intuition play an important role in debates about causal exclusion, especially in Kim's defense of upward exclusion. Indeed, Kim (1998) claims that upward exclusion "arises from the very notion of causal explanation and ... a perfectly intuitive and ordinary understanding of the causal relation" (p. 67). One reading of this claim is that according to Kim, the view that multiply realized properties are causally impotent is itself an intuitive position. Another, perhaps more plausible reading is that while upward exclusion may not itself be intuitive, it follows via the exclusion argument from causal principles that are themselves intuitive. When defending the premises of his argument, Kim does appeal to their supposedly intuitive status, claiming, for example, that causal inheritance is "highly plausible" (1998, p. 54) and that non-overdetermination is "virtually an analytic truth with not much content" (2003, p. 51).

[1]When we say that property A has certain effects, we mean that an event instantiating A causes those effects in virtue of instantiating A. And when we say that property A causes property B, we mean that an event that instantiates A causes an event instantiating B in virtue of instantiating A.

[2]As Zhong (2011) emphasizes, Kim's writings contain several versions of the exclusion argument. In Section 4, we discuss another version of the argument.

[3]More precisely, the principle states that "no single event can have more than one sufficient cause occurring at any given time—unless it is a genuine case of causal overdetermination" (Kim, 2003, p. 42). This effectively precludes an event's being caused both by a higher-level property and its realizer, as for Kim genuine overdetermination involves two *independent* causes producing an effect, and a higher-level property and its realizer are not independent.

Appeals to intuitive plausibility also appear in List and Menzies's (2009) defense of downward exclusion. While they do not treat downward exclusion as intuitive itself, they argue that it follows from a principle of *proportionality* requiring causes to be "proportional" to their effects—that is, specific enough to account for those effects but no more specific (see also Yablo, 1992; Woodward, 2010). One consideration they advance in favor of downward exclusion is that it fits intuitions about particular cases. Consider a pigeon conditioned to peck at red stimuli, and presented with a crimson stimulus (example from Yablo, 1992). Intuitively, the stimulus's redness, not its crimsonness, causes the pecking. This intuition is consistent with the principle of proportionality, according to which crimsonness is "too specific" to properly account for the effect.

So far, no attempt has been made to examine whether laypeople (i.e., non-philosophers) find these various positions—or the causal principles from which they allegedly follow—intuitively plausible. One aim of this paper is to begin to fill this gap. We report the results of three experiments investigating how laypeople judge the causal status of multiply realized higher-level properties. Of course, if a certain position conflicts with intuitive judgments, it does not follow that it should be rejected. It could still be defensible on grounds other than intuitiveness, for example, because it follows from a scientifically or philosophically respectable theory of causation. (In fact, List and Menzies's case for proportionality proceeds largely that way: Their main argument is that a proportionality requirement on causation fits with the scientifically and practically useful dictum that causes are difference-makers (2009, pp. 489–490)). Even so, an accurate understanding of laypeople's views on causal exclusion is valuable, particularly because it can illuminate the dialectical structure of the debate (e.g., by showing that a certain position on the problem conflicts with ordinary intuitions and so *must* be defended on other grounds).

A second aim of this paper is to contribute to an ongoing project in philosophy and psychology to identify the principles and concerns that govern people's choice of causal representations and explanations (Lombrozo, 2012, 2016). Out of all the variables that could be used to represent the causal structure of a situation, which variables do people tend to favor, and what cognitive and practical goals do these "variable choices" advance (Woodward, 2016)? One leading hypothesis (Lombrozo & Carey, 2006; Lombrozo, 2010) is that these choices are largely driven by the need to identify dependence relationships that are "exportable," that is, support prediction and control across a wide range of cases. This notion can be developed in potentially conflicting ways. A concern for exportability could favor generalizations that are more *general* in the sense that they have more instances—for example, "All cats have whiskers" rather than "All Siamese cats have whiskers" (see Johnston, Sheskin, Johnson, & Keil, 2018; Lombrozo, 2016). On the other hand, a concern for exportability could instead favor generalizations that are especially *stable* (Woodward, 2010) in the sense that they require few unspecified background conditions to hold.[4] (Moreover, as Blanchard, Vasilyeva, and Lombrozo (2018) argue, the notion of stability itself can be developed in two different ways—as favoring either *broad* generalizations that hold in many conditions, or generalizations that provide guidance in the sense that they build in their application conditions.)

Situations of inter-level causation (involving variables at different "levels") constitute a particularly telling test case for investigating these competing considerations. Consider again Figure 1, and suppose that an instantiation of P_1 (and M) causes an instantiation of M*. A preference for variables allowing for more general causal claims might lead one to privilege a description of the cause in terms of M rather than P_1 (as every instance of the $P_1 \rightarrow$ M* relationship also instantiates the M → M* relationship, but not vice versa), whereas a preference for stability might lead one to favor the alternative description, as the $P_1 \rightarrow$ M* relationship is in a

[4]Vasilyeva et al. (2018) provide empirical evidence of a preference for stable causal relationships.

certain respect more stable than the M → M* relationship.[5] Thus, examining laypeople's causal intuitions in these situation is especially interesting for the project of elucidating the concerns that drive people's practices of causal representation.

Prior work in psychology raises additional hypotheses that cases of inter-level causation can potentially adjudicate. First, a preference for lower-level explanations could emerge from a *reductive* preference. In the context of scientific explanations, Hopkins, Weisberg, and Taylor (2016) find that non-experts prefer reductive explanations (e.g., favoring neuroscientific explanations for psychological effects). This predicts a preference for describing the cause of M* as P_1 rather than M, regardless of whether M is uniquely or multiply realized. Second, Johnson and Keil (2014) propose a "level-matching heuristic," according to which people assume that "a cause and its effect will tend to be at the same level of the event hierarchy," such that a high-level effect will have a high-level cause, and a low-level effect a low-level cause. They find that when causal relationships between elements in a hierarchical structure are left unspecified, participants prefer causes at the same level as effects. This predicts a preference for taking M rather than P_1 as causing M*, regardless of whether M is uniquely or multiply realized.

A final introductory remark: Because philosophy of mind has been the primary forum for discussion of non-reductive physicalism, the debate about the status of higher-level causation has been concerned primarily with *mental* properties. But it is also noted in the literature that the relevant positions and arguments apply to any kind of higher-level property (e.g., chemical, biological, or social) multiply realized by the physical.[6] Specific challenges arise for investigating the mental that do not apply to other higher-level properties, especially the possibility that laypeople may be non-physicalists about (phenomenal) mental states (see for example, Gottlieb & Lombrozo, 2018; Knobe & Prinz, 2008; Sytsma & Machery, 2010). To avoid those issues, we present participants with materials concerning biological and geological properties, for which physicalism is plausibly assumed.

2 | EXPERIMENT 1

Experiment 1 investigated whether laypeople envision higher-level causation in line with Kim's view. We presented participants with vignettes in which a higher-level property M might be thought to cause a certain phenomenon M* (see Figure 2). Participants were assigned to one of

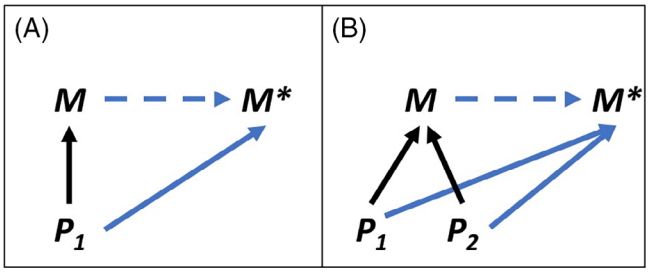


FIGURE 2 Setup of Experiment 1: (a) unique realizer condition; (b) multiple realizer condition [Color figure can be viewed at wileyonlinelibrary.com]

[5]To see why, let R_1 and R_2 be the ranges of background circumstances that must hold for (respectively) P_1 and P_2 to cause M*. Because M is equivalent to P_1-or-P_2, it is only if *both* R_1 and R_2 are actual that the M → M* relationship holds (at least for a deterministic relationship). In that sense, more background circumstances must hold for the M → M* relationship than for the P_1 → M* relationship, hence the latter is more stable.

