# Actual Causation and Compositionality

Jonathan Livengood and Justin Sytsma

**Abstract:** Many theories of actual causation implicitly endorse the claim that if *c* is an actual cause of *e*, then either *c* causes *e* directly or every intermediary by which *c* indirectly causes *e* is itself both an actual cause of *e* and also an actual effect of *c*. We think this compositionality constraint is plausible. However, as we show, it is not always satisfied by the causal attributions ordinary people make. After showing that the compositionality constraint is not always satisfied by the causal attributions ordinary people make, we step back to consider what philosophers working on causation should do when the deliverances of their theories diverge from what ordinary people say.

In this paper, we identify a structural constraint—the *compositionality constraint*—that is implicitly endorsed by many accounts of actual causation in the philosophical literature, and we present evidence suggesting that the causal attributions ordinary people make sometimes violate the compositionality constraint. In Section 1, we articulate the constraint and argue that many accounts of causation in the literature satisfy it. In Sections 2 and 3, we argue that there is reason to predict that ordinary causal attributions do not tend to respect the compositionality constraint in all cases, and we put our prediction to the test. Finally, in Section 4, we step back to reflect on the compositionality constraint, the goals of philosophical work on actual causation, and the implications of our results.

## 1. Articulating the Compositionality Constraint

Causation comes in at least two varieties—structural causation and actual causation. Structural causal relations are something like causal laws. They generate patterns of statistical association and underwrite the efficacy of interventions. By contrast, actual causation is about accounting for

things that have already happened: for attributing moral and legal responsibility, for explaining historical events, and for diagnosing complex systems.

Suppose that *c*, *d*, and *e* are events that actually occur. An initially plausible requirement on theories of actual causation is that they satisfy the following *compositionality constraint*: if *c* is an actual cause of *e*, then either *c* causes *e* directly or every intermediary *d* by which *c* indirectly causes *e* is itself an actual effect of *c* and an actual cause of *e*. To illustrate, suppose Hazel sets up a long chain of dominoes and then knocks them over by giving the first domino a flick. Hazel's flick caused the first domino in the chain to fall *directly*, not by way of any intermediary. The first domino falling caused the second in the chain to fall and so on to the end of the chain. Actual causation may not be transitive, but in this case it seems that Hazel's flick not only caused the *first* domino to fall, it also caused the *last* domino to fall. But Hazel's flick did not *directly* cause the last domino to fall; rather, Hazel's flick *indirectly* caused the last domino to fall by way of the other dominoes in the chain. The compositionality constraint says both that if Hazel's flick actually caused the last domino to fall, then either the flick directly caused the last domino to fall or the flick indirectly caused the last domino to fall and also that if the flick indirectly caused the last domino to fall, then it actually caused (either directly or indirectly) some intermediary domino to fall and the intermediary domino falling actually caused (either directly or indirectly) the last domino to fall.

Many theories of actual causation entail the compositionality constraint. For example, consider a simple mark-transmission account in the spirit of Reichenbach (1956), where some event *c* is a cause of another event *e* iff *c* transmits a mark to *e*. Naively, an event *c* transmits a mark to another event *e* only if *c* exhibits a sign that is acquired by *e* and there is a (possibly

empty) sequence of events—all of which exhibit the sign—connecting *c* and *e*.[1] When the sequence of events is empty, the transmission is direct in the sense of the compositionality constraint, and when the sequence is non-empty, the transmission is indirect. Now, if *c* causes *e* by way of a sequence of events, then *c* must transmit the mark to each of the events in the sequence in order to successfully transmit it to *e*. As such, the simple mark-transmission account entails the compositionality constraint. The conserved-quantity accounts of Salmon (1994, 1997) and Dowe (1995, 2000) make the notion of connection considerably clearer, but they retain the core idea here—and with it the compositionality constraint.

Ehring's (1997) trope persistence theory takes a precise version of the compositionality constraint as a *definition* of causation. Ehring first defines a symmetric relation of *strong causal connection* in terms of the relations of identity, parthood, and lawful connection for tropes. In order to secure the asymmetry of causation, he adds a notion of causal priority, and then he defines causation as follows (2004, 73-74):

> Trope *P* at *t* causes trope *Q* at *t´* if (A) *P* at *t* is strongly causally connected to *Q* at *t´* and *P* at *t* is causally prior to *Q* at *t´* [or] … (B) there is a set of properties ($R_1$, …, $R_n$) such that *P* is a cause of $R_1$ under clause (A), …, and $R_n$ is a cause of *Q* under clause (A).

[1] We do not have a rigorous account of "connection" to offer. A fully-articulated mark-transmission account needs to make "connection" precise—as Salmon (1994, 1997) and Dowe (1995, 2000) attempt to do—without appealing to causation. In an earlier draft, we said that for a sign exhibited by *c* to be acquired by *e*, the events must either have no gap between them or be connected by a sequence of events that all exhibit the sign. An anonymous referee expressed puzzlement, remarking: "Presumably one could in principle have mark transmission across a spatio-temporal gap, so that should not be what is meant by 'gap'." We agree and have dropped the term. However, we think that *in the context of causal process accounts of causation*, the idea that there should be no spatio-temporal gaps is very inviting. As Armstrong (2004, 447) puts it: "There is a deep predisposition to think that where cause and effect are not contiguous in both space and time we can always trace an intermediate chain of causes that links the cause to its distant effect." Armstrong defends the predisposition against the objection that causation by omission violates it. Unremarked by Armstrong is that one may satisfy the *spirit* of the predisposition without requiring spatio-temporal contiguity for causes and effects. One may cut away the concern with spatio-temporal contiguity from Armstrong's predisposition and obtain our compositionality constraint, which is satisfied by accounts of causation that have no special problem with causation by omission.

Ehring's (A) and (B) clauses correspond, respectively, to direct and indirect causation in the compositionality constraint. Whenever *P* causes *Q*, it either does so directly or it does so indirectly by way some intermediaries $R_i$, each of which is an effect of *P* and a cause of *Q*.

If, as Hall (2004) maintains, there are two fundamentally different varieties of causation—a relation of causal production and a relation of causal dependence—then one might worry that the compositionality constraint is an artifact of production accounts. But standard accounts of causal dependence are also committed to the compositionality constraint. To illustrate, consider Lewis's (1973) counterfactual theory.

According to Lewis, an event *e* depends causally on an event *c* if and only if the following two counterfactuals are true: (1) if *c* were to occur, then *e* would occur; and, (2) if *c* were not to occur, then *e* would not occur. Lewis maintains that if *e* causally depends on *c*, then *c* is a cause of *e*. But he rejects the converse, writing (563), "Causation must always be transitive; causal dependence may not be; so there can be causation without causal dependence." In order to secure transitivity, Lewis does not immediately analyze causation in terms of causal dependence. Instead, he tells us that a causal chain is a finite sequence $< c, d, e, ... >$ of actual particular events such that *d* causally depends on *c*, *e* causally depends on *d*, and so on. And then he says that an event *c* is a cause of an event *e* if and only if there is a causal chain beginning with *c* and ending with *e*. Now, any initial segment of a causal chain is itself a causal chain, and any tail segment of a causal chain is *also* a causal chain. Therefore, we have the following version of the compositionality constraint on Lewis's account: if *c* is a cause of *e*, then either there is no intermediary between *c* and *e*, or there are some intermediaries $d_i$ on a causal chain $< c, d_1, d_2, ..., d_2, e >$, where for each $d_i$, *c* is a cause of $d_i$, and $d_i$ is a cause of *e*.[2]

[2] The account in Lewis (2000) also satisfies the compositionality constraint because (as before) Lewis takes causation to be the ancestral of causal dependence, now understood as *influence* (191).

The compositionality constraint is also entailed by some probabilistic accounts of causation. For brevity, we'll focus on an obvious example—Lewis's (1986) analysis of chancy causation.[3] As with his counterfactual theory, Lewis analyzes chancy causation in two stages. He first cashes out chancy causal dependence in terms of counterfactual conditionals about single-case probabilities. According to Lewis, *e* chancily causally depends on *c* iff the probability that *e* occurs would be greater if *c* were to occur than if *c* were not to occur. Then, he takes chancy causation to be the ancestral (the transitive closure) of chancy causal dependence. Since *c* chancily causes *e* iff there is a chain of chancy causal dependences running from *c* to *e*, it follows that either *e* chancily depends on *c* directly or for any event *d* on the chain of dependences running from *c* to *e*, *c* is a chancy cause of *d* and *d* is a chancy cause of *e*.

