

# Intuitions and The Metaphysics of Causation

(forthcoming in *Experimental Metaphysics*, ed. David Rose, Bloomsbury)

Sara Bernstein

Should intuitions play a role in a theory of causation? Metaphysicians often bristle at the idea, while at the same time utilizing intuitions (“Intuitively, Billy rock is not the cause of the window’s shattering” or “Intuitively, Barack Obama’s failure to water my plant is not a cause of my plant’s death”) as evidence for or against their theories. Metaphysicians feel that whether or not  $c$  is a cause of  $e$  has nothing to do with our intuitions or mere thoughts about them, yet most metaphysical and causal theorizing centrally involves intuitions. Some metaphysicians recognize the relevance of intuitions to metaphysics more generally (e.g. Schaffer forthcoming, Paul 2010), whereas others (e.g. Ladyman and Ross 2007) see the use of such intuitions as exemplifying what’s wrong with contemporary metaphysics.

Some hold that that since intuitions seem to play a role in debates about causation, the nature, pervasiveness, and stability of these intuitions should be formally tested. Recently there has been a profusion of experimental work on folk intuitions about causation. Empirical data gathered by Knobe (2009), Hitchcock and Knobe (2009), and Henne et. al (forthcoming) shows that the folk concept of causation is deeply intertwined with, and effected by, normative considerations. Work by Halpern and Hitchcock (2015), Alicke, Rose, and Bloom (2011), and Clarke et. al (2015) show that judgments of blame, typicality, and norm-violation play a central role in folk causal judgments.

Those who gather data on folk intuitions rarely about causation can be viewed as falling into several camps about the dialectical role of such data. (They often do not explicitly self-identify with categories in this classificatory scheme—which is meant to clarify the intellectual terrain). According to one camp, intuitions, whatever they turn out to be, are irrelevant to the question of what causation really is. (This camp distinguishes between the question of what causation really is and the project of conceptual analysis.) Perhaps such intuitions are intrinsically interesting, this camp holds, but they have no bearing on a theory of causation of the sort that metaphysicians aim to develop. A second

camp remains carefully neutral on such issues, gathering data but making no methodological claims about how it should be used. (Much of the recent empirical work falls into this camp.) Another camp holds that empirical data about the folk concept of causation places pressure on metaphysical theories to comport with such data. Finally, the most radical camp holds that data on folk intuitions serve to debunk views which take such intuitions seriously.

Those in the last three camps face a greater explanatory burden than those in the first two camps. The carefully neutral camp owes an account of the philosophical utility of such data, if it is not to be incorporated into a theory of causation. And experimental philosophers who hold that theorists should do justice to intuitions in causal theories must provide reasons for such a view. It is not enough to lob data about intuitions in the direction of causal theorists. Rather, they must provide an answer to the question: *why* should such data be considered relevant to the project of what causation really *is*? This is a methodological question about the relationship between metaphysics and intuition, rather than a first-order question for theories of causation. And if metaphysicians are to take such intuitive data seriously, arguments must be provided for why theories must incorporate this kind of data.

Such arguments require understanding the range of goals undertaken by causal theorists, and how intuitive data might link up with such goals. This paper clears the ground on the relationship between metaphysical theories of causation and intuitions about causation, including those gathered by experimental philosophers. I will not argue for the rightness of certain theories nor the centrality of intuitions to metaphysics more generally; rather, I will isolate and clearly articulate the roles that intuitions can (and to some extent, already do) play in theories of causation. The goal will be to clarify several types of answers to the central question, to sketch adequacy conditions for these answers, and to clear the ground for further progress.

In section A, I clarify the questions at hand and I give a taxonomy of roles for intuitions in the metaphysics of causation, focusing on case studies of intuition-based problems for theories of causation. In section B, I sketch several such reasons that metaphysicians might be compelled to incorporate folk intuitions into their theories. In

section C, I lay out some obstacles to utilizing folk data in metaphysical theories of causation.

A few preliminaries will be helpful. I take empirical data about causation to constitute formal studies as performed by experimental philosophers and empirical psychologists about the operation and nature of the concept of causation. I take metaphysical theories of causation to be theories that partially or wholly aim at discovering what causation “really is”. For a metaphysical theory to “accord with” or “answer to” intuitions is, minimally, for it to not contradict a set of intuitions about how causation works.<sup>1</sup> Maximally, for data to accord with a theory is for that theory to provide an explanation of why causation works the way the folk concept dictates. These features are distinct from a metaphysics of causation which depends on intuitions: for example, a theory of causation which postulates that causation just is whatever the folk think it is.

#### A. Intuitions and Causal Theories: The Logical Space of Views

Inquiring about the role of intuitions in the metaphysics of causation encompasses several questions. One question is broadly methodological: should any metaphysics of causation answer to intuitions, and if so, why? A more specific version of the question is whether any particular theory of causation (for example, the counterfactual account of causation) should answer to intuitions. Finally, we might ask: should a causal theory answer to folk intuitions, as opposed to “philosophically trained” intuitions, i.e., the putatively more carefully considered intuitions of professionally trained philosophers?