[6]List and Spiekermann (2013) apply List's and Menzies's take on the exclusion problem to social properties.

two conditions. In the *unique realizer* condition, M is always realized by the same lower-level property P_1, which always leads to M* via the same physical mechanism. In the *multiple realizer* condition, M is realized either by P_1 or some other physical property P_2, where P_1 and P_2 each lead to M* via different physical mechanisms. All participants were then presented with a case in which some instance of M and P_1 was followed by an instance of M*, and they were asked to choose between two hypotheses—that M* was caused by M, and that M* was caused by P_1.

Kim's view is consistent with the claim that M is a cause of M* in the unique realizer condition (on the assumption that M is identical to P_1), but not in the multiple realizer condition (where M cannot be identical to P_1 or P_2). Thus, if non-philosophers find upward exclusion intuitive, participants should be more likely to judge that M* occurred because of P_1 than because of M in the multiple realizer condition. And participants should not be more likely to judge that M* occurred because of M in the multiple than in the unique condition.

2.1 | Method

2.1.1 | Participants

Sixty-eight participants (35% female, 65% male, mean age = 34, age range 19–72) were recruited on Amazon Mechanical Turk and paid $0.50 for participating. An additional 12 were excluded for failing comprehension checks (explained below). In all experiments, participation was restricted to users with an IP address within the United States and an approval rating ≥95% based on ≥50 tasks.

2.1.2 | Materials, design, and procedure

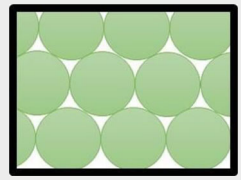
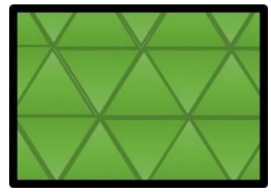
Participants were placed in the role of a scientist studying either a kind of plant ("yorgis") or a kind of rock ("kehlins") on a fictional planet. We illustrate with yorgis, but the structure of the experiment was the same in both conditions (see Table 1). Participants read that when certain lizards ingest yorgis, they get "arteritis," a harmless condition involving the formation of small tears on the linings of the lizards' arteries. Participants read about the molecules yorgis are made of and the mechanism(s) by which they produce arteritis. The information differed across two conditions, the *unique realizer* and the *multiple realizer* conditions ("unique" and "multiple" for short).

In the unique condition, all yorgis have the same molecular constitution and cause arteritis via the same mechanism. Participants read that their colleagues recently discovered that "all yorgis are of the same kind," and are "entirely made of small, round and heavy molecules called *alpha-mol*ecules."[7] Participants were also presented with a picture of such molecules and read the following information about why lizards get arteritis when they ingest a yorgi:

> When a lizard ingests a yorgi, stomach acid dissolves the plant, and the alpha-molecules are released in the bloodstream. As they travel with the bloodstream, alpha-molecules repeatedly bump against the linings of the arteries. Because alpha-

[7]In Experiment 1, the text that participants read omitted hyphens after "alpha" and "beta." These hyphens were included in subsequent experiments.

TABLE 1 Materials for Experiments 1 and 2 [Color table can be viewed at wileyonlinelibrary.com]

	Item	Yorgis	Kehlins (Exp. 1)	Kehlins (Exp. 2)
	Effect	Lizards get arteritis when ingesting yorgis.	Kehlins bounce when dropped.	Kehlins emit a constant buzzing sound.
Unique realizer condition	Lower-level realizer and mechanism	All yorgis are made of *alpha-molecules* (pictured below). Alpha-molecules are heavy and bump against artery linings. 	All kehlins are made of *psi-particles*. Their bouncing is produced by psi-particles behaving like coil springs.	All kehlins are made of *psi-particles*, which undergo constant internal vibrations that produce sound.
Multiple realizer condition	Lower-level realizers and mechanisms	Yorgis are made of either *alpha-molecules* or *beta-molecules* (pictured below). Beta-molecules are pointy and create small tears in artery linings. 	Kehlins are made of either *psi-particles* or *theta-particles*. Theta-particles are electrically charged, and when kehlins hit the ground their electrical discharges cause bouncing.	Kehlins are made of either *psi-particles* or *theta-particles*. Theta-particles produce sound when impinged on by air.

molecules are heavy, these bumps distend the linings and create small cracks in them.

In the multiple condition, yorgis are multiply realized: they can be made of two different kinds of molecules, each causing arteritis via a different mechanism. In this condition, participants were told that "there are two kinds of yorgis: *alpha*-yorgis and *beta*-yorgis." Participants read that alpha-yorgis are made of alpha-molecules, and were given the same information as the unique condition regarding why lizards develop arteritis when they ingest alpha-yorgis. Participants also learned that beta-yorgis are made of beta-molecules, which are triangle-shaped and have sharp, pointy edges, and that:

> When a lizard ingests a beta-yorgi, stomach acid dissolves the plant, and the beta-molecules are released in the bloodstream. As beta-molecules travel with the bloodstream, their edges often bump against the linings of the arteries. Because the edges of beta-molecules are pointy, these bumps create small tears in the artery linings.

In both conditions participants further learned that:

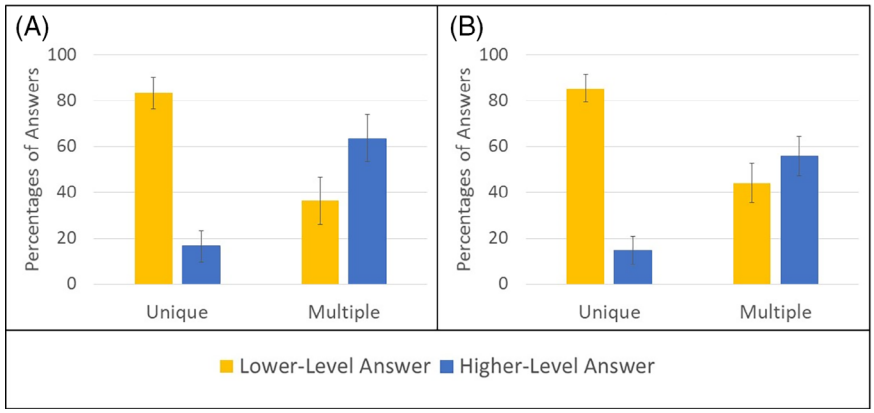


FIGURE 3 Answers to the causal question in Experiment 1: (a) among participants who endorsed supervenience; (b) among all participants. Error bars represent the standard error of proportion in either direction [Color figure can be viewed at wileyonlinelibrary.com]

> Your research assistants recently brought back a yorgi to the lab. They have labeled it "Y38." This yorgi is made of alpha-molecules. One night, a lizard crawls into the lab and eats Y38. Sure enough, the lizard has arteritis the next day.

After this information, participants were asked which of two causal statements they most agreed with (multiple choice, only one response allowed): that "Y38 caused arteritis because it is a yorgi" or that "Y38 caused arteritis because it is made of alpha-molecules." We refer to this question as the *causal question*, and to the two possible answers as the *higher-level* and *lower-level* answers, respectively.[8]

Participants were further asked a *supervenience question*: whether they agreed that "A yorgi is nothing more than an assemblage of alpha-molecules" (unique condition) or that "A yorgi is nothing more than an assemblage of alpha or beta molecules" (multiple condition). This was intended to gauge whether participants shared the assumption—needed for the exclusion problem to arise—that the relevant higher-level property (being a certain kind of plant) supervenes on lower-level properties (being made of certain types of molecule).

Finally, participants answered two true/false comprehension questions about the scenario (e.g., "According to what you have read, eating yorgis causes lizards to shed their skin".) Participants who answered either incorrectly were excluded.

2.2 | Results

We first tested for differences across vignettes (yorgis and kehlins). As they did not differ in Experiment 1 or 2, we report results collapsed across them.