Finally, all of the purely structural theories of actual causation that have been advanced in the graphical causal modeling tradition (so far as we know) are committed to the compositionality constraint. In the graphical causal modeling tradition, structural causal relations hold between variables in a causal model, which is often (partially) represented pictorially by a directed graph. A purely structural theory of actual causation provides conditions for counting an evaluated variable as an actual cause of some other evaluated variable, where those conditions are given in terms of the structural causal relations and the results of possible manipulations of the variables in the model.[4] For example, Woodward's (2003) account says that $C = c$ is an actual cause of $E = e$ iff the following two conditions both hold:

[3] The probabilistic account in Menzies (1989) also entails the compositionality constraint. The account in Glynn (2011) does not entail the constraint, but he tells us (personal communication) that he hopes his account is consistent with the constraint.

[4] The details of the relationship between actual causation and structural causation are a matter of ongoing debate (see Hitchcock 2001, Woodward 2003, Chockler and Halpern 2004, Halpern and Pearl 2005, Hitchcock 2007a and 2007b, Glymour et al. 2010, Livengood 2013, Halpern and Hitchcock 2015, Halpern 2016, and Blanchard and Schaffer 2017). Purely structural accounts, such as the three definitions of actual causation given in Chapter 2 of Halpern (2016), are to be contrasted with accounts that supplement the structural machinery with considerations of defaults, typicality, or normality. See Halpern and Hitchcock (2015) for a recent account that supplements structural

(W1) The actual value of $C$ is $c$, and the actual value of $E$ is $e$.
(W2) There exists a path P from $C$ to $E$, and there exist manipulations $do(C = c^*)$ for $c^* \neq c$ and $do(\boldsymbol{W}=\boldsymbol{w})$ for $\boldsymbol{w}$ in the redundancy range of P such that $E_{\boldsymbol{W}=\boldsymbol{w},\, C=c^*} \neq e$.

Where $\boldsymbol{w}$ is in the redundancy range of a path P iff carrying out the manipulations denoted by $do(\boldsymbol{W} = \boldsymbol{w})$ leaves all of the variables on P at their actual values.

In the language of graphical causal modeling, the compositionality constraint may be stated precisely as follows:

> If $C = c$ is an actual cause of $E = e$, then **either** $C$ is a direct structural cause of $E$ **or** there is at least one path P from $C$ to $E$ such that for every variable $D$, if $D$ is on P and $D = d$, then $C = c$ is an actual cause of $D = d$ and $D = d$ is an actual cause of $E = e$.

If $C = c$ is an actual cause of $E = e$ according to Woodward's account, then there is a path from $C$ to $E$ in the causal graph. Moreover, there are manipulations of variables other than $C$ such that $E$ retains its initial value but would have a different value were $C$ to be manipulated appropriately in that context. Pick a variable on that path. Call it $D$. Since the manipulations of variables other than $C$ were in the redundancy range for the path P, the value of $D$ remains at its actual value $d$ after the imagined manipulation. But the value of $D$ would have to change if $C$ were manipulated appropriately. Otherwise, the value of $E$ would not change following a manipulation of $C$ as it should. Hence, $C = c$ is an actual cause of $D = d$. Similarly, if we carry out a manipulation of variables that satisfies condition (W2), then we bring about a context in which there is a way of manipulating $D$ such that $E$ takes on a new value. Therefore, Woodward's account entails the compositionality constraint.

machinery with considerations of defaults, typicality, and normality. See Livengood, Sytsma, and Rose (2017) for further experimental work on such "DTN accounts." Halpern also discusses defaults, typicality, and normality in Chapter 3 of his book.

## 2. Ordinary Causal Attributions and the Compositionality Constraint

The compositionality constraint strikes us as a plausible requirement on any adequate theory of actual causation. But how might we test to see whether what strikes us as plausible is correct? One very tempting approach is to adopt what Livengood, Sytsma, and Rose (2017) call the *folk attribution desideratum* (FAD) as a condition of adequacy for theorizing about actual causation and require that what a theory of actual causation says about concrete, everyday cases accords with ordinary causal attributions. As noted in that article, "ordinary causal attribution" refers to the use of language like "X caused Y" (see Sytsma et al. 2019 for a further discussion of causal language more broadly construed). While our primary focus in this paper is on ordinary causal attributions, we discuss the broader question of causal cognition in Section 4.

Ordinary causal attributions are central to recent empirical work on actual causation. And as detailed in Livengood, Sytsma, and Rose (2017), many philosophers and other researchers developing theories of actual causation have expressed some commitment to the FAD.[5] So, do ordinary causal attributions satisfy the compositionality constraint? On the basis of existing empirical research, we conjecture that they do not always abide by this constraint, such that it is not universally treated as a constraint on ordinary causal attributions. One robust finding in recent empirical work on ordinary causal attribution is that injunctive norms play a substantial role in people's judgments about concrete cases.[6] People's judgments about what someone or something *ought* to do in a given situation have a significant impact on ordinary causal attributions. We have argued that these findings are best explained by what we call *the responsibility view*, which maintains that the default concept of causation at play in ordinary

[5] See Bernstein (2017), however, for a dissenting view.
[6] See Hilton and Slugoski (1986); Alicke (1992); Knobe (2006); Hitchcock and Knobe (2009); Sytsma, Livengood, and Rose (2012); Kominsky et al. (2015); and, Livengood, Sytsma, and Rose (2017), among others.

causal attributions has some indispensable evaluative content alongside whatever non-evaluative content it has (Sytsma, Livengood and Rose 2012; Livengood, Sytsma, and Rose 2017; Sytsma and Livengood MS; Sytsma et al. 2019; Sytsma MS).[7] Hence, on our view, this concept is a kind of thick ethical concept akin to our concept of responsibility. If so, then ordinary causal attributions will tend to violate the compositionality constraint for cases where someone or something is responsible for an effect by way of an intermediary that does not share in the responsibility.[8] We tested several cases of this type and report our results in this and the following section. As predicted, we found that ordinary causal attributions by and large do not satisfy the compositionality constraint in these cases.

### *2.1 Poisoned Cup*

In our first study, we gave participants the following vignette:

> Amy wants to kill her daughter, Jessica, but she doesn't want to go to prison for murder. As such, Amy hatches a plan. She arranges for a baby sitter, Courtney, to take care of Jessica while she is out of town on business. Before leaving, Amy laces one of Jessica's sippy cups with a deadly poison that is very difficult to detect. That evening, Courtney gives Jessica juice in the poisoned sippy cup. Jessica drinks the juice and dies two hours later.

We then asked participants to rate their level of agreement or disagreement with the following two causal attributions on a seven-point scale anchored at 1 with "strongly disagree," at 4 with "neutral," and at 7 with "strongly agree": (1) Amy caused Jessica's death; (2) Courtney caused Jessica's death.[9]

---

[7] McGrath (2005) defends a similar idea, which she calls the NORMAL PROPOSAL, for causation by omission.
[8] One might raise other worries about whether ordinary causal attributions satisfy the compositionality constraint. See Livengood (2013) for one direction this might go.
[9] An anonymous referee observed that while we characterize compositionality in binary terms—with every event either being an actual cause or failing to be an actual cause—our measure is a seven-point scale. The referee then wondered whether violations of compositionality might come in degrees: for example, where some event *c* is thought to be a very strong cause of some event *e* but a significantly weaker cause of a required intermediary event *d*. This might arise despite the average ratings being greater than the mid-point. We think this is an interesting

The Poisoned Cup case seems like a straightforward example of a simple sequence connecting a cause to an effect through a salient intermediary: Amy brings about Jessica's death via Courtney. Given this simple model, according to the compositionality constraint, if Amy is an actual cause of Jessica's death, then Courtney is an actual cause of Jessica's death as well. Assuming ordinary causal attributions uphold the compositionality constraint, participants should tend to affirm the second statement if they affirm the first. However, in the Poisoned Cup case, the responsible party brings about an effect through an intermediary who does not share in the responsibility. Amy *dupes* Courtney into delivering the poison to Jessica. Hence, the responsibility view predicts a failure of compositionality. If ordinary causal attributions align with responsibility judgments, then participants should affirm that Amy caused Jessica's death but deny that Courtney caused Jessica's death.