Such questions often lurk in the background of metaphysical theories of causation, but are rarely directly addressed. Many metaphysicians aim to discover what causation “really is”, and yet the same philosophers blanch when their views are revealed to have counterintuitive commitments. Others use counterintuitive consequences as an objection to rival views. And still others openly commit themselves to counterintuitive claims, yet admit that these commitments are shortcomings of their theories.

Complicating matters is that causation is unique among metaphysical projects. It seems to enjoy a resistance to the methodological concerns that have plagued other ontological debates such as constitution and composition. These debates are vulnerable to

the suspicion that there is no fact of the matter about (e.g.) whether the statue is distinct from the lump of clay that constitutes it, or whether there is an object rather than particles-arranged-object-wise. Causation, in contrast, enjoys a more privileged status as a “science-y” metaphysical posit, both used and respected by many of our best scientific theories. Nonetheless, metaphysicians often aim to model a theory of causation which respects intuitions and can guide moral assessment, a goal sometimes at odds with the putative science-y nature of the posit.

Whether intuitions should play a role in theories of causation, and what that role should be, is partially a matter of the project in which a causal theorist is engaged. Only certain metaphysical explananda are amenable to the task. Consider the following non-exhaustive list of projects:

1. Strictly ontological. An objective, mind-independent causal relation in the world.
2. Conceptual analysis. A causal relation derived from a priori analysis of the causal concept.
3. Hybrid ontological. A causal relation in the world that accords with human intuitions about causation.
4. Scientifically posited. A causal relation posited by, and included in, the best complete physical theory of the world.
5. Scientifically plausible. A causal relation compatible with the best physical theory of the world.
6. Normative. A causal relation that tracks moral judgments relating to promises, ethical norm violations, and other moral concepts.

Each of these projects comes with an associated *success condition*, or condition under which the explanatory goal of the project is satisfied.

Recognizing distinctions between different projects is an essential starting point for structuring debates about the role of intuitions in studies of causation: if the project is strictly ontological, for example, objections based on counterintuitive theoretical commitments have no grip.<sup>2</sup> A project that aims to uncover whether the causal relation is fundamental likely has no use for intuitions. Accounting for intuitions is irrelevant to the

success conditions for strictly ontological accounts of causation.

A brief aside: here I set aside tricky questions about what the constraints on a theory of causation *are*. Suppose that a theory of causation doesn't obey intuitive constraints at all. For example, consider a theory that holds that *c* is a cause of *e* iff *c* and *e* both occur on a Thursday.<sup>3</sup> This is technically a theory of causation, but one that disobeys intuitive constraints so radically that it changes the subject entirely. Theories of causation thus presumably must obey some constraints in order to count as such. In this discussion, I will assume that we have a grip on what it is for a theory to be about causation in some sense, while denying or failing to comport with some intuitions about them—for example, counterfactual accounts of causation according to which one's birth is a cause of one's death.

Let us return to the topic of which projects are vulnerable to which objections. Projects primarily concerned with the operation of our concepts have reason to care about intuitions and their natures. These include conceptual analysis and projects aimed at uncovering intuitions. Unlike strictly ontological views of causation, conceptual analyses of causation need not consider theoretical virtues such as “elegance” and ontological parsimony. If it turns out that our concept of causation is irredeemably complex, so much the worse for those virtues, since conceptual analysis aims at accurate reflection.

What is confusing is that many projects are either intentionally or unintentionally *hybrid*: some projects which claim to be strictly ontological nonetheless take intuitions into account; other views which largely center around causal perceptions and intuitions nonetheless pay mind to what causation “really is”. An example will be helpful.

Lewis (2004) holds that causal theorizing aims at conceptual analysis, according to which theories unearth and make precise the tools and assumptions that competent users of the concept utilize. According to Lewis' famous “Canberra Plan”, we handle suspicious metaphysical posits by utilizing folk platitudes in order to find the best referents for the key terms. Given that Lewis thinks the job of theorizing is that of conceptual analysis, one might expect him to appeal to the Canberra Plan for causation as well. Roughly, the Canberra Plan proceeds in two steps. The first step involves the

creation of a Ramsey sentence, a statement that postulates the existence of the entities the world must have to have in order satisfy our best folk theories about it. The second stage checks the best theories of reality to see whether the referents exist. Intuitions play a central role in the Canberra Plan insofar as platitudes about metaphysics involve intuitive theorizing. The Canberra plan aims to link up intuitions about the world with what the world would have to be like in order to satisfy the intuitions.

Notably, however, Lewis suggests that that the plan does not apply to causation, for two reasons. First, the apparent heterogeneity of the causal relation renders it resistant to capture in a single homogenous relation. My tipping over the coffee cup causing it to spill, the mortgage crisis causing the recession, and the brain state causing the arm movement are all instances of causation. Any relation generated from the Canberra Plan, Lewis holds, would be unnatural and disjunctive. Second, causation by omission, which is part of the folk concept of causation, creates a problem of missing relata: an omission is not a relatum, and so cannot be united by a causal relation. There is considerable debate about whether the Canberra Plan can be applied to causation.<sup>1</sup> For our purposes, it is sufficient to note that the heterogeneity of concepts of causation poses an in-principle obstacle to a hybrid account of causation which seeks to account for intuitions about how the world is, whether or not such a hybrid account is Canberra Plannable.