Our main question was whether participants' responses would align with Kim's upward exclusion view. We restricted our initial analyses to participants who answered the

[8]A reviewer notes that since the answers to the causal question contain the word "because," they may probe participants' intuitions not about what *caused* arteritis, but about what best *scientifically explains* it. But for the purpose of our hypotheses these explanatory judgments are germane because they indicate that Y38 caused the lizard's arteritis because the former instantiated a certain property. This mirrors the way the exclusion problem is routinely presented in the literature—namely, as the question whether an event causes an outcome in virtue of instantiating a certain higher-level property or a certain lower-level property.

supervenience question positively ($N = 52$, 76.5% of participants), and hence are plausibly regarded as endorsing a physicalist view of the higher-level kinds or properties described in the vignette (see Figure 3a). In the multiple condition, participants were no more likely to select the lower-level than the higher-level answer ($p = .361$). In the unique condition, by contrast, they were significantly more likely to select the lower-level answer ($p < .001$). A chi-square test revealed a significant difference across conditions, $\chi^2\ (1) = 12.076$, $p = .001$, with a higher proportion of lower-level responses in unique than in multiple. Notably, this relationship is in the reverse direction to that consistent with Kim.

We also considered whether responses to the supervenience question varied as a function of condition. A significantly higher percentage of participants endorsed supervenience in unique than in multiple (88% versus 65%), $\chi^2\ (1) = 5.231$, $p = .022$. However, even pooling all participants, we find that participants in multiple were significantly more likely to choose the higher-level causal answer than participants in unique, $\chi^2\ (1) = 12.621$, $p < .001$ (see Figure 3b). Again, responses differed from chance in unique ($p < .001$), but not in multiple ($p = .627$).

2.3 | Discussion

Experiment 1 revealed a significant effect of whether a higher-level kind is multiply realized on people's judgments of that higher-level kind's causal relevance, but one contrary to what Kim would predict: we found that higher-level kinds are *more* likely to be accepted as causes of an effect when they are multiply realized compared to when they are uniquely realized.

3 | EXPERIMENT 2

While Experiment 1 provided evidence that people are not upward exclusionists, this evidence was not entirely conclusive. Because participants faced a forced choice between the higher-level and lower-level answers (without the option of choosing both), participants' choice of a specific answer gave no information as to whether they regarded the other answer as also correct, and hence whether they regarded the causal relevance of one property as *incompatible* with the causal relevance of the other. One goal of Experiment 2 was to address this shortcoming. We thus asked participants to choose between *three* possible answers: that the *higher-level* but not the lower-level kind caused the effect; that the *lower-level* but not the higher-level kind caused the effect; or that *both* caused the effect. This allowed us to assess participants' views about whether causation at the lower level *excludes* causation at the higher level.

In addition, the majority of participants who endorsed the higher-level causal answer in the multiple condition of Experiment 1 might have done so because they endorsed downward exclusion, or because they regarded both the higher- and lower-level kind as causally relevant to the effect but also considered the higher-level answer preferable, perhaps based on considerations of breadth or level matching. In Experiment 2, downward exclusion and conciliationism were presented as alternative answers, allowing us to disentangle which view participants find more intuitive. Finally, as noted above, Kim is more plausibly read as claiming intuitive status primarily for the premises of his exclusion argument (not necessarily its conclusion). A final advantage of this design is that it allowed us to investigate whether participants do find these premises (non-overdetermination and causal inheritance) intuitive.

Our second goal was to investigate an issue that regularly arises in discussions of exclusion—namely, that the status of higher-level causation may depend on one's theory of causation.

Philosophers standardly distinguish between two broad families of views. On *difference-making* views, A causes B if the occurrence of B depends on the occurrence of A (usually understood as *counterfactual* dependence). On *production* views, causation consists in the existence of a spatio-temporally continuous process from A to B (perhaps involving a physically conserved quantity). Several philosophers (e.g., Loewer, 2002) have argued that the problem of exclusion arises only on production views; on dependence views, higher-level causation is unproblematic insofar as the effect clearly counterfactually depends on the higher-level kind's occurrence. Likewise, List and Menzies (2009) challenge Kim's non-overdetermination principle on the grounds that, on a difference-making approach (which they regard as superior to a production approach), the principle fails to hold in a number of situations (including typical cases of multiple realization). Indeed, they argue that on a difference-making view, *down*ward exclusion is the natural stance. While Kim insists that his argument goes through whatever view of causation one endorses (Kim, 1998, p. 66), he acknowledges that the production view is in the background of the argument (Kim, 2002, p. 675). If a tight link does exist between upward exclusion and production, this suggests several hypotheses and questions about the results of Experiment 1. Specifically, the results might have been driven by a dependence view of causation, or by a causal concept that incorporates considerations of dependence and production (e.g., Hall, 2004; Lombrozo, 2010). If so, then participants may still endorse upward exclusion when instead thinking about causation in terms of production. To investigate these possibilities, we varied the language in which the causal question was presented by formulating it either in terms of difference-making or production.

3.1 | Method

3.1.1 | Participants

Three hundred and twenty-two participants (47% female, 51% male, 1% unspecified, mean age = 36, age range 19–69) were recruited on Amazon Mechanical Turk and paid $0.60 for participating. An additional 87 were excluded for failing comprehension checks like Experiment 1's.

3.1.2 | Materials, design, and procedure

Differences from Experiment 1 were as follows. First, participants were assigned to one of three language conditions, in which the causal question was formulated in terms of difference-making, production, or simply "cause." To illustrate with yorgis, participants were asked to evaluate whether Y38 either *made a difference to*, *caused*, or *physically produced* the arteritis because it is a yorgi or because it is made of alpha-molecules. Second, participants were told that two colleagues, Alice and Bob, were debating whether Y38 made a difference to (or caused, or produced) arteritis because it is a yorgi or because it is made of alpha-molecules (see Table 2). Participants were then asked to choose one of three statements (presented in randomized order): that Alice's answer is true but Bob's is not, that Bob's is true but Alice's is not, or that both answers are true. This framing allowed us to evaluate whether participants think the causal influence of either the higher-level or lower-level kind *excludes* that of the other.

TABLE 2 Causal question in Experiment 2 across language conditions

Difference-making	Cause	Production
Your colleagues Alice and Bob are debating the following question: Why did Y38 *make the difference* to whether or not the lizard got arteritis? Alice's answer is that Y38 made the difference to whether or not the lizard got arteritis *because it is a yorgi*. Bob's answer is that Y38 made the difference to whether or not the lizard got arteritis *because it is made of alpha-molecules*. In your view, which of Alice's and Bob's answers is true? (A) Alice's answer is true, but Bob's answer is not (B) Bob's answer is true, but Alice's answer is not (C) both Alice's and Bob's answers are true	Your colleagues Alice and Bob are debating the following question: Why did Y38 *cause* arteritis in the lizard? Alice's answer is that Y38 caused arteritis in the lizard *because it is a yorgi*. Bob's answer is that Y38 caused arteritis in the lizard *because it is made of alpha-molecules*. In your view, which of Alice's and Bob's answers is true? (A) Alice's answer is true, but Bob's answer is not (B) Bob's answer is true, but Alice's answer is not (C) Both Alice's and Bob's answers are true	Your colleagues Alice and Bob are debating the following question: Why did Y38 *physically produce* arteritis in the lizard? Alice's answer is that Y38 physically produced arteritis in the lizard *because it is a yorgi*. Bob's answer is that Y38 physically produced arteritis in the lizard *because it is made of alpha-molecules*. In your view, which of Alice's and Bob's answers is true? (A) Alice's answer is true, but Bob's answer is not (B) Bob's answer is true, but Alice's answer is not (C) Both Alice's and Bob's answers are true

Participants also typed a few sentences in a text box explaining their choice, to verify that they understood the task.

Third, we made changes to our second vignette prompted by the introduction of the language condition (see Table 1): We changed it so that the effect was not a state of the kehlin but the state of another object. Specifically, in Experiment 2, the effect was the fact that on a particular occasion, one of the participants' colleagues heard a buzzing sound when she put O27 to her ear.