We collected responses for the Poisoned Cup case online from 34 native English-speakers, 18 years of age or older, who completed the survey, had not taken a survey through the website previously, and had at most minimal training in philosophy.[10] The same restrictions were used for each study we report. The results are striking and strongly align with our prediction (Figure 1): each participant completely agreed with the claim that Amy caused Jessica's death (M=7), and two-thirds of participants (22/34) completely disagreed with the claim that Courtney

possibility, but we do not have space to explore it experimentally. One might wonder whether (and if so, under what conditions) participants conflate an attribution that [X is a cause of Y to degree D] with an attribution to degree D that [X is a cause of Y]. While we think it is *possible* that participants would confuse the two in some settings, we doubt that they typically do so. But even if participants frequently conflate them, we do not think it threatens the points we want to make in this paper. We do wonder, however, whether there is any interesting relationship between attributions of what we might call *full causation* and attributions of what we might call *partial or graded causation*. Perhaps the two are related in the same way as full belief and partial belief. To our knowledge, no one has investigated the relationship here, though there is some excellent work on partial causation in the psychology literature, e.g. Lagnado, D., Gerstenberg, T., and Zultan, R. (2013).

[10] Responses were collected through the Philosophical Personality website (philosophicalpersonality.com). Participants were counted as having more than minimal training in philosophy if they were philosophy majors, had completed a degree with a major in philosophy, or had taken graduate-level courses in philosophy. Participants were 70.6% women, average age 35.6, ranging from 20 to 55

caused Jessica's death (M=2.06). The central tendency of the ratings for Amy was statistically greater than the neutral value of 4—indicating that participants agreed with the claim that Amy caused Jessica's death. Whereas, the central tendency of the ratings for Courtney was statistically smaller than the neutral value of 4—indicating that participants disagreed with the claim that Courtney caused Jessica's death.[11] The effect sizes were very large for Amy and large for Courtney.[12]

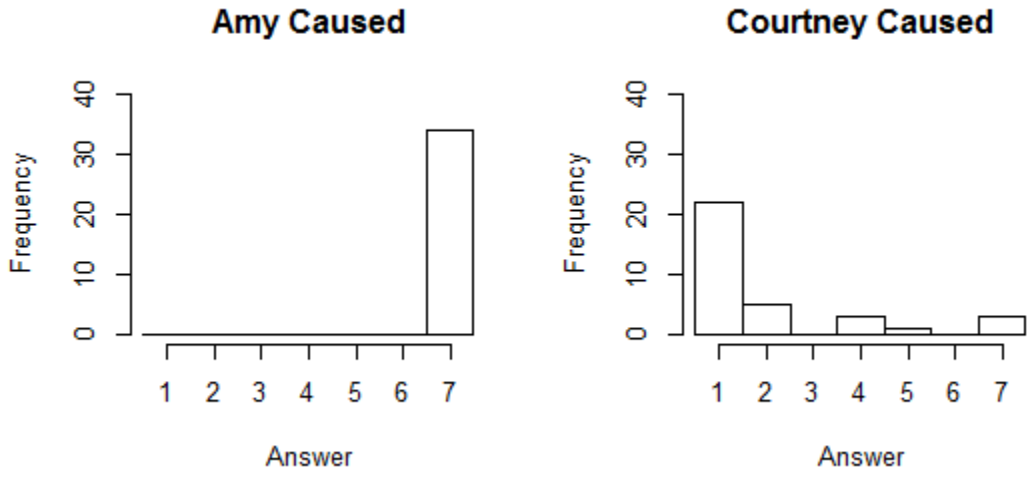


**Figure 1:** Histograms for Study 1

The causal attributions we observed in our first study by and large violate the compositionality constraint, since (1) Amy does not bring about Jessica's death directly but does

[11] Wilcoxon matched-pairs signed rank tests rejected the hypothesis that the central tendency for "Courtney" was greater than or equal to 4 ($V=58$, $p=3.568e^{-5}$) and the hypothesis that the central tendency for "Amy" was less than or equal to 4 ($V=595$, $p=2.922e^{-9}$).

[12] We used Rogmann's (2013) orddom package in R to calculate the "probability of superiority" and Cliff's delta for our comparisons. The within-subject probability of superiority of "Amy" against an equal-length vector of 4's was 1.00, meaning that every individual's evaluation of the claim that Amy caused the death was strictly greater than the neutral value of 4. Cliff's delta for Amy compared with the neutral vector was also 1. The within-subject probability of superiority of the neutral vector against "Courtney" was 0.794. Cliff's delta for the neutral vector compared with Courtney was 0.676 with a 95% confidence interval of [0.376, 0.848]. From now on, we will simply write 95% confidence intervals in square brackets after a point estimate. In this paper, we will use the following ordinary language descriptions of Cliff's delta results: 0-20% is practically no effect; 20-40% is a small effect; 40-60% is a moderate effect; 60-80% is a large effect; and greater than 80% is a very large effect. See McGraw and Wong (1992) for the theory.

so *by way of* Courtney, (2) participants all affirm that Amy caused Jessica's death, and (3) and the vast majority of participants deny that Courtney caused Jessica's death. We predicted these results and explain them by appeal to the responsibility view. As a check on our explanatory framework, we ran two further studies. According to the responsibility view, if we increase Courtney's responsibility for the outcome, her causal ratings should increase, and if we decrease Amy's responsibility, her causal ratings should decrease. We tested these predictions in our second and third studies, respectively.[13] Responses were collected from 59 participants.[14] The results agreed with the predictions (Figure 2). In Study 2, approximately 60 percent of participants completely agreed with the claim that Amy caused Jessica's death (17/29, M=5.97) and the claim that Courtney caused Jessica's death (18/29, M=6.03). In Study 2, the central tendencies of the ratings for Amy and for Courtney were statistically greater than 4.[15] By contrast, in Study 3, more than half of the participants completely disagreed with the claim that Amy caused Jessica's death (16/30, M=2.43) and the claim that Courtney caused Jessica's death (20/30, M=2.03). In Study 3, the central tendencies of the ratings for Amy and for Courtney were statistically less than 4.[16]

[13] In our second study we modified the vignette from our first study to read: "Amy wants to kill her daughter, Jessica. As such, Amy hatches a plan. She arranges for a contract killer, Courtney, to poison Jessica while she is out of town on business. After Amy leaves, Courtney laces one of Jessica's sippy cups with a deadly poison that is very difficult to detect. That evening, Courtney gives Jessica juice in the poisoned sippy cup. Jessica drinks the juice and dies two hours later." In our third study, the vignette read: "Amy has a daughter, Jessica. Amy is going out of town on business. She arranges for a baby sitter, Courtney, to take care of Jessica. Before leaving, Amy purchases a new sippy cup for Jessica. Unbeknownst to Amy, the sippy cup has been laced with a deadly poison that is very difficult to detect. That evening, Courtney gives Jessica juice in the poisoned sippy cup. Jessica drinks the juice and dies two hours later."

[14] 64.4% women, average age 38.4, ranging from 18 to 65

[15] AMY: V=362, $p=8.564e^{-5}$, $\Delta=0.759$ [0.422, 0.911]; COURTNEY: V=343, $p=5.206e^{-5}$, $\Delta=0.793$ [0.488, 0.925].

[16] AMY: V=84.5, p=0.001408, $\Delta=0.633$ [0.279, 0.836]; COURTNEY: V=21.5, $p=1.545e^{-5}$, $\Delta=0.667$ [0.357, 0.844].