Similarly, some causal theorists engaged in the causal modeling program aim both to uncover the “real” causal relation by testing interventions on causal paths in systems with objective, mind-independent features, while determining and modifying default and deviant variables in order to fit desired contexts and “normality” intuitions. According to this method, a causal scenario is represented by a formal equation whose “default” and “deviant” variables are set relative to one’s interests. A causal model predicts whether  $c$  is a cause of  $e$  given normality and typicality constraints. The hybrid causal modeling approach is endorsed by Hitchcock (2007), who writes:

“We can afford to let judgments of token causation be infected by pragmatic criteria without giving up on the objectivity of causation generally: objectivity can

---

<sup>1</sup> For one example, see Liebesman (2011) for an argument that the Canberra plan need not yield a relation.

be retained at the level of token causal structure.”

There is a fair bit of controversy over the nature of the causal modellers’ project. Some modellers are more of the projectivist variety, seemingly aiming to give formal

Even “energy transfer” theories of causation, according to which causation is a transfer of energy from cause to effect, may be viewed as straddling the divide between ontology and intuitions. On the one hand, these theories automatically rule out a large swath of intuitive causal instances: causation by omission. In such cases, an absence or failure of something to occur causes something else to occur. Absence causation is ubiquitous, and many take these cases to be Moorean facts for which theories must account. And yet energy transfers theorists clearly aim to account for the pretheoretic, intuitive “biff” causation at the center of folk intuitions about causation.

Another unclarity in the debate lies in how intuitive problems for metaphysical theories of causation are to be viewed: are they problems for intuitions, for the metaphysical theories that hold them, or for both? Consider what Menzies (2004) terms the problem of profligate omissions: accepting a simple counterfactual account of causation, according to which  $c$  is a cause of  $e$  if had  $c$  not occurred  $e$  would not have occurred, yields the result that *all* omissions count as causes. For example: suppose that I promise to water your plant, fail to water it, and the plant dies. The counterfactual “If I hadn’t failed to water your plant, the plant would not have died” is true. But the counterfactual “If the Queen of England hadn’t failed to water your plant, the plant wouldn’t have died” is also true. Intuitions, as well as well-established empirical data (Henne et. al (forthcoming), Clarke et. al (2015)) about such intuitions, supports the claim that only my failure to water the plant is causally relevant to the plant’s death.

But what kind of problem is the problem of profligate omissions? Exactly whom is it a problem for? Here are a few options:

- (1) A problem for the metaphysics of causation insofar as the metaphysics of causation should account for such intuitions.

(2) A problem for the metaphysics of causation insofar as metaphysical theories alone can't account for intuitive distinctions between omissions, and should.

(3) A problem about intuitions insofar as intuitions do not track the truth of counterfactuals or the reality of profligate omissive causation.

(4) A problem for neither intuitions nor metaphysics.

Each of these options incorporates a significant methodological assumption about the relationship between metaphysics and intuitions. Option (1) assumes that metaphysics *should* account for intuitions. But that assumes an answer to the very debate in question. There must be arguments or reasons why metaphysics should pay heed to such a thing. (2) makes the same assumption, but constitutes a *prima facie* positive reason for metaphysics to incorporate intuitions: intuitions can do philosophical work that metaphysics cannot. For example, if metaphysics alone cannot distinguish between intuitively irrelevant and relevant omissions, and only intuitions can do such work, then one might argue that intuitions do extra explanatory work that metaphysics cannot. Option (3) assumes that metaphysics should ignore intuitions, or at the very least, intuitions should be amenable to revision in light of metaphysical theories. And (4) is an assumption about the methodological independence of metaphysics and intuitions. Even this common problem in the causation literature is an instance of the methodological assumptions that infuse debates about causation.

It would be helpful to have an independent grip on whether intuitions are to be considered relevant or irrelevant to causal theories, or else metaphysicians and experimental philosophers are bound to continue developing philosophical projects in parallel rather than in tandem. I now turn my attention to this topic: what roles intuitions can play in the metaphysics of causation, and reasons for including such intuitions in metaphysical projects.

## B. Roles for Intuitions in Metaphysical Theories

One reason for taking intuitions to be relevant to theories of causation involves theory choice. Metaphysical debates which do not aim to incorporate intuitions have



fewer clear standards of adjudication than those that do. Theoretical virtues such as simplicity, elegance, and fertility only take us so far in choosing between theories that are equally empirically adequate. For example, let us suppose that the energy transfer theory of causation and the counterfactual theory of causation exhibit equal theoretical virtues: they are equally simple, equally elegant, and equally fertile. How should we choose between them? Pitting the strictly ontological theories against each other takes us only so far. Arguably, intuitiveness is already wielded as a theoretical virtue in debates among equally virtuous strictly ontological theories. Even in debates about what causation “really is”, intuitions often rear the heads when other dimensions of adjudication have been exhausted. Explicitly counting intuitiveness as a theoretical virtue provides an extra way to weigh causal theories against each other.