3.2 | Results

The most noteworthy result is that conciliationism was overwhelmingly preferred in all conditions (see Table 3 and Figure 4). Participants who endorsed supervenience ($N = 214$, 66.5% of participants; see Table 3a and Figure 4a) were significantly more likely than chance to choose conciliationism rather than upward or downward exclusion, both in unique (χ^2 (1) = 53.156, $p < .001$) and in multiple (χ^2 (1) = 24.022, $p < .001$). The same is true if we consider all participants (unique: χ^2 (1) = 78.274, $p < .001$, multiple: χ^2 (1) = 35.333, $p < .001$). Conciliationism thus appears to be the default view in our sample.

We now examine whether participants were more likely to endorse Kim's view in multiple, and whether they were even more likely to do so when the causal question was formulated in terms of production. We first consider participants who endorsed supervenience.[9] To examine whether the effect of realizer on endorsement of upward exclusion was moderated by the language

[9]In contrast to Experiment 1, participants in Experiment 2 did not endorse supervenience significantly more in the unique compared to multiple realizer condition, χ^2 (1) = .751, $p = .386$.

TABLE 3 Percentages of answers to the causal question in Experiment 2 among: (a) participants who answered supervenience question positively; (b) all participants

	Dependence		Cause		Production	
	Unique	Multiple	Unique	Multiple	Unique	Multiple
(a)						
Upward exclusion	9.3	22.6	19.4	10.0	21.1	12.9
Downward exclusion	2.3	22.6	0.0	20.0	5.3	9.7
Conciliationism	88.4	54.8	80.6	70.0	73.6	77.4
(b)						
Upward exclusion	1.9	21.2	2.0	20.4	5.1	18.9
Downward exclusion	7.5	19.2	19.6	13.0	18.6	7.5
Conciliationism	90.6	59.6	78.4	66.6	76.3	73.6

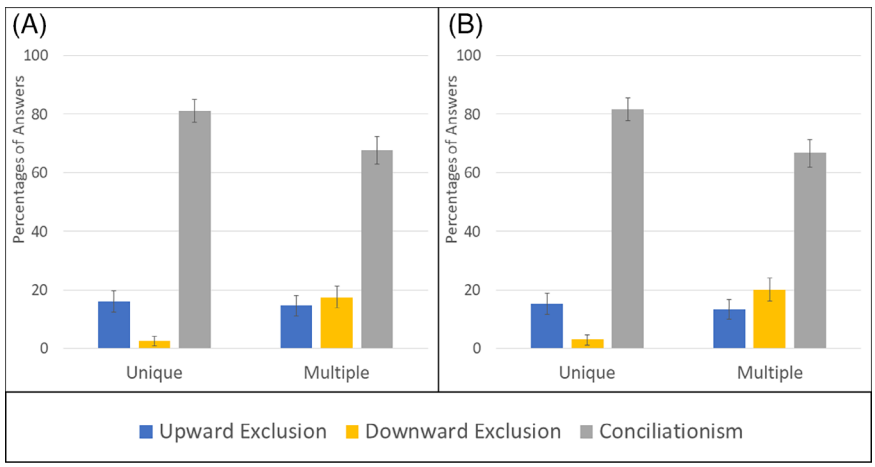


FIGURE 4 Answers to the causal question in Experiment 2, collapsed across language conditions: (a) among participants who answered the supervenience question positively; (b) among all participants. Error bars represent the standard error of proportion in either direction [Color figure can be viewed at wileyonlinelibrary.com]

of the causal question, we collapsed participants' answers to that question into a variable with two values (one representing endorsement of upward exclusion, and another representing endorsement of downward exclusion or conciliationism) and conducted a binary logistic regression on that variable using realizer, causal language, and an interaction term as predictors.[10] The resulting model did not significantly improve over the intercept-only model, χ^2 (5) = 4.855, p = .434: Participants were not more likely to endorse upward exclusion in multiple than in unique when the question was formulated in terms of production. Indeed, the trend was in the other direction, although not significant (see Table 3a). To test for a main effect of realizer condition on endorsement of upward exclusion, we conducted a further logistic regression omitting the realizer *x* language interaction.

[10]This strategy—collapsing two categories of answer into a single one—is the one we generally adopted to analyze the results of Experiments 2 and 3. This partially helps to avoid the issue that our data in Experiments 2 and 3 contain cells with zero or very small *n*s, making the fit of the models yielded by multinomial regressions on those data uncertain.

The resulting model was also not significant, χ^2 (3) = .375, p = .945; the realizer variable had no significant effect on endorsement of upward exclusion (p = .809).

We found similar results when considering all participants, including those who rejected supervenience (see Table 3b and Figure 4b). A logistic regression using realizer, language, and realizer *x* language as predictors yielded a model that did not significantly improve over the intercept-only model, χ^2 (5) = 7.581, p = .181. Likewise for the model yielded by a logistic regression using only realizer and language as predictors, χ^2 (3) = .775, p = .856.

We turn now to results concerning downward exclusion (considering participants who endorsed supervenience first). Our first question was whether participants were more likely to endorse downward exclusion in multiple than in unique. Unfortunately, a logistic regression to test for an interaction could not be conducted, as one cell (endorsement of downward exclusion in the unique realizer *x* cause condition) was empty. A further logistic regression on endorsement of downward exclusion among participants who endorsed supervenience using only realizer and language as predictors yielded a model that significantly improved over the intercept-only model, χ^2 (3) = 15.422, p = .001, and explained 15% of the variance (Nagelkerke R^2). The realizer condition was the only variable that had a significant effect: Participants were more likely to endorse downward exclusion in multiple than in unique, Exp (B) = 7.911, B = 2.067, p = .001. (A similar result holds for the corresponding logistic regression including all participants: The relevant model was significant, χ^2 (3) = 25.361, p = .002, and explained 15% of the variance, with participants more likely to choose downward exclusion in multiple than in unique, Exp (B) = 7.985, B = 2.078, p < .001.)

3.3 | Discussion

These results shed light on participants' preference for the higher-level answer in the multiple compared to the unique realizer condition observed in Experiment 1. They reveal that while most participants are willing to regard *both* the higher-level and the lower-level property as causally relevant, even when the latter is only one of multiple realizers, participants are also significantly more likely to endorse downward exclusion when the higher-level property is multiply compared to uniquely realized. These results also provide further support for the conclusion suggested by Experiment 1: that Kim's view finds little support among laypeople (even when the problem is formulated in terms of physical production). Most importantly, these results speak to the intuitive status of the premises of the exclusion argument itself: They suggest that non-overdetermination enjoys little intuitive support among laypeople, as the overwhelming majority grant causal efficacy to both the higher- and lower-level property in the multiple realizer condition, as they do in the unique realizer condition. The fact that significantly more participants endorse downward exclusion in the multiple than in the unique condition (though still not as many as endorse conciliationism) also casts doubt on the intuitive plausibility of the causal inheritance principle, since downward exclusion is incompatible with a higher-level property inheriting causal powers from its lower-level realizer(s).

4 | EXPERIMENT 3

Our goal in Experiment 3 was to examine a different—and arguably stronger—form of the exclusion argument. As Zhong (2011) observes, one drawback of the argument already

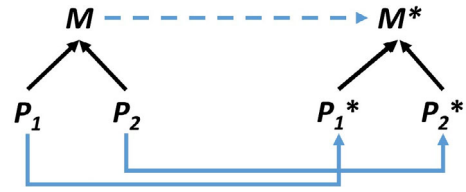


FIGURE 5 Blue arrows represent causal relations and black arrows supervenience relations (For a colour figure see the online version) [Color figure can be viewed at wileyonlinelibrary.com]

presented is its reliance on the causal inheritance principle, which fits uneasily with dependence views of causation. To see this, note that in the configuration of Figure 1, M* would not be instantiated if M was not instantiated, so on a dependence view the latter is a cause of the former. But it is not necessarily the case that if P_1 (or P_2) were not instantiated, M* would not be instantiated—perhaps the other realizer would be instantiated, in which case M* would still be instantiated. In that case, M* would depend on neither P_1 nor P_2. So dependence views do not support causal inheritance.