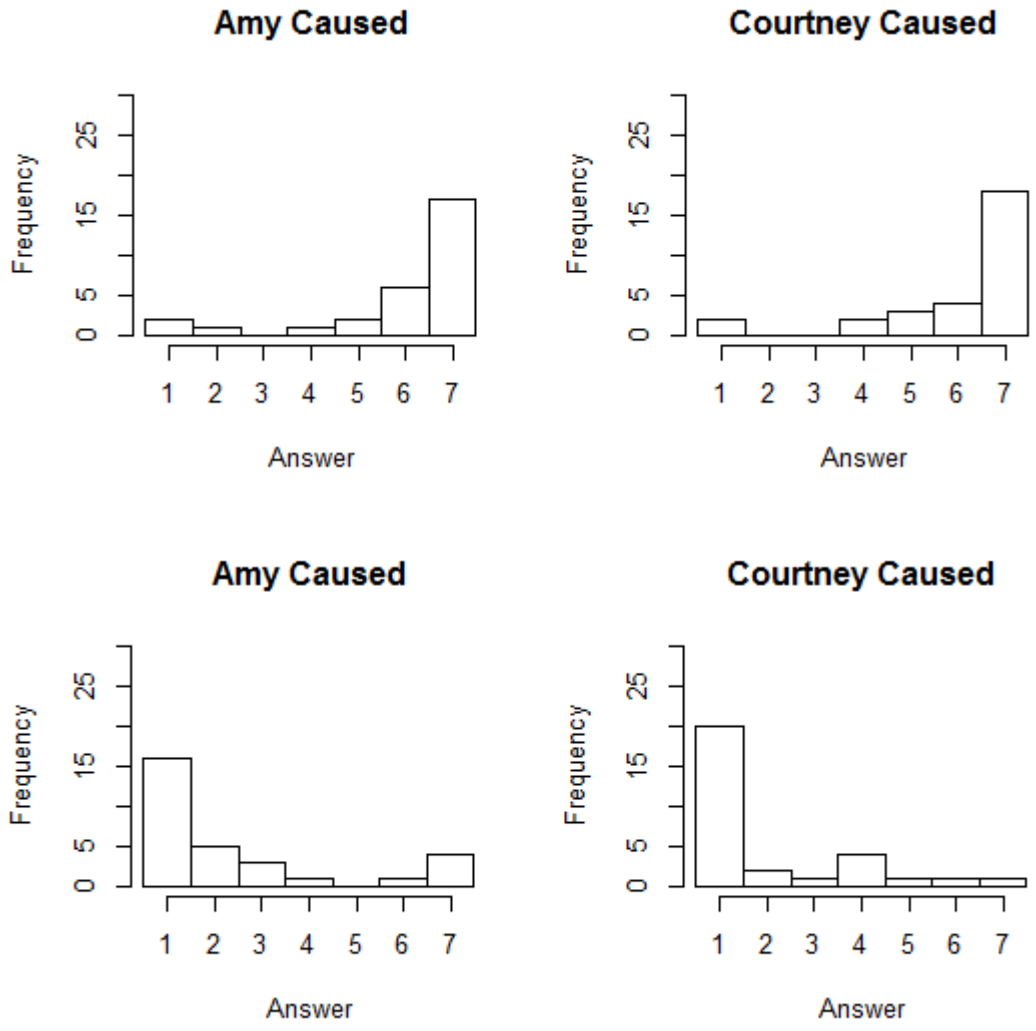


**Figure 2:** Histograms for Study 2 (top) and Study 3 (bottom)

*2.2 Objections and Replies*

Even if our preferred *explanation* of the results for the Poison Cup case is incorrect, we think our results provide a clear example of a failure of compositionality for ordinary causal attributions. If you disagree, we suspect it is for one of three reasons. First, you might think that our participants are confusing *causation* with *responsibility* in a way that makes their answers irrelevant to evaluating the compositionality constraint. Second, you might think that we should have asked

about *actions* or *events* or *facts*, rather than *agents*. Third, you might think that we have the wrong causal model for the case and that the underlying causal structure in the Poison Cup case should be represented by a *collider* rather than a *chain*.

### 2.2.1 Causation versus Responsibility

You might be worried that participants in our study were confusing causation with responsibility. On our view, ordinary people are not confused. Rather, the default concept of causation at play in ordinary causal attributions has both descriptive content and normative, evaluative content. We think that the responsibility view best accords with the available empirical data. However, one might defend what we call a *bias view* according to which the default concept does not have any evaluative content but ordinary causal attributions are strongly biased, e.g. by a desire to assign blame (Alicke 1992). Accepting a bias view, one might argue that the default concept is the same concept targeted by philosophical accounts.

We have three replies. First, explaining the results of the present study would require a self-underminingly large bias. If the compositionality constraint is a highly intuitive principle underlying the concept of causation that philosophers have traditionally targeted and if that concept is the default concept at play in ordinary causal attributions, then any bias strong enough to override the compositionality constraint would lead to incredibly large effects in other cases in the literature. The compositionality constraint does appear to be a highly intuitive principle underlying the concept of causation *that philosophers working have traditionally targeted*. But we do not see incredibly large effects in other cases in the literature. Hence, the concept of causation that philosophers have traditionally targeted is not the default concept at play in ordinary causal attributions. Second, the burden of proof is on proponents of the bias view.

Comparing the responsibility view and the bias view raises the difficult challenge of distinguishing between the systematic misapplication of a concept and the correct application of an alternative concept. Neither we nor proponents of the bias view have articulated any *general* way of drawing such a distinction. But in the absence of some general way of drawing the distinction, both charity and simplicity favor the responsibility view, since the responsibility view takes ordinary causal attributions to result from the correct application, while the bias view does not. Third, we have elsewhere conducted studies aimed at experimentally distinguishing the responsibility view and the bias view (for which, see Sytsma and Livengood MS), and those studies favor the responsibility view.

### 2.2.2 The wrong relata?

One might worry that while philosophical accounts of actual causation have typically focused on causation between events, we asked participants about whether an *agent* (Amy, Courtney) caused an event (Jessica's death). Perhaps if we asked about the agents' actions, then participants' judgments would shift into line with the compositionality constraint. As noted above, however, our focus in this paper is on accounts that follow the FAD (see Section 4 for further discussion). And although it has not been much remarked on, one clear finding in recent empirical work on causal attribution is that ordinary people often count agents as causes. Further, thought experiments in the causation literature often involve cases where an agent is closely tied to the action she performs (a hiker ducking, a kid throwing a rock, etc.), and sometimes, authors shift back and forth between saying that an action or event caused something and saying that a

relevant agent did.[17] As such, we believe that a failure of compositionality for attributions to agents would be problematic for many accounts of actual causation, even if there were not a corresponding failure for attributions to events. That said, we expect that our findings will generalize.

Previous research suggests that asking participants about an agent's action does not dramatically alter ordinary causal attributions relative to asking about the agent directly (Livengood, Sytsma, and Rose 2017). As such, we expect that the way we phrased our test statements is not driving the responses. Nonetheless, this is an empirical claim that is easily checked. In our fourth study we gave participants the same probe as in the first study, but we changed the two test attributions as follows: (1) Amy caused Jessica's death by poisoning the sippy cup; (2) Courtney caused Jessica's death by giving her juice in the sippy cup. Responses were collected from 137 participants.[18] The results were not significantly different from the results for the original study (Figure 3): almost every participant (129/137) completely agreed with the first claim (M=6.84), and roughly two-thirds (92/137) completely disagreed with the second claim (M=2.06). As in our first study, the ratings for Amy were statistically greater than 4, while the ratings for Courtney were statistically smaller than 4.[19]

[17] Hume famously does this. See also McDermott (1995, 527) and Halpern and Pearl (2005, 868 and 871). Paul and Hall (2013, 4) mention the possibility of confusion but reject agents as proper causes, writing, "While Suzy might cause a window to break, she does so only in virtue of the way she is involved in an event."
[18] 66.4% women, average age 33.7, ranging from 18 to 71
[19] AMY: V=595, $p=2.922e^{-9}$, $\Delta=0.941$ [0.862, 0.975]; COURTNEY: V=1246.5, $p=3.568e^{-5}$, $\Delta=0.667$ [0.539, 0.764].

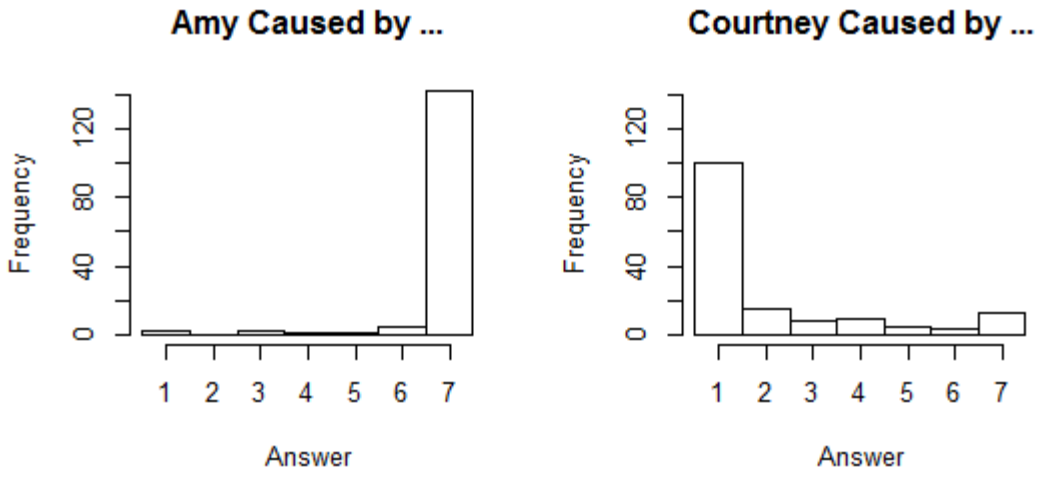


**Figure 3:** Histograms for Study 4

In our fifth study we went further, changing the two attributions to ask directly about the agent's actions: (1) Amy's action of poisoning the sippy cup caused Jessica's death; (2) Courtney's action of giving Jessica juice in the sippy cup caused Jessica's death. Responses were collected from 56 participants.[20] Almost every participant (49/56) completely agreed with the first claim (M=6.52), and almost half of the participants (26/56) completely disagreed with the second claim (M=3.36). The results again suggest that the causal attributions of most ordinary people do not satisfy the compositionality constraint for this case (Figure 4).[21] Comparing the ratings for Amy and Courtney directly, we found that they were statistically different and that the effect was moderate.[22] However, we would be remiss if we did not point out that in this study a notable minority of participants (16/56) completely *agreed* with the second claim, which is in