Here are two illuminating parallels in other metaphysical debates. First, consider the well-worn debate over whether particles-arranged-table-wise compose a table. Mereological nihilists hold that nihilism has the benefits of theoretical parsimony and elegance. Those who believe in composition believe that their views do more justice to intuitions (namely, intuitions like “There is a table there”) than nihilism. Nihilists such as van Inwagen respond by holding that paraphrases such as “particles-arranged-tablewise” do as much justice to our folk concept of the object as the world as “table”. And Dorr and Rosen (2002) go to great lengths to argue that mereological nihilism can still vindicate our ordinary picture of the world. Similarly, compatibilists and hard determinists about free will often agree on the causal history of a particular human action (for example, my arm being raised), but disagree on whether their views do justice to intuitions about what it takes to have free will. Formally adding intuitiveness to the many theoretical virtues often under consideration makes explicit considerations already implicitly in use in various metaphysical debates, including the debate over the nature of causation.

While some reasons to use intuitions are methodological, others have to do with the causal projects themselves. The first and most obvious instance of this is in causal projects that involve conceptual analysis. What conceptual analysis *is* is tricky, controversial business, but let us assume that conceptual analysis involves unmasking and making explicit the nature and behavior of a particular philosophical concept, including

its correct extension. Testing the nature of the causal concept deployed by competent users is an important reinforcement of conceptual analyses of causation, or else such theories threaten to reveal only what a particular theorist unearths about her own intuitions. Intuitions held by competent users of the concept affect the viability of causal theories for which conceptual analysis is central: if a particular theory does not accord with widespread intuition, then it is simply not a good analysis. Imagine how strange it would be to call a conceptual analysis “counterintuitive”. It’s almost a contradiction in terms. What it is for a theory to contain conceptual analysis is for that theory to accurately reflect and model a concept.

Intuitions can play an important role in solving problems for which metaphysical structure cannot fully do the job of making intuitive causal distinctions. In the problem of profligate omissions, for example, one might hold that intuitions are required for achieving a theory of causation according to which I am the cause of the plant’s death but the Queen of England is not. Here, intuitions play the role of “going the extra theoretical mile” in distinguishing causally relevant from irrelevant omissions. Similarly, one might hold that intuitions play a similar role in distinguishing causes from background conditions. For example: the airplane rather than the initial state of the universe caused my arrival in Paris.

Now, does this mean that such theories of causation incorporate or depend on such intuitive distinctions? Not necessarily. One plausible view of the relationship is that a metaphysical theory of causation lays out “what there really is”—say, profligate causation by omission or causation involving wide background conditions—while relying on empirically-backed intuitions to draw distinctions that metaphysics alone cannot. This is similar to what Hitchcock and Knobe have in mind, when they note:

It has long been recognized that ordinary causal judgments make use of information that goes beyond anything that might be included in causal structure. People seem to rely on extra-structural information to select certain candidate causes over others—claiming, for example, that the spark was a “cause” of the fire, while the oxygen was merely a background “condition.” (2009, p. 5)

Through intuitions, we can draw human-created but nonetheless philosophically

significant boundaries and distinctions for which metaphysical theories alone cannot account. And this need not conflict with the mission of hybrid ontological or scientifically plausible accounts of causation. The intuitions do not “create” the ontology; they serve as guides for distinctions made in addition to ontological models of the world. Lewis’ method is compatible with this approach:

We sometimes single out one among all the causes of some event and call it “the” cause, as if there were no others. Or we single out a few of the “causes,” calling the rest mere “causal factors” or “causal conditions.” Or we speak of the “decisive” or “real” or “principle” cause. We may select the abnormal or extraordinary causes, or those under human control, or those we deem good or bad, or just those we want to talk about. I have nothing to say about these two principles of invidious discrimination. I am concerned with the prior question of what it is to be one of the causes (unselectively speaking). (Lewis, 1973, pp. 558-559.)

Here, intuitions are to be utilized as distinction-creators and line-drawers. The metaphysical models are mind-independent, whereas the “invidious distinctions” are human-created.

This approach works particularly well with the numerous empirical studies that suggest a strong conceptual link between causation and normativity. Several studies have shown that whether or not the folk count  $c$  as a cause of  $e$  is shaped in by whether  $c$  was normatively required to bring about  $e$  (for example, she promised to water the plant, but didn’t; or Professor Smith was not supposed to take a pen, but did.) If norms and normality are mind-dependent, and whether or not  $c$  is a cause of  $e$  is bound up with these concerns, then intuitions can demarcate causal relevance where metaphysics cannot.

Certain first-order views of causation automatically incorporate intuitions. One example is causal idealism. According to causal idealism, causation is a matter of individual human projection of the causal relation. (Bernstein, forthcoming) More formally:  $c$  is a cause of  $e$  only if at least one observer mentally projects a causal relation between  $c$  and  $e$ . In the framework of causal idealism, the task of probing our intuitions is probing the process by which we project causation onto the world. Similarly, intuitions are clearly relevant to certain forms of contextualism about causation, according to which the truth conditions of causal claims are partially or wholly a matter of human-created

context. If context is a matter of convention and usage, then intuitions about such things is helpful for unearthing the truth conditions of causal claims.