But as Zhong further notes, Kim's writings contain another version of the argument that relies not on causal inheritance, but on a principle of "downward causation," according to which a property A can cause another B (on a certain occasion) only by causing the realizer of B instantiated on that occasion. To see how this version of the argument works, consider Figure 5, where both M and M* are multiply realized, and in which each realizer of M causes one (and only one) of M*'s realizers. Given that P_1 causes P_1* (and P_2 causes P_2*), non-overdetermination entails that M cannot cause P_1* (or P_2*). By downward causation, this means that M cannot cause M* either. (The fact that the higher-level effect is multiply realized is important here. If M* was uniquely realized by a single physical property P*, dependence theorists could reject the assumption that P_1 and P_2 cause P* on the same ground that they reject causal inheritance: If one of the realizers did not occur, perhaps the other would occur, and P* would still occur.)

This version of the exclusion argument is preferable to that presented in the introduction because at least some dependence views of causation—including Lewis's (1973) influential theory—are committed to the principle of downward causation. On Lewis's view, if *c* and *e* are actual events, *e* depends on *c* just in case *e* would not have happened if *c* had not happened. This entails that if an instantiation of M* depends on (and hence is caused by) an instantiation of M, then so does the instantiation of the property that realizes M* on that occasion. If M* would not have been instantiated if M had not been instantiated, necessarily the relevant realizer of M* would not have been instantiated either. So Lewis's view endorses downward causation.[11]

In Experiment 3, we turn to this stronger version of the exclusion argument and the principles that underlie it.[12] Participants encountered structures in which the higher-level cause and

[11]Not all dependence views endorse downward causation, however. For *e* to depend on *c*, not only must it be true that *e* would not occur if *c* did not occur; the counterfactual "if *c* occurred, *e* would occur" must be true. For Lewis, if *c* and *e* are actual, the latter counterfactual is automatically true. But if one rejects this (as some dependence theorists do) one might reject downward causation. In a situation where M is realized by P_1 (and M* by P_1*), "If *M* had been instantiated, *M** would have been instantiated" may be true but "If *M* had been instantiated, P_1* would have been instantiated" false (perhaps M could have been realized by P_2, in which case P_2* would have been instantiated instead of P_1*).

[12]In Experiment 2, we found that the non-overdetermination principle enjoys little intuitive support, but only with respect to the version of the argument in the introduction, where the principle precludes *higher-level* effects from having two sufficient causes. In the version now under consideration, the principle precludes a *realizer* (physical) property from having two sufficient causes. The results of Experiment 2 suggest that participants are willing to accept overdetermination for higher-level properties, but do not show that they accept overdetermination for lower-level properties.

effect were both multiply realized, and we asked them to identify not only the cause(s) of M*, but also the cause(s) of its realizer.

4.1 | Method

4.1.1 | Participants

A total of 218 participants (41% female, 58% male, 1% unspecified, mean age = 32, age range 19–65) were recruited on Amazon Mechanical Turk and paid $0.70 for participating. An additional 53 were excluded for failing comprehension checks as above.

4.1.2 | Materials, design, and procedure

The experiment differed from Experiment 2 as follows. We included only the yorgi vignette and changed it so that in the multiple condition, *both* yorgis and arteritis are multiply realized. In the multiple condition, participants were told that there are two kinds of arteritis: "distension-arteritis," which consists in a distension of artery linings, and "hole-arteritis," which consists in the formation of tiny holes in artery linings. Participants learned that their lab had recently discovered "the mechanisms by which each of these types of arteritis arise when a lizard eats a yorgi." Yorgis made of round and heavy alpha-molecules are associated with distension-arteritis, the mechanism of which was described as in Experiments 1 and 2. Yorgis made of triangle-shaped and so sharp and pointy beta-molecules, however, are associated with hole-arteritis, the mechanism of which was described as follows:

> When a lizard ingests a yorgi made of beta-molecules, stomach acid dissolves the plant, and the beta-molecules are released in the bloodstream. As beta-molecules travel with the bloodstream, their edges often bump against the linings of the arteries. Because the edges of beta-molecules are pointy, these bumps create tiny holes in the artery linings.

In the unique condition, participants read that arteritis was recently discovered to consist in the formation of small holes in the arteries, and that all yorgis are made of beta-molecules. In both conditions, participants read that Y38—a yorgi made of beta-molecules—was recently brought to the lab, and that a lizard developed tiny holes in its artery linings upon ingesting it.

Third, we added a further question (see Table 4). While participants were asked to answer the same causal question as in Experiment 2 (except that their colleagues were now named "Antonia" and "Brandon"), they were also asked to evaluate a disagreement between two other colleagues (Alice and Bob) about whether the lower-level effect (the formation of holes in the lizard's artery linings) was caused by the higher-level property, the lower-level property, or both (with answers presented in randomized order). We refer to this question as the *lower-level* causal question (and to the question about what caused arteritis as the *higher-level* causal question). Participants were presented with the higher-level and lower-level questions in random order and, for each, given the option to type a few sentences explaining their choice, to verify they understood the task.

TABLE 4 Lower-level effect question in Experiment 3 across language conditions

Difference-making	Cause	Production
Your colleagues Alice and Bob are debating the following question: Why did Y38 *make the difference* to whether or not the lizard got tiny holes in his artery linings? Alice's answer is that Y38 made the difference to whether or not the lizard got tiny holes in his artery linings *because it is a yorgi*. Bob's answer is that Y38 made the difference to whether or not the lizard got tiny holes in his artery linings *because it is made of alpha-molecules*. In your view, which of Alice's and Bob's answers is true? (A) Alice's answer is true, but Bob's answer is not (B) Bob's answer is true, but Alice's answer is not (C) Both Alice's and Bob's answers are true	Your colleagues Alice and Bob are debating the following question: Why did Y38 *cause* tiny holes in the lizard's artery linings? Alice's answer is that Y38 caused tiny holes in the lizard's artery linings *because it is a yorgi*. Bob's answer is that Y38 caused tiny holes in the lizard's artery linings *because it is made of alpha-molecules*. In your view, which of Alice's and Bob's answers is true? (A) Alice's answer is true, but Bob's answer is not (B) Bob's answer is true, but Alice's answer is not (C) Both Alice's and Bob's answers are true	Your colleagues Alice and Bob are debating the following question: Why did Y38 *physically produce* tiny holes in the lizard's artery linings? Alice's answer is that Y38 physically produced tiny holes in the lizard's artery linings *because it is a yorgi*. Bob's answer is that Y38 physically produced tiny holes in the lizard's artery linings *because it is made of alpha-molecules*. In your view, which of Alice's and Bob's answers is true? (A) Alice's answer is true, but Bob's answer is not (B) Bob's answer is true, but Alice's answer is not (C) Both Alice's and Bob's answers are true

4.2 | Results

As in Experiment 2, participants were overwhelmingly more likely to prefer the conciliationist answer to the higher-level causal question than either of its alternatives (see Table 5 and Figure 6). Participants who answered the supervenience question positively ($N = 149$, 68% of participants) were more likely than chance to endorse conciliationism in both unique (χ^2 (1) = 43.200, $p < .001$) and multiple (χ^2 (1) = 8.926, $p = .003$). The same pattern holds among all participants (unique: χ^2 (1) = 43.293, $p < .001$; multiple: χ^2 (1) = 11.921, $p < .001$). Conciliationism thus remains the intuitive view even when both higher-level cause and effect are multiply realized.