[20] 71.4% women, average age 35.2, ranging from 18 to 74

[21] AMY: V=1450.5, p=8.176e$^{-11}$, Δ=0.839 [0.637, 0.933]; COURTNEY: V=513.5, p=0.04302, Δ=0.214 [-0.045, 0.447].

[22] A Wilcoxon matched-pairs signed rank test rejected the hypothesis that the central tendency for "Amy" was equal to the central tendency for "Courtney": V=708, p=3.595e$^{-7}$, Δ=0.571 [0.392, 0.709].

line with the compositionality constraint. What individual differences might account for our results is an interesting question that deserves to be studied further, but since the clear majority opinion is inconsistent with the compositionality constraint, we will not pursue it further here.

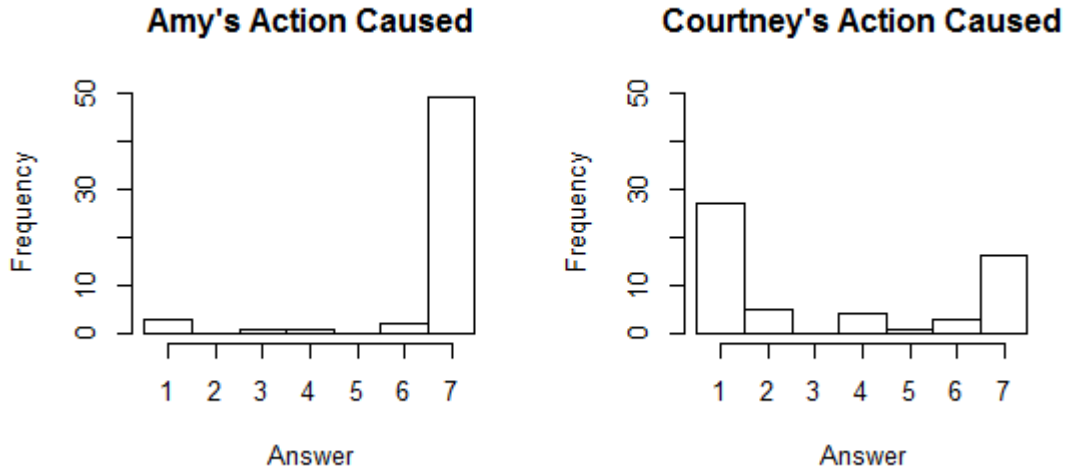


**Figure 4:** Histograms for Study 5

2.2.3 Collider versus Chain

One might worry that we haven't picked out the right causal structure and that on an alternative causal structure responses to the Poison Cup case do not indicate a failure of compositionality. Specifically, one might argue that Amy's action and Courtney's action should be thought of as operating conjointly to bring about Jessica's death, rather than in sequence. Restricting attention to a causal model with variables for Amy's action, Courtney's action, and Jessica's status, the two alternatives are pictured in Figure 5.

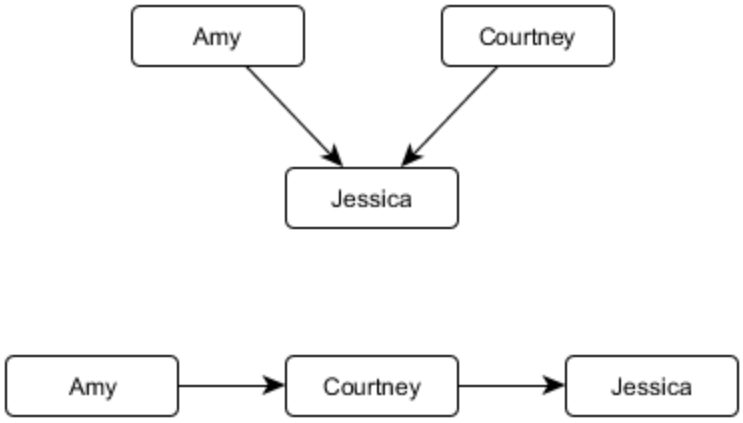


**Figure 5:** Collider Model (top) and Chain Model (bottom) for the Poison Cup Case

In line with the actions specified in the follow-up study above, we might construe the relevant actions as Amy poisoning the cup and Courtney giving Jessica juice in the cup. Construed in this way, either event might have occurred without the other. On such a construal, the Poison Cup case is best modeled with a collider. If participants were operating with such a model, then their responses would not violate the compositionality constraint, since Courtney is not an intermediary between Amy and Jessica in a collider model.[23]

There are reasons to believe that the folk model in the Poison Cup case really is sequential (i.e. a chain model). One reason to think that the folk model is sequential is that it was

[23] An anonymous referee argued that the normatively correct model has both a collider and a chain. On the one hand, Amy hires Courtney to babysit, and hence, Amy causes Courtney to cause Jessica's death. On the other hand, the events *the cup contains poison* and *the cup is given to Jessica* are independent in the sense that each might happen without the other happening, and the variables representing those events are independently manipulable. Therefore, in the normatively correct model, variables for the poisoning and for the giving should form a collider at a variable representing the event *Jessica ingests poison* or the event *Jessica dies*. We agree that one can model the system in this way. But we are not sure that the referee's model is normatively correct. This depends on whether there is a unique, correct choice of variables to use in building the model, and we are not confident that there is. If we use the variables *Amy* [plans to kill/doesn't], *Poison Cup* [yes/no], *Hire Courtney* [yes/no], *Courtney Gives Cup to Jessica* [yes/no], and *Jessica Dies* [yes/no], then the referee is correct in thinking that there is both a causal chain from *Amy* to *Jessica Dies* by way of *Hire Courtney* and *Courtney Gives Cup to Jessica*. But there are plausible alternative choices of variables. We try to induce such an alternative in Study 6.

the first model that we constructed for the scenario. Only after thinking about the case for an extended period did one of us start to worry that the case might be modeled by a collider, and we (the authors) continue to disagree about how plausible the collider model actually is. But the fact that we both initially wanted to model the case with a chain and only later worried about a collider structure is some (admittedly weak) reason to think that participants in our studies modeled the case with a chain. Moreover, the chain model better captures a key element of the mechanism in the Poison Cup case, so it is more likely that participants operated with such a model. Amy *used* Courtney as a delivery device for the poison. Courtney is part of the mechanism that Amy used to bring about Jessica's death. Hence, Courtney was a salient intermediary on the path the poison takes from Amy's hand to Jessica's stomach.

One might accept that the chain model is the most natural, however, while still worrying that participants might have been operating with a collider model. Whether participants in our studies used a collider model or a chain model to represent the Poison Cup case is an empirical matter. But exactly how to settle the question is not obvious. One plausible option is to emphasize the transfer of the poison, describing Courtney as giving Jessica the *poisoned* cup. Described in this way, a collider model seems far less plausible, since Courtney could not have given Jessica the poisoned cup if Amy had not first laced the cup with poison. By emphasizing the transfer of the poison, we implicitly encourage participants to use the variables *Amy* [plans to kill/doesn't], *Poison Cup* [yes/no], *Hire Courtney* [yes/no], *Courtney Gives Poisoned Cup to Jessica* [yes/no], and *Jessica Dies* [yes/no]. Given that collection of variables, the normatively correct model for the story has a collider at *Courtney Gives Poisoned Cup to Jessica*, but it is not possible to assert that Amy poisoning the cup caused Jessica's death and also that Courtney

giving the poisoned cup to Jessica did *not* cause Jessica's death without violating the compositionality constraint.