Another view that naturally incorporates intuitions holds that we already implicitly know what causes what because platitudes of causation are analytic. Intuitions merely reflect this knowledge. As Nolan writes:

If I am confident that I competently deploy the concept of causation (and competently use the English word “cause”), I can be confident that I already know the platitudes about causation, at least implicitly, and all I need to do is whatever self-examination is required to make them explicit. [...] If I had good reason to believe both that the platitudes were analytic and that the analyticities were all things I implicitly knew, I could have methodological confidence in armchair philosophy that might otherwise be harder to come across. (Nolan 2009, p. 19)

According to this view, platitudes like “causation is intrinsic” are analytic, and intuitions reflect knowledge that we already have about causation. On this methodology, testing for intuitions about the meanings of causal terms can bolster the folk platitudes by providing evidence for their analyticity.

Intuitions are also relevant to debates in which convention, usage, and assertability play a central role, as in debates about special or strange causal counterfactuals. Consider the debate over counterpossibles, counterfactual conditionals with impossible antecedents. Examples of counterpossibles include “If  $2+2$  had equaled 5, geometry textbooks would not have been different” and “If Sara was a unicorn, she would have had curly hair.” The central debate over counterpossibles concerns their *vacuity*. Why think they are vacuous? Generally, two reasons are given. First, with counterpossibles, there are no  $c$  worlds in which  $e$  occurs, because there are no possible  $c$  worlds simpliciter. Second, if the antecedent is impossible, “anything goes.” For example, a world in which  $2+2=5$  doesn’t seem to rule out any sort of consequent: it might be plausible that a world in which mathematical laws are different is one in which squaring the circle causes trees to bloom. At stake is whether counterpossibles have non-vacuous content—a matter largely dependent on intuitions about counterpossibles.

There is strong intuitive evidence that not all counterpossibles are vacuous. In the example “If  $2+2$  had equaled 5, geometry textbooks would not have been different”, there

is clearly a strong intuition that this is false. (Geometry textbooks would have been different if mathematical laws were different.) And consider a causal counterpossible such as “If the mathematician hadn’t failed to prove that  $2+2=5$ , her advisor would not have remained unimpressed.”<sup>4</sup> Here, there is a clear intuition that failing to prove that  $2+2=5$  is a cause of the advisor’s nonplussed attitude. Recent empirical work by Ripley (ms) shows that the folk judge counterpossibles to be true and false rather than vacuous. The counterpossibles debate exemplifies a case in which gathering and analysis of folk intuitions reinforces an existing view in a metaphysical debate whose opposing sides already depend on intuitions. Many debates in which convention, usage, and acceptability are relevant to the truth of counterfactuals will follow this pattern.

To recap: intuitions can play several roles in metaphysical theories of causation. Methodologically, taking intuitiveness to be a theoretical virtue provides extra guidance about theory choice. Theories that partly involve conceptual analysis involve evaluating folk platitudes about causation, which largely draw on intuitions. Intuitions can draw distinctions when causal structure alone cannot do the job. Theories of causation based on human thought and projection, such as causal idealism, naturally incorporate intuitions. And some topics in causation that involve usage and convention centrally involve intuitions in debates over truth and assertability.

### C. Obstacles to Utilizing Experimental Data in Metaphysical Theories

Even if intuitions and empirical data about such intuitions are relevant to some theories of causation, there are methodological obstacles to the systematic use of them in metaphysical theories. Here I discuss several obstacles and objections to applying empirical data to metaphysical theories of causation. Some obstacles are methodological in nature, owing to limitations in how relevant intuitions are to certain debates. Other obstacles are due to shortcomings in the concepts or natures of intuitions themselves.

First, intuitions likely have little role to play in adjudicating strictly ontological debates about the nature of causation--debates over what causation really is in which participants already assume the mind-independence of the causal relation. For example, metaphysicians whose dispute is over the putative fundamentality of the causal relation

have no use for folk intuitions about this matter. Similarly for theorists who argue about whether grounding is like causation. There the argument is whether the posit satisfying one technical term of art (grounding) bears formal similarity to another technical posit (causation). Folk intuitions likewise have no bearing on the relationship between causation and natural laws, or whether causation is just energy transfer. One might argue that intuitions have bearing insofar as theories should account for intuitions. But as I suggested above, such a matter is a second-order matter of theory of choice rather than an objection to theories themselves. If intuitiveness is taken to be a virtue of theories more generally, then whether or not a theory is intuitive will be evidence for or against that theory. But if intuitiveness is not a dimension of theory choice, then intuitions have no bearing on many first-order ontological debates, including debates exclusively over the nature of mind-independent causation.