Our main question in Experiment 3 was whether participants were more likely to endorse upward exclusion in response to the higher-level causal question in the multiple than in the unique condition. We first consider participants who endorsed supervenience. The model yielded by a logistic regression on endorsement versus non-endorsement of upward exclusion in response to the higher-level causal question using realizer, language, and an interaction term as predictors was not a significant improvement over the intercept-only model, χ^2 (5) = 8.637, $p = .124$. As in Experiment 2, the effect of the realizer condition on endorsement of upward exclusion was not significantly moderated by the language in which the question was formulated. A further regression to test for the main effect of realizer using realizer and language as predictors also did not significantly improve over the intercept-only model, χ^2 (3) = 3.254, $p = .071$. Although participants tended to select upward exclusion more in multiple than in

TABLE 5 Percentages of answers to the higher-level causal question in Experiment 3 among: (a) participants who answered supervenience question positively; (b) all participants

	Dependence		Cause		Production	
	Unique	Multiple	Unique	Multiple	Unique	Multiple
(a)						
Upward exclusion	13.8	30.4	23.8	16.0	5.9	29.4
Downward exclusion	0.0	17.4	4.8	20.0	5.9	5.9
Conciliationism	86.2	52.2	71.4	64.0	88.2	64.7
(b)						
Upward exclusion	16.2	30.6	33.3	24.3	10.3	27.3
Downward exclusion	0.0	22.2	5.6	18.9	5.1	6.1
Conciliationism	83.8	47.2	61.1	56.8	84.6	66.6

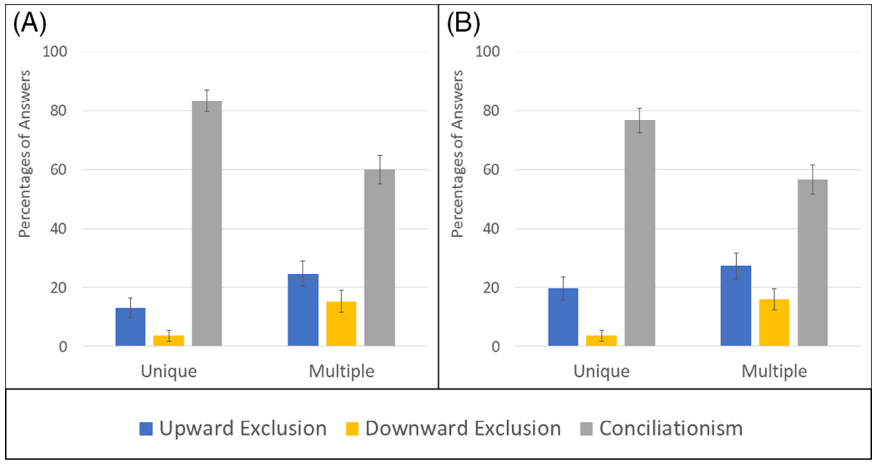


FIGURE 6 Answers to the higher-level causal question in Experiment 3, collapsed across language conditions: (a) among participants who answered the supervenience question positively; (b) among all participants. Error bars represent the standard error of proportion in either direction [Color figure can be viewed at wileyonlinelibrary.com]

unique, this trend was not significant ($p = .074$). (Similar results hold if we include all participants; see Table 5b and Figure 6b). A regression on endorsement of upward exclusion using realizer, language, and their interaction did not yield a significant improvement over the intercept-only model, χ^2 (5) = 8.719, $p = .121$, nor did a regression using only realizer and language, χ^2 (3) = 4.009, $p = .261$.)

We now consider participants' endorsement of downward exclusion in response to the higher-level causal question. As in Experiment 2, a regression to test whether the effect of realizer on endorsement of downward exclusion was moderated by the language of the causal question was not possible, as one cell (endorsement of downward exclusion in the unique realizer *x* dependence condition) was empty. As Figure 6 shows, there was a trend in favor of downward exclusion in multiple versus unique, both among participants who endorsed supervenience and among all participants. However, considering only participants who endorsed supervenience,

TABLE 6 Percentages of answers to the lower-level causal question in Experiment 3 among: (a) participants who answered supervenience question positively; (b) all participants

	Dependence		Cause		Production	
	Unique	**Multiple**	**Unique**	**Multiple**	**Unique**	**Multiple**
(a)						
Upward exclusion	10.3	65.2	19.0	64.0	17.6	70.6
Downward exclusion	0.0	0.0	0.0	8.0	0.0	5.9
Conciliationism	89.7	34.8	81.0	28.0	82.4	23.5
(b)						
Upward exclusion	10.8	69.4	30.6	62.2	20.5	72.8
Downward exclusion	2.7	0.0	0.0	8.1	0.0	3.0
Conciliationism	86.5	30.6	69.4	29.7	79.5	24.2

the model yielded by a regression using realizer and language as predictors did not significantly improve over the intercept-only model ($p = .062$), though the effect of realizer was significant, Exp (B) = 4.456, $B = 1.514$, $p = .028$. If we consider all participants, a regression using the same variables as predictors did yield a model that significantly improved over the intercept-only model, χ^2 (3) = 12.367, $p = .006$. The realizer condition was the only significant predictor, with participants significantly more likely to endorse downward exclusion in multiple than in unique, Exp (B) = 1.627, $B = 5.089$, $p = .005$.

We now consider participants' answers to the *lower-level* causal question, considering participants who endorsed supervenience first (see Table 6 and Figure 7). A binary regression on participants' endorsement versus non-endorsement of upward exclusion in response to the lower-level causal question using realizer, language, and realizer x language as predictors yielded a model that significantly improved over the intercept-only model, χ^2 (5) = 42.856, $p < .001$, but the interaction was not significant ($p = .755$). A further regression using realizer and language as predictors also yielded a model that significantly improved over the intercept-only model, χ^2 (3) = 42.291, $p < .001$, and explained 34% of the variance (Nagelkerke R^2). The realizer condition was the sole significant predictor. Participants were more likely to endorse upward exclusion in multiple than in unique, Exp (B) = 11.085, $B = 2.406$, $p < .001$. Indeed, while a large majority of participants chose the conciliatory answer in unique, a majority chose the upward exclusion answer in multiple.[13] A similar binary logistic regression including all

[13]This pattern tells against two deflationary interpretations of the findings of Experiments 2 and 3. The first is that in multiple, participants read "because it is a yorgi" in Alice/Antonia's answer to the higher-level causal question as meaning "because it is an *alpha-yorgi*" (that is, as referring to the lower-level kind and not the higher-level one). In this case, Alice's answer would be identical with Bob's, and the fact that most participants endorsed both answers would tell against neither upward nor downward exclusion. But if so, we should expect participants to interpret Alice's answer to the *lower-level* question in Experiment 3 similarly, and hence to overwhelmingly choose conciliationism for that question in *both* the unique and multiple conditions. The second interpretation is that participants regarded one of the two answers to the higher-level causal question as literally false, but as close enough to the truth to be accepted. Yet, we find it difficult to see why, if this hypothesis is correct, participants did not also regard both of their colleagues' answers to the *lower-level* causal question as "true enough," in which case conciliationism should once again have been overwhelmingly preferred.

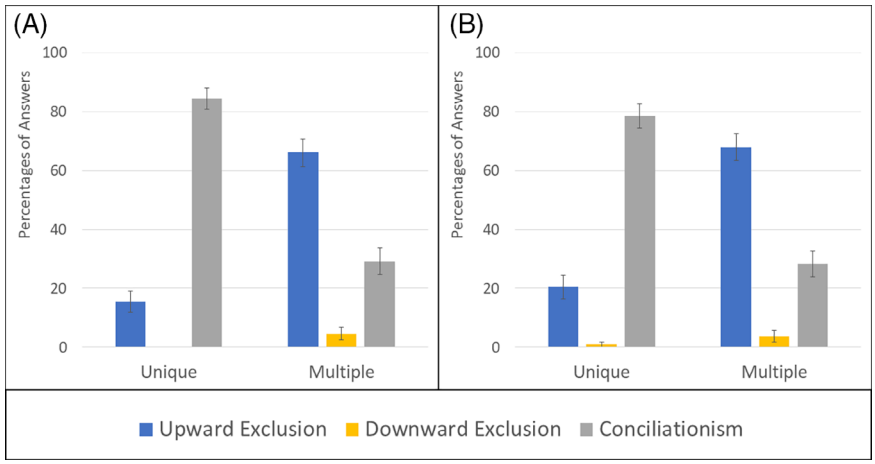


FIGURE 7 Answers to the lower-level causal question in Experiment 3, collapsed across language conditions: (a) among participants who answered the supervenience question positively; (b) among all participants. Error bars represent the standard error of proportion in either direction [Color figure can be viewed at wileyonlinelibrary.com]

participants also yielded a model that significantly improved over the intercept-only model, χ^2 (3) = 52.870, $p < .001$, with the realizer condition once again the sole significant predictor, Exp (B) = 8.318, $B = 2.118$, $p < .001$.[14]