Hence, in our sixth study, we gave participants the vignette used in our first study, but we changed the statements that we asked participants to evaluate in order to emphasize the transfer of the poison, as follows: (1) Amy caused Jessica's death by lacing the cup with poison; (2) Courtney caused Jessica's death by giving her the cup laced with poison. Responses were collected from 27 participants.[24] Again, almost every participant (22/27) completely agreed with the first claim (M=6.56), and more than half (15/27) completely disagreed with the second claim (M=2.48), which again suggests that the causal attributions of most ordinary people do not satisfy the compositionality constraint for the Poisoned Cup case (Figure 6).[25]

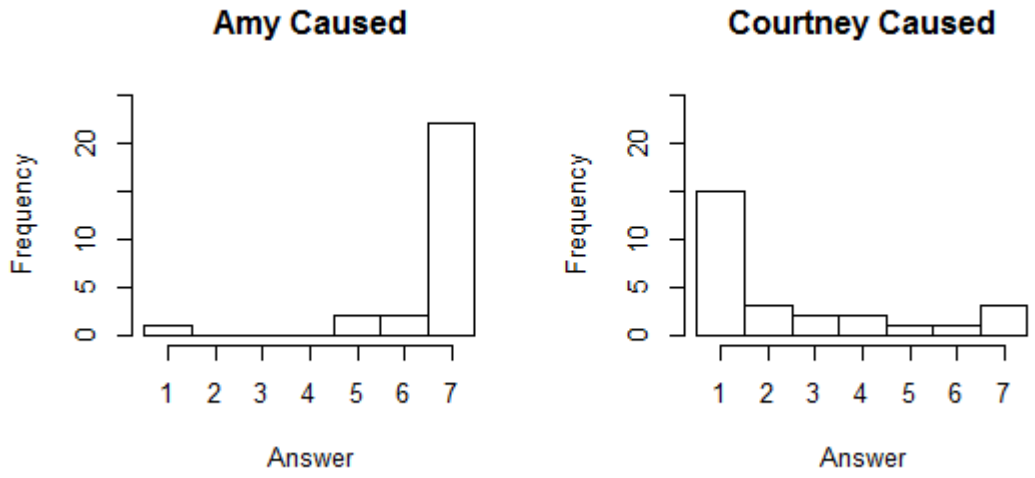


**Figure 6:** Histograms for Study 6

[24] 59.3% women, average age 34.8, ranging from 18 to 64. The number of participants was small in this study because we were only confirming the existence of a large effect, not trying to precisely estimate the effect size.
[25] AMY: V=362, p=0.0003539, Δ=0.926 [0.605, 0.988]; COURTNEY: V=57, p=0.001538, Δ=0.556 [0.181, 0.789].

## 3. Further Studies

At this point, we think we have made a prima facie case that ordinary causal attributions do not tend to satisfy the compositionality constraint for all cases. In this section, we present some further studies that expand on our initial findings and strengthen our prima facie case.

### *3.1 Poisoned Candy*

We begin with a case that is similar to the Poison Cup case but includes both a collider and a chain. In our seventh study, participants were given the following vignette:

> Amy is a seven-year-old girl visiting an amusement park on a school trip. Amy's mother, Gwen, signed the permission slip allowing Amy to go on the school trip. Elizabeth is a park worker who is paid to give out free candy samples to children as part of a marketing strategy. Elizabeth has no idea that the candy she is giving out has been poisoned by the evil Cruella, who hates both children and amusement parks.
>
> As it happens, Elizabeth offers Amy some candy, which was poisoned by Cruella. Amy takes the candy, eats it, and dies.

Participants were then asked to rate their level of agreement or disagreement with four causal attributions using the same seven-point scale as in the previous studies: (1) Cruella caused Amy's death by poisoning the candy; (2) Gwen caused Amy's death by allowing her to go on the school trip; (3) Amy caused her own death by eating the candy; (4) Elizabeth caused Amy's death by giving her the candy.

The Poison Candy case differs from the previous scenario in that the poisoning isn't directed at a specific individual. Another difference is that we've added a description of an additional agent (Gwen) who makes a decision that ends up putting the victim (Amy) into harm's way. In this case, four key events were involved in bringing about Amy's death: Cruella poisoned the candy, Gwen allowed Amy to go to the amusement park where Elizabeth could give her the poisoned candy, Elizabeth gave Amy the poisoned candy, and Amy ate the poisoned

candy. What is the causal structure for this case? Following the reasoning given for the previous case, it seems that Cruella, Elizabeth, and Amy stand in a causal chain, since each is involved in transferring the poison. Gwen is not part of this chain, however, since she had nothing to do with the poison. Nonetheless, she did play a role in bringing about the outcome since she put Amy into position to eat the poisoned candy. Cruella poisoning the candy *and* Gwen allowing Amy to go to the park were necessary for Elizabeth to be able to give Amy poisoned candy.

We predicted that participants would judge that Cruella caused Amy's death. Assuming the model suggested above, compositionality would then predict that people would also judge that Elizabeth caused Amy's death by giving her the candy and that Amy caused her own death by eating the candy. In contrast, based on the responsibility view we predicted that participants would only affirm that Cruella caused Amy's death, since Cruella, and Cruella alone, violated a clear injunctive norm and should be held responsible for the outcome.

Responses for the Poison Candy case were collected from 60 participants.[26] Again, the results were in line with the prediction based on the responsibility view and by and large did not abide by the compositionality constraint (Figure 7): almost every participant (56/60) completely agreed with the claim that Cruella caused the death (M=6.78), while half completely disagreed that Elizabeth caused the death (M=2.57) and a majority completely disagreed that Amy caused her own death (36/60; M=2.28). In addition, a large majority completely disagreed that Gwen caused Amy's death (48/60; M=1.57).[27]

[26] 75.0% women, average age 33.2, ranging from 18 to 60

[27] CRUELLA: $V=1796.5$, $p=1.523e^{-13}$, $\Delta=0.933$ [0.766, 0.982]; ELIZABETH: $V=250$, $p=3.351e^{-6}$, $\Delta=0.483$ [0.245, 0.667]; AMY: $V=162.5$, $p=4.997e^{-8}$, $\Delta=0.583$ [0.356, 0.746]; GWEN: $V=46.5$, $p=6.082e^{-12}$, $\Delta=0.8$ [0.602, 0.905].