Conceptual analysis clearly has use for intuitions, and related empirical testing about their nature and structure. But sometimes conceptual analysis falls short due to limitations of the concepts themselves. Consider the example of whether the folk judge cases of “double prevention” to always count as causation. In cases of double prevention, one event prevents another event from preventing the occurrence of an outcome, e.g.:

(Fighter pilots): Suzy and Billy are fighter pilots. Suzy’s mission is to bomb a village. An enemy plane approaches. Billy shoots it down, preventing the enemy plane from preventing Suzy from bombing the village. If Billy hadn’t shot it down, Suzy would not have been able to bomb the village.

Intuitions differ about whether Billy is a cause of the village’s bombing, given that he isn’t connected to the bombing of the village via physical process. Lombrozo (2010) shows that folk intuitions differ between different cases of double prevention. In “obvious” cases of double prevention such as the firing of a gun (which involves a double preventive structure) causing a bullet wound, the folk take these sorts of cases to be clear cases of causation. But the folk are less likely to judge Fighter Pilots as straightforwardly causal. One conclusion we might draw from Lombrozo’s work, as Woodward (2014) points out, is that the folk concept of causation is more complex and subtle than previously imagined. But I see this rather as a case where the causal concept is simply not

complex enough to weigh in on whether double prevention counts as causation. As Lewis says of similarly confounding cases of overdetermination, such cases are “spoils to the victor” given the deficiency of the causal concept. (Shortly, I address the question of whether such cases are problematic because there are multiple causal concepts in play.) In cases when the details of the worldly phenomenon outstrip the concept, conceptual analysis can fall short in delivering data suited for reinforcing metaphysical views. The lesson is generalizable for many complex theories of causation: the folk concept might not be suited to weigh in on, for example, differing diagnoses of causal cases characterized by 20+ “neurons” in a complex neuron diagram, complex multi-variable causal models, or causal theories with numerous and complex theoretical posits.

Another limitation concerns the topic of experimental studies. Empirical studies of causal concepts are arguably relevant to selection problems, problems about which among many candidates are causes, or whether or not causation occurs in a given scenario. But these topics are arguably distinct from the nature of the causal relation itself. Studies often test whether or not the folk think that  $c$  is a cause of  $e$  in specific instances, i.e., the extension of the folk causal concept. But empirical studies don’t test what the causal concept *is*; that is, they don’t test the intrinsic nature of the causal relation apart from its real-world instances. This tendency manifests in several ways.

First, it manifests in an ambiguity with respect to which causal concept is in use. In the causation literature, much has been made of Ned Hall’s so called “two concepts” of causation. According to Hall (2004), there is a *productive* concept of causation encompassing transfer of energy from cause to effect, generally through spatiotemporally local causal chains. And there is a distinct *dependent* concept of causation encompassing counterfactual dependence of one effect on another. The dependent account covers, among other things, causation by omission and cases of causation without energy transfer.

Empirical studies of causation generally do not test for which causal concept is in operation.<sup>2</sup> If there are different concepts in play, there will be a question about which

---

<sup>2</sup> Some empirical studies do test this issue independently. Lombrozo (2010), for example, establishes that folk causal ascriptions manifest both concepts of causation depending on the mode of presentation. But studies that test causal intuitions more generally often do not test for specific causal concepts in play.

concept or concepts experimental results about causation reflect in the many experimental studies on the topic. An experiment meant to limn the relationship between causal and normative intuitions likely elicits both dependent and counterfactual concepts; a study on causation by omission might be confounded because only the dependent causal concept is in play.

As an example of the ambiguity in which causal concept is deployed, consider the following pair of cases:

(Victim) Two independent evil scientists plot to kill Victim. At time  $t$ , each scientist delivers an electric shock to Victim. Either shock would have been sufficient to kill Victim.

(Resistant Victim) Two independent evil scientists plot to kill Victim. At time  $t$ , each scientist delivers an electric shock to Victim. Unbeknownst to each evil scientist, Victim is particularly resistant to electric shocks, and it takes both shocks to cause Victim's death.

Victim is a case of *causal overdetermination*, in which multiple causes are individually sufficient to bring about an outcome, whereas Resistant Victim is a case of *joint causation*, in which multiple causes are necessary to bring about an outcome. The underlying question is whether the individual assassins in (Victim) are *more* or *less* causally responsible than those in (Resistant Victim) for Victim's death.<sup>3</sup> According to an energy transfer concept of causation, overdetermination involves "more" causation on the part of each cause, given that there are multiple energy transfers from each cause to the outcome. Thus each scientist in Victim is more responsible for the death than in Resistant Victim. But according to a dependent concept of causation, each joint cause is more causally responsible for Victim's death than each overdetermining cause, since had one joint cause not occurred, the outcome would not have occurred. The correct diagnosis of the contrast cases depends on which causal concept one employs. That different descriptions evoke different causal concepts in such a simple set of cases suggests that it

---

<sup>3</sup> I discuss this pair of cases and this question at length in my "Causal Proportions and Moral Responsibility" (forthcoming) and "Causal and Moral Indeterminacy" (forthcoming).



is ambiguous which causal concept is being employed in intuitive attributions of causation and moral evaluation, especially in cases of collective action. In studies of folk judgments of causation more generally, it is often unclear which causal concept is being elicited.