This pattern of responses to the lower-level causal question is consistent with the hypothesis that participants tend to find the first premise of Kim's second exclusion argument—the non-overdetermination principle—fairly intuitive when applied to *lower-level* effects. What about downward causation? Among participants in the multiple condition who endorsed supervenience, those who accepted upward exclusion for the higher-level effect were also more likely to accept it for the lower-level effect, χ^2 (1) = 4.319, $p = .038$, $\varphi = .258$. The same pattern holds if we consider all participants, χ^2 (1) = 11.617, $p = .001$, $\varphi = .331$ (see Table 7). This is consistent with the hypothesis that participants who endorse upward exclusion for the higher-level effect do so partly because they find downward causation intuitive.

However, participants in the multiple condition who regard the higher-level property as a cause of the higher-level effect (i.e., those who endorse conciliationism or downward exclusion) were not more likely to give a similar answer to the lower-level causal question (conciliationism or downward exclusion) than to endorse upward exclusion, $p = .138$ ($p = .199$ among participants who endorse supervenience). In fact, the pattern is in the opposite direction: These participants tended to judge that only the antecedent lower-level property was a cause of the lower-level effect (although the trend is not significant; see Table 7). This suggests that overall, participants do not find downward causation particularly intuitive—that is, they are willing to count a property as a cause of a higher-level effect even if the property does not cause the realizer of that effect.

[14]Here also, a logistic regression using realizer, language, and realizer *x* language as predictors yielded a model that significantly improved over the intercept-only model, χ^2 (5) = 57.292, $p < .001$, but in which the interaction variable was not significant ($p = .115$).

TABLE 7 Numbers of responses to the higher-level and lower-level causal questions in the multiple condition in Experiment 3 among: (a) participants who answered supervenience question positively; (b) all participants

		Answer to lower-level causal question		
		Upward exclusion	**Downward exclusion**	**Conciliationism**
(a)				
Answer to higher-level causal question	Upward exclusion	14	0	2
	Downward exclusion	5	2	3
	Conciliationism	24	1	14
(b)				
Answer to higher-level causal question	Upward exclusion	27	0	2
	Downward exclusion	9	2	6
	Conciliationism	36	2	22

4.3 | Discussion

Experiment 3 suggests that the second version of Kim's exclusion argument fares no better than the first. We find no effect of unique realization on endorsement of upward exclusion when both higher-level putative cause and higher-level effect are multiply realized. The results also indicate that one of the two central premises of Kim's second exclusion argument—downward causation—enjoys little intuitive support among non-philosophers (although some participants' endorsement of upward exclusion for the higher-level effect may be driven in part by an endorsement of this principle). Interestingly, we do find intuitive support for the non-overdetermination principle, but only with respect to the lower-level effect; as in Experiment 2, most participants reject non-overdetermination with respect to the higher-level effect. This suggests that laypeople do not object to causal overdetermination of higher-level effects, even if they are leery of overdetermination at lower levels.

5 | GENERAL DISCUSSION

The implications of non-reductive physicalism for higher-level causation have been hotly disputed in the philosophy of mind and metaphysics. Much of the debate has centered on Kim's contention that multiply-realized properties are causally impotent—a contention which, according to him, emerges naturally from "a perfectly intuitive and ordinary understanding of the causal relation" (1998, pp. 66–67). The two versions of Kim's exclusion argument discussed in this paper attempt to make explicit how upward exclusion supposedly arises from intuitive causal principles—non-overdetermination, causal inheritance (in the first version of the argument), and downward causation (in the second version).

TABLE 8 Principles in Kim's exclusion arguments in relation to results of Experiments 1–3

	Experiment 1	Experiment 2	Experiment 3
Non-overdetermination A multiply realized property and its realizers cannot both be causes of an effect.	Conflicts	Conflicts	Conflicts when effect is high-level: Consistent when effect is low-level
Causal inheritance Higher-level properties inherit their causal powers from their realizers.	Consistent	Consistent	Consistent
Downward causation A causes B by causing B's realizers.	N/A	N/A	Conflicts

Our results suggest, contrary to Kim, that upward exclusion and the causal principles on which it relies enjoy little intuitive support. We failed to find any significant support for upward exclusion regarding higher-level effects—even in situations where the effect is multiply realized, and so which fall within the ambit of Kim's second and arguably stronger exclusion argument. We neither found intuitive support for upward exclusion itself, nor for the premises of Kim's exclusion arguments (see Table 8). The causal inheritance principle may be intuitive to laypeople, as the judgments of the majority of participants in all experiments are consistent with this principle. But causal inheritance yields upward exclusion (in the first version of Kim's argument) only when conjoined with non-overdetermination, according to which effects of higher-level properties cannot be caused both by a property and its realizer. And non-overdetermination does not seem to be intuitive to laypeople. In Experiment 1, we find that participants are not less likely to judge that the cause of a higher-level effect is a higher-level property when that higher-level property is multiply realized. And in Experiments 2 and 3, we find that most non-philosophers are perfectly willing to countenance the explicit causal overdetermination of a higher-level effect by a higher-level property and its realizers. Admittedly, the results of Experiment 3 suggest that laypeople are willing to reject overdetermination of a *lower-level* effect by a higher-level property and its realizer (and also that they regard the latter rather than the former as the cause). But this yields upward exclusion (in the second version of Kim's exclusion argument) only when conjoined with downward causation, and the results of Experiment 3 suggest that this principle enjoys at best limited support among non-philosophers.

At the opposite extreme, some philosophers have sought to defend higher-level causation by arguing that causal exclusion occurs downward rather than upward. This, according to List and Menzies (2009), follows naturally from a principle of proportionality requiring causes to vary one-to-one with their effects. One of their arguments for proportionality is its supposedly intuitive status, and we did find one effect consistent with the downward exclusionary view in Experiment 2. But the effect was small and the majority of participants still preferred conciliationism in all conditions. Indeed, our results suggest that most non-philosophers are willing to countenance lower-level realizers of a multiply realized property as causes of that property's effects, even though they are not proportional with those effects.

We hasten to acknowledge two limitations of our results regarding downward exclusion. First, our supervenience question was formulated in terms ("nothing more than") that some non-reductive physicalists—especially downward exclusionists—may reject, as it has hints of

reductionism. This strengthens our case against upward exclusion (since participants who endorsed this formulation of supervenience may be expected to favor a reductionistic view such as Kim's). But it weakens our findings against downward exclusion. Perhaps a weaker formulation of supervenience would lead to a stronger effect in favor of downward exclusion (though we find this unlikely, as the effect we found for downward exclusion was still small even when including all participants).

Second, in our vignettes the putative higher-level cause only had two realizers, and the mechanism linking these realizers to the effect was described in detail. People may be less inclined to regard lower-level properties as causes when a higher-level property has a large number of potential realizers, or when the mechanism linking lower-level realizers to the effect is not easily specifiable (as in, e.g., the causal relationship between increase in interest rates and decrease in inflation). This is an important open question for future research.