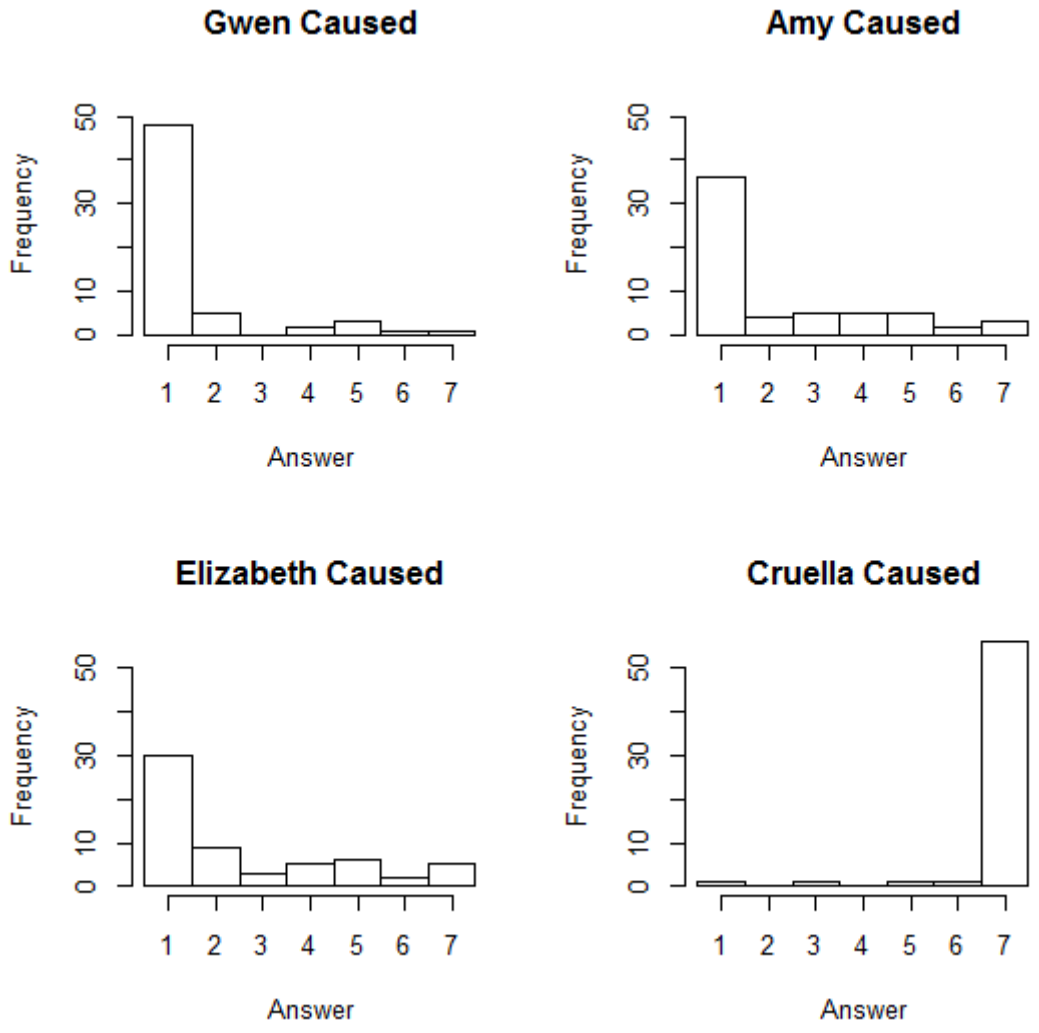


**Figure 7:** Histograms for Study 7

*3.2 Revolver*

The intermediaries in the cases we've looked at so far have been agents, but we expect similar considerations to hold when they are what philosophers would consider non-agentive objects.[28] Specifically, we predict that when an agent brings about an outcome via a series of properly

[28] Research by Bloom (2007), Rose (2015), and Rose and Schaffer (2017), among others, suggests that people tend to take an agentive perspective on nature as a whole, which provides reason to expect that people will hold "non-agents" responsible.

functioning mechanical intermediaries, people won't treat the intermediaries as being responsible for the outcome, and hence won't judge that they caused the outcome. In our eighth study, we gave participants the following vignette:

> Trent has decided to kill his father, Brad. He aims his loaded revolver at Brad and pulls the trigger, releasing the hammer. The hammer strikes the cartridge, igniting the gun powder. The gun powder explodes, driving the bullet from the gun. The bullet hits Brad in the head. He dies instantly.

Participants were then asked to rate their level of agreement or disagreement with four causal attributions using the same seven-point scale as in the previous studies: (1) Trent caused Brad's death; (2) the hammer caused Brad's death; (3) the gun powder caused Brad's death; and (4) the bullet caused Brad's death.

Our causal model for this story is a chain: Trent pulls the trigger, which causes the hammer to strike, which causes the gunpowder to explode, which causes the bullet to hit Brad, which causes Brad to die. Given this model, if ordinary causal attributions uphold the compositionality constraint and people say that Trent caused Brad's death, then they should say that the hammer, the gunpowder, and the bullet all caused Brad's death as well.

Responses for the Revolver case were collected from 51 participants.[29] Again, the results were striking (Figure 8): 43 participants completely agreed that Trent caused Brad's death, and another five responded with a six (M=6.71). The central tendency of the responses for Trent was statistically significantly above 4, and the difference was very large.[30] By contrast, the modal answer for the hammer and for the powder was complete *dis*agreement. The central tendency of the responses for the hammer was statistically *below* 4, and the difference was moderate.[31] Similarly, the central tendency of the responses for the powder was statistically below 4, and the

[29] 74.5% women, average age 36.2, ranging from 18 to 68
[30] TRENT: V=1296.5, $p=5.277e^{-11}$, $\Delta=0.961$ [0.785, 0.993]
[31] HAMMER: V=182.5, $p=2.989e^{-5}$, $\Delta=0.529$ [0.280, 0.712]

difference was large.[32] The responses for the bullet were bimodal with slightly more participants answering “completely agree” than answering “completely disagree.” The central tendency of the responses for the bullet was not statistically different from 4.[33]

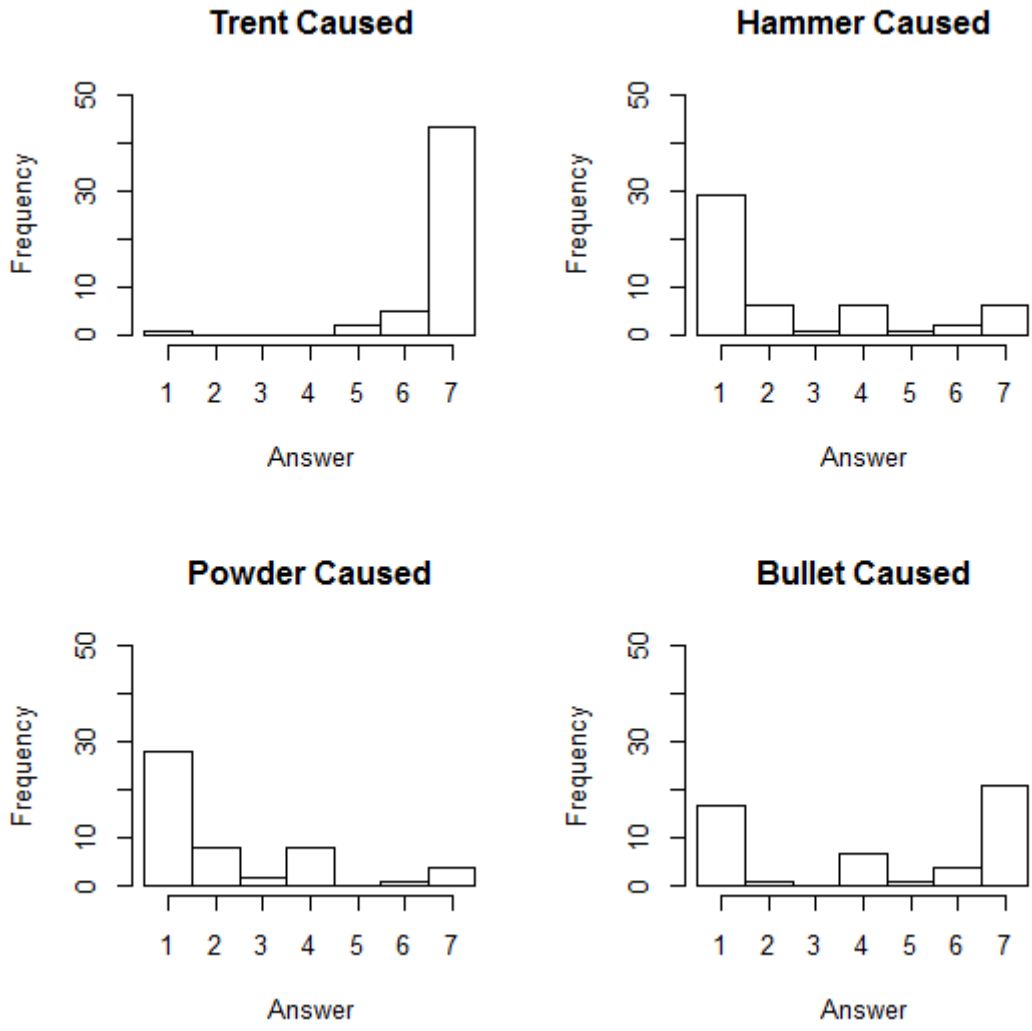


**Figure 8:** Histograms for Study 8

[32] POWDER: V=117, p=2.963e$^{-6}$, Δ=0.647 [0.428, 0.794]
[33] BULLET: V=552.5, p=0.469, Δ=0.157 [-0.105, 0.398]

*3.3 GFCI*

In the Revolver case, we asked participants to consider a case where the causal intermediaries are mechanical, and we saw the same pattern of results as in previous studies. However, the Revolver case still has a person acting to set events in motion and a person being harmed. Many people (including an anonymous referee) have expressed concern that lay people might treat causation differently when they are considering cases involving agents than when they are considering cases involving non-agents. They worry that if people think differently about causation when agents are involved, then the results of studies like those we have conducted are not informative about the ordinary concept of causation. One way to express the worry is as a variation on the first objection to our initial study. People are confusing causation with responsibility. By presenting cases involving agents, we are encouraging that confusion. Moreover, the objection continues, if we had presented cases that did not involve agents, we would not have seen notable violations of the compositionality constraint.