A second, related limitation concerns the clarity of the folk causal concept apart from norms of human action. It is widely established that the folk causal concept is often inextricably bound up with normative considerations. This point is demonstrated by the now-famous Knobe Effect, which shows that the folk judge damaging action to be intentional, and beneficial action to be less so. And the empirical work of Henne et. al (among many others) shows that the folk endorse a dependence of causation on normativity. For example, the reason that my omission rather than the Queen of England caused the plant's death is that I was the one who promised to water it. The problem is not necessarily the link between the causal concept and the moral one. The worry here is that the folk semantics of omissive claims such as "I failed to water the plant" largely tracks folk concepts of norms rather than a folk concept of causation. The risk is that *causation*, as a sort of metaphysical structure distinct from norms, falls out of the picture entirely.

To see the problem, consider a modified example of the problem of profligate omissions that includes no promises or related norms of human action:

(Rain) The failure of rain to fall caused the isolated patch of wild grass in southern France to die. But if anyone had travelled to the remote patch of grass to water it, the patch would not have died.

Here, there are countless humans that could have prevented the death of the patch of grass: had any of them not failed to water it, it would not have died. But there is no norm to guide selection of a particular cause of the grass' death: no one promised to or was expected to water it. Yet there are countless omissions such that had they not occurred, the grass would have lived. The only structural difference between this case and the classic plant-watering case is that there is a norm in play in the plant case (namely, someone promised to water the plant) whereas there is no norm in play in the rain case.

The folk judge the promiser to be the cause of the plant's death in the classic case. No studies that I know of have tracked folk intuitions about the non-normative case, but it seems clear that the folk concept will not line up with the truth of the numerous counterfactuals that create the problem of profligate omissions. For example, the folk are unlikely to accept that the Queen of England is a cause of death of the isolated patch of grass. Such are the limitations of a normative concept of causation: it falls short when these norms are not in operation. Studies establishing the relationship between causation and normative concerns may be relevant to the metaphysics of causation, but they cannot limn the mind-independent causal structure of the world minus norms. And it is worth noting that a large portion of the metaphysics of causation, across many different projects, does not concern norms at all. What metaphysicians are looking for is, e.g., an answer to the question: "In virtue of what at the bottom level of reality is the rain a cause of the plant's death while the Queen of England is not?" Reinforcing an intuitive link between causation and moral responsibility does not make progress on this strictly metaphysical question.

Finally, another obstacle to applying intuitions to a theory of causation involves cultural variation of the causal concept. Cultural variation in intuitions about knowledge, reference, and other central philosophical posits has been a linchpin of experimental philosophy since Weinberg, Nichols, and Stich (2001), but cultural variation in causal intuitions has been comparatively ignored by the experimental philosophy community. There are some signs of the topic in empirical psychology. A study from Choi et. al (1999) shows cultural variation in attributions of causation, specifically with respect actions' causal roots in dispositions and personality traits. Bender and Beller (2011) also suggest that causal attributions are subject to cultural influences. If the causal concept is subject to cultural variation, then what is being tested is often a culturally specific causal concept, rather than "the" causal concept. For example, if a study tests whether the folk take  $c$  to be a cause of  $e$  in a given scenario and the answers fall on either side of a cultural divide, then the causal concept will have different extensions in different cultures.<sup>5</sup>

#### D. Conclusion

The role of intuitions in metaphysical theories of causation has often been unclear: theories which purport to ignore intuitions often take them into account, and theories which aim to account for intuitions often seek to be about the world rather than our concepts exclusively. In this discussion, I have aimed to clear the ground about this complex and sometimes contentious relationship. Hybrid theoretical approaches (such as the famous Canberra Plan) aim to incorporate intuitions while still remaining, in some sense, “about the world”. Methodological space is bookended by different approaches and goals. At one end of the space is conceptual analysis, which aims to exclusively analyze the operation and structure of the human causal concept. Intuitions are clearly relevant to such a project. At the other end is fundamental metaphysics, which aims to limn the structure of mind-independent reality. No amount of empirical data-gathering about intuitions on such matters will answer these questions (even if the folk intuitions on such data is of intrinsic interest.) Finally, it is helpful to have a sense of the limitations of experimental philosophy in aiding metaphysics, even those projects friendly to intuitions. Even if experimental philosophy aims to reinforce conceptual analysis by gathering data on intuitions, the method can fall short in various ways. Knowing this does not lessen the impact of empirical studies, but does deliver a clearer picture of the delicate union of metaphysics and intuition.