Our findings have further implications for the status of the proportionality principle: They suggest that causal judgments do not penalize all "non-proportional" causal factors equally. In Experiment 3, the majority of participants in the multiple condition refused to countenance the *higher-level* factor as a cause of the *lower-level* effect, in line with proportionality (as the occurrence of the higher-level factor is not systematically associated with the occurrence of the lower-level effect). But in Experiments 2 and 3, the majority of participants in the multiple conditions were willing to regard the lower-level realizer as causing the higher-level effect, despite the absence of proportionality (the alternative realizer still leads to the higher-level effect). This suggests that non-proportional causes may be penalized more when they are insufficient for the effect than when their absence is insufficient for the effect's absence. (Alternatively, perhaps all non-proportional causes are penalized equally, but for lower-level causes and higher-level effects other factors counterbalance this penalty.) This finding is potentially significant not only for List and Menzies's view, but for any view on which proportionality plays an important role in causal attribution, even if it is not a necessary condition for it (e.g., Woodward, 2010).

In addition to their bearing on proportionality, our findings have implications for the principles and concerns that underlie causal representation and explanation. Notably, they suggest that people value causal variables that stand in suitably broad relationships with their effects (explaining why participants are willing to regard higher-level factors as causes), but also that breadth is not the only concern governing causal attribution. In the multiple condition of Experiment 1, participants were equally likely to choose the higher-level and lower-level factors as causes, despite the latter's lower generality. It could be that breadth competes with a reductive preference, though there is little evidence of a global preference for lower-level attributions across our studies. Our findings here suggest that the strength and manifestation of effects of breadth and reduction may depend on how the explanandum is realized: People may be more inclined toward higher-level causes that capture broad regularities (of a sort that hold despite variation in implementation) in cases of multiple versus unique realization, but not necessarily to the exclusion of lower-level factors.

It is important to acknowledge several limitations of our work. Most notably, we solicited judgments from a limited sample using highly controlled stimuli. It would be valuable to solicit judgments from a more diverse sample, and with more varied materials. Still, the systematic pattern of results observed here suggests that our population had reliable intuitions. Even if our findings are limited to some populations under some conditions, they raise questions about the role of appealing to intuitions in drawing metaphysical conclusions about causation.

Perhaps Kim's exclusion argument can be reframed in terms of metaphysical principles that make no appeal to ordinary intuition, but as stressed in the introduction, this is not how the

argument has been portrayed in the literature. Insofar as the causal exclusion argument does rely on intuition, our results cast doubt on the claim that the non-overdetermination principle is "virtually an analytic truth with not much content" (Kim, 2003, p. 51), and they suggest that while upward exclusion may be "a perfectly intuitive and ordinary understanding of the causal relation" (1998, pp. 66–67) to Kim, it is not to most laypeople.

ACKNOWLEDGEMENTS

We thank the audience at the 2019 Workshop on Experimental Philosophy of Science at Aarhus University for helpful comments. This work was supported by the Varieties of Understanding Project, funded by the John Templeton Foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Templeton Foundation.

ORCID

Thomas Blanchard https://orcid.org/0000-0001-5002-7517

REFERENCES

Bennett, K. (2003). Why the exclusion problem seems intractable, and how, just maybe, to tract it. *Noûs*, *37*, 471–497. https://doi.org/10.1111/1468-0068.00447

Blanchard, T., Vasilyeva, N. & Lombrozo, T. (2018). Stability, breadth and guidance. *Philosophical Studies*, *175*, 2263–2283. https://doi.org/10.1007/s11098-017-0958-6

Gottlieb, S. & Lombrozo, T. (2018). Can science explain the human mind? Intuitive judgments about the limits of science. *Psychological Science*, *29*, 121–130.

Hall, N. (2004). Two concepts of causation. In J. Collins, L. Paul & N. Hall (Eds.), *Causation and counterfactuals* (pp. 225–276). Cambridge, MA: MIT Press.

Hopkins, E., Weisberg, D. & Taylor, J. (2016). The seductive allure is a reductive allure: People prefer scientific explanations that contain logically irrelevant reductive information. *Cognition*, *155*, 67–76. https://doi.org/10.1016/j.cognition.2016.06.011

Jackson, F. & Pettit, P. (1990). Program explanation: A general perspective. *Analysis*, *50*, 107–117. https://doi.org/10.1093/analys/50.2.107

Johnson, S. & Keil, F. (2014). Causal inference and the hierarchical structure of experience. *Journal of Experimental Psychology: General*, *143*, 2223–2241. https://doi.org/10.1037/a0038192

Johnston, A., Sheskin, M., Johnson, S. & Keil, F. (2018). Preferences for explanation generality develop early in biology but not physics. *Child Development*, *89*, 1110–1119. https://doi.org/10.1111/cdev.12804

Kim, J. (1993). *Supervenience and mind*. Cambridge, MA: Cambridge University Press. https://doi.org/10.1017/CBO9780511625220

Kim, J. (1998). *Mind in a physical world*. Cambridge, MA: MIT Press. https://doi.org/10.1017/S0031819100210103

Kim, J. (2002). Responses. *Philosophy and Phenomenological Research*, *65*, 671–680. https://doi.org/10.1111/j.1933-1592.2002.tb00231.x

Kim, J. (2003). *Physicalism, or something near enough*. Princeton, NJ: Princeton University Press. https://doi.org/10.1515/9781400840847

Knobe, J. & Prinz, J. (2008). Intuitions about consciousness: Experimental studies. *Phenomenology and the Cognitive Sciences*, *7*, 67–83. https://doi.org/10.1007/s11097-007-9066-y

Lewis, D. (1973). Causation. *Journal of Philosophy*, *73*, 556–567. https://doi.org/10.2307/2025310

List, C. & Menzies, P. (2009). Non-reductive physicalism and the limits of the exclusion principle. *Journal of Philosophy*, *106*, 475–502. https://doi.org/10.5840/jphil2009106936

List, C. & Spiekermann, K. (2013). Methodological individualism and holism in political science: A reconciliation. *American Political Science Review*, *107*, 629–643. https://doi.org/10.1017/S0003055413000373

Loewer, B. (2002). Comments on Jaegwon Kim's *Mind in a physical world*. *Philosophy and Phenomenological Research*, *65*, 655–662. https://doi.org/10.1111/j.1933-1592.2002.tb00229.x

Loewer, B. (2007). Mental causation, or something near enough. In B. McLaughlin & J. Cohen (Eds.), *Contemporary debates in philosophy of mind* (pp. 243–264). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199580781.001.0001

Lombrozo, T. (2010). Causal-explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions. *Cognitive Psychology*, *61*, 303–322. https://doi.org/10.1016/j.cogpsych.2010.05.002

Lombrozo, T. & Carey, S. (2006). Functional explanation and the function of explanation. *Cognition*, *99*(2), 167–204.

Lombrozo, T. (2012). Explanation and abductive inference. In R. Morrison & K. Holyoak (Eds.), *Oxford handbook of thinking and reasoning* (pp. 260–276). Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199734689.013.0014

Lombrozo, T. (2016). Explanatory preferences shape learning and inference. *Trends in Cognitive Sciences*, *20*, 748–759. https://doi.org/10.1016/j.tics.2016.08.001

Sytsma, J. & Machery, E. (2010). Two conceptions of subjective experience. *Philosophical Studies*, *151*, 299–327. https://doi.org/10.1007/s11098-009-9439-x

Vasilyeva, N., Blanchard, T. & Lombrozo, T. (2018). Stable causal relationships are better causal relationships. *Cognitive Science*, *42*(4), 1265–1296.

Woodward, J. (2010). Causation in biology: Stability, specificity, and the choice of levels of explanation. *Biology and Philosophy*, *25*, 287–318. https://doi.org/10.1007/s10539-010-9200-z

Woodward, J. (2016). The problem of variable choice. *Synthese*, *193*, 1047–1072. https://doi.org/10.1007/s11229-015-0810-5

Yablo, S. (1992). Mental Causation. *Philosophical Review*, *101*(2), 245–280.

Zhong, L. (2011). Can counterfactuals solve the exclusion problem?. *Philosophy and Phenomenological Research*, *83*(1), 129–147.

How to cite this article: Blanchard T, Murray D, Lombrozo T. Experiments on causal exclusion. *Mind & Language*. 2022;37:1067–1089. https://doi.org/10.1111/mila.12343