Where do we go from here? Having a clearer picture of the relationship between various metaphysical projects and intuitions shifts the debate into new terrain. Instead of asking: what is the role for intuitions in metaphysics?, the question becomes: how, exactly, should each causal project be classified? Even for theories that explicitly purport to be doing one sort of project over another, there is room for debate about how unadulterated by intuition these projects in fact are. As we have seen, projects that aim to be strictly ontological are often hybrid, and projects which aim to model mind-independent reality nonetheless lean on intuitions for certain problems of demarcation. Causation theorists can progress by clarifying their own methodologies within the preceding classificatory schema, and experimental philosophers can progress by taking into account these differences. <sup>6</sup>



## References

- Alicke, M., Rose, D., & Bloom, D. (2014). Causation, Norm Violation, and Culpable Control\*. Volume 2 Experimental Philosophy, 229-252.
- Bender, A., & Beller, S. (2011). Causal Asymmetry Across Cultures: Assigning Causal Roles in Symmetric Physical Settings. Frontiers in Psychology Front. Psychology, 2.
- Bernstein, S. (forthcoming a) Causal Proportions and Moral Responsibility. Oxford Studies in Agency and Responsibility, ed. David Shoemaker.
- Bernstein, S. (forthcoming b) Causal and Moral Indeterminacy. Ratio.
- Bernstein, S. (forthcoming c) Causal Idealism. In Idealism: New Essays in Metaphysics, eds. Tyron Goldschmidt and Kenny Pearce, Oxford University Press.
- Bernstein, S. (forthcoming d) Omission Impossible. Philosophical Studies.
- Choi, I., Nisbett, R., and Norenzayan, A. (1999) Causal Attributions Across Cultures: Variation and Universality. Psychological Bulletin Vol. 125, No. 1., pp. 47-63.
- Clarke, R. et al. (2015). Causation, Norms, and Omissions: A Study of Causal Judgments. Philosophical Psychology 28 (2):279-293.
- Dorr, C. and G. Rosen (2002) Composition as a Fiction, in R. Gale (ed.), The Blackwell Guide to Metaphysics, Oxford: Blackwell.
- Hall, N. (2004) Two Concepts of Causation in Counterfactuals and Causation, edited by John Collins, Ned Hall, and L. A. Paul, MIT Press, Cambridge MA, pp. 139-180.
- Joseph Y. Halpern & Christopher Hitchcock (2015). Graded Causation and Defaults. British Journal for the Philosophy of Science 66 (2):413-457.
- Henne, P. et. al. (forthcoming) Cause by Omission and Norms. Australian Journal of Philosophy.
- Hitchcock, C. (2007) "Prevention, Preemption, and the Principle of Sufficient Reason," Philosophical Review, 116: 495–532.
- Hitchcock, C & Knobe, J. (2009). Cause and Norm. Journal of Philosophy 106 (11):587-612.
- Ladyman, J. and Ross, D. (2007) Every Thing Must Go. Oxford: Oxford University Press.
- Liebman, D. (2011) Causation and the Canberra Plan. Pacific Philosophical Quarterly (92) 232-242.
- Lewis, D. K. (1973). Causation. Journal of Philosophy, 70, 556-567.
- Lombrozo, T. (2010) Causal-Explanatory Pluralism: How Intentions, Functions, and Mechanisms Influence Causal Ascriptions. Cognitive Psychology 61: 303-32.

Menzies, P. (2004) Difference-making in Context, in Counterfactuals and Causation, edited by John Collins, Ned Hall, and L. A. Paul, MIT Press, Cambridge MA, pp. 139-180.

Nolan, D. (2009) Platitudes and Metaphysics. In David Braddon-Mitchell & Robert Nola (eds.), Conceptual Analysis and Philosophical Naturalism. Cambridge: MIT Press.

Paul, L. (2010) A New Role for Experimental Work in Metaphysics. Review of Philosophy and Psychology 1:461-4761

Ripley, D. (ms) Exploring Counterpossibles Experimentally.

Schaffer, J. (forthcoming) Cognitive Science and Metaphysics: Partners in Debunking Goldman and his Critics, eds. Kornblith and McLaughlin: Wiley-Blackwell.

Weinberg, J., S. Nichols, and S. Stich. (2001) “Normativity and Epistemic Intuitions.” Philosophical Topics 29 (429–60).

Woodward, J. (2014) Causal Reasoning: Philosophy and Experiment in Oxford Studies in Experimental Philosophy, eds. Tania Lombrozo, Joshua Knobe, and Shaun Nichols.

---

<sup>1</sup> Note that this definition is compatible with a theory contradicting some intuitions. Suppose that a sum of intuitive judgments are inconsistent. A theory must contradict some of them. But the theory may still ‘answer to’ or ‘accord with’ the intuitions by being compatible with the best mix of intuitions. Thanks to Jason Turner for this point.

<sup>2</sup> David Rose points out that one way in which such objections could have a grip is if intuitions play a role in strictly ontological projects, including some strains of conventionalism and social constructivism. Here I set aside such projects, as I view them as other than something strictly ontological.

<sup>3</sup> Thanks to Jason Turner for this example, and for pressing me on this issue.

<sup>4</sup> Bernstein (forthcoming d) discusses nonvacuity of causal counterpossibles in detail.

<sup>5</sup> Here I set aside the even stronger conclusion that there is *no* project of conceptual analysis if there is cultural variation in the causal concept.

<sup>6</sup> Thanks to Paul Henne, David Rose, and Jason Turner for extensive feedback on this paper.