An anonymous referee suggested that we could provide evidence against an evaluative reading of "caused" by presenting cases involving only inanimate objects. As the referee put it, "With no agents involved, the worry that a cause question is being interpreted as a blame question would be largely addressed, except perhaps for the possibility of metaphorical attributions of blame to inanimate occurrences." Hence, in our ninth study, we tried to minimize the active role of agents while still having a case that we expected to show violation of the compositionality constraint. We gave participants the following vignette:

> John is a scientist conducting a very important experiment on an unusual species of plant. His experiment requires growing his plants under a special light, which is plugged into an outlet with a ground fault circuit interrupter (GFCI) safety mechanism. The pipes running to John's laboratory were correctly manufactured and installed, and the system was protected from any changes in weather condition.

> Despite there being nothing wrong with the pipes, one day a pipe burst in John's laboratory. Water ran into the outlet powering the special light. A properly functioning GFCI safety mechanism will break the circuit so that no power flows through its outlet if exposed to water in this way. And in fact, the GFCI safety mechanism did break the circuit. The special light turned off and the experiment was ruined.

Participants were then asked to rate their level of agreement or disagreement with two causal attributions using the same seven-point scale as in our other studies: (1) The pipe bursting caused the experiment to be ruined; and (2) the GFCI safety mechanism breaking the circuit caused the experiment to be ruined.

Our causal model for this story is a chain: a pipe bursts, which causes water to run into the outlet, which causes the GFCI to break the circuit, which causes the light to turn off, which causes the experiment to be ruined. The pipe bursting doesn't *directly* cause the experiment to be ruined. If the pipe bursting causes the experiment to be ruined, it does so by way of the GFCI. Given this model, if folk causal attributions satisfy the compositionality constraint and people say that the pipe bursting caused the experiment to be ruined, then they should also say that the GFCI safety mechanism breaking the circuit caused the experiment to be ruined. The driving thought behind the GFCI case was that in cases where some (ultimately bad) random event occurs, if a mechanism does what it is *supposed to do* and in virtue of doing what it is supposed to do, a bad outcome results, people will treat causal responsibility as being transmitted by the mechanism while denying that the mechanism is itself causally responsible for the bad outcome.

Responses for the GFCI case were collected from 163 participants.[34] The results were mostly consistent with our account of when and why compositionality fails with respect to ordinary causal attributions (Figure 9): 72 participants completely agreed that the pipe bursting caused the experiment to be ruined, and another 30 responded with a six. The central tendency of

[34] 72.4% women, average age 45.1, ranging from 16 to 77

the responses for the pipe bursting (M=5.48) was statistically significantly above 4, and the difference was moderate in size.[35] By contrast, the modal answer for the GFCI was complete *dis*agreement. Moreover, the central tendency of the responses for the GFCI (M=3.67) was statistically *below* 4, though the difference was practically non-existent.[36]

However, directly comparing responses for the pipe to responses for the GFCI paints a different picture. Direct comparison suggests that we are probing two sub-populations that think differently about causation. Ratings for the pipe and for the GFCI were statistically different, though the effect was small.[37] However, when we plot the pairwise differences, we see two things. First, we see (as we expected) that many individuals were more willing to say that the pipe bursting as opposed to the GFCI breaking the circuit caused the experiment to be ruined. But second, we see that 48 participants gave the *same rating* for both the pipe and the GFCI, which is consistent with the compositionality constraint. Moreover, a difference of zero was the mode for this distribution! To be sure, more people gave a higher rating for the pipe: 87 in total gave a higher rating for the pipe than for the GFCI, and 59 of those gave a rating for the pipe that was 4, 5, or 6 points greater than their rating for the GFCI. But as you can see from the bottom row of Figure 9—where we have plotted the pairwise differences for our first study next to the pairwise differences for Study 9—those 48 participants are striking.

[35] PIPE: V=9498, $p=2.73e^{-15}$, $\Delta=0.564$ [0.434, 0.671]
[36] GFCI: V=4191, p=0.025, $\Delta=0.117$ [-0.029, 0.257].
[37] Wilcoxon matched-pairs signed rank test rejected the hypothesis that the central tendency for "Pipe" was equal to the central tendency for "GFCI" (V=19473, $p=6.335e^{-14}$). The probability of superiority for "Pipe" over "GFCI" was 0.534, $\Delta=0.362$ [0.239, 0.473].

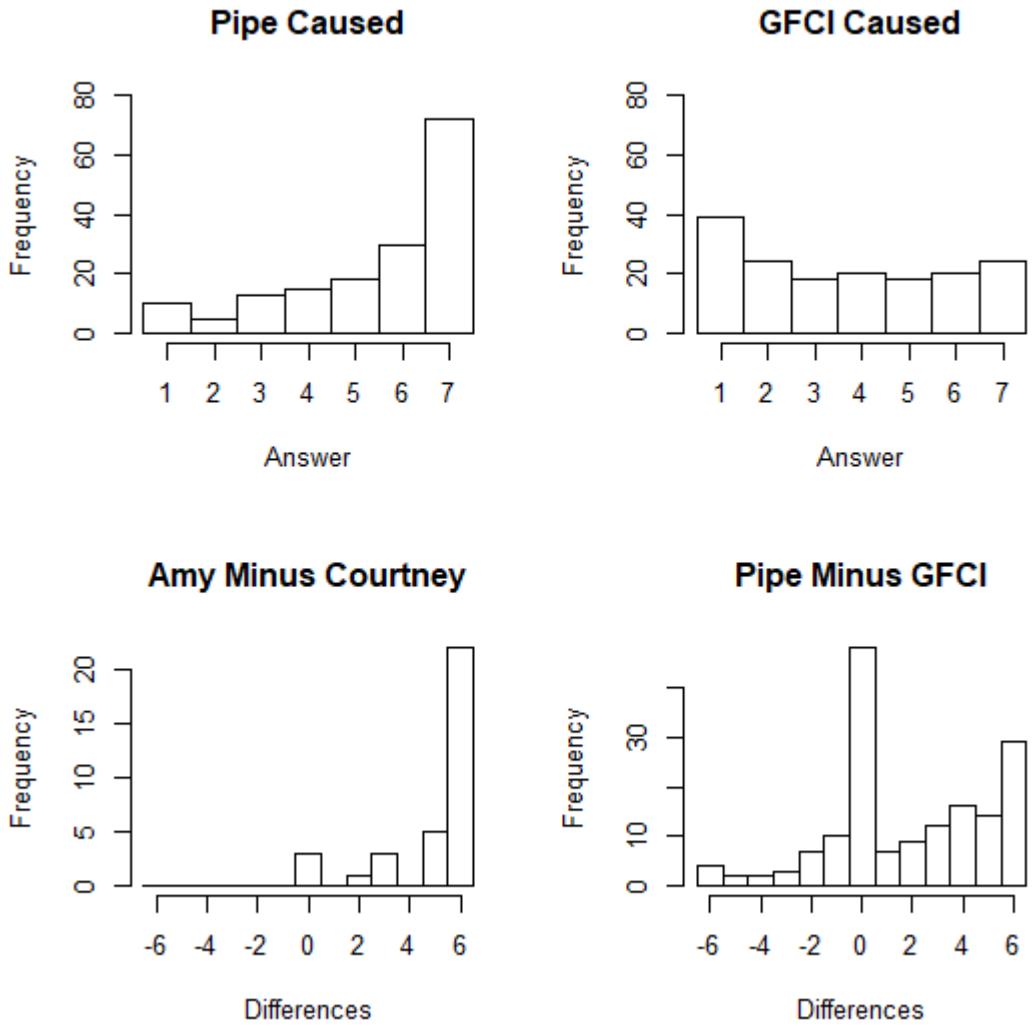


**Figure 9:** Histograms for Study 9

## 4. Concluding Remarks à la Mode

We have seen that many theories of actual causation are implicitly committed to the compositionality constraint, and we find the constraint plausible. What, then, should we say in light of the fact that we found that most ordinary causal attributions violate the constraint across the four cases tested? Here are two obvious answers: (1) Follow the FAD and reject the constraint, or (2) accept the constraint and abandon the FAD. We think both answers are too

extreme and come from neglecting the variety of projects under the umbrella of research on causation. In closing, we consider four such projects—conceptual analysis, conceptual engineering, descriptive psychology, and realist metaphysics. We suggest that philosophers may accept or reject the FAD to differing degrees with respect to different projects and that it is possible to consistently pursue several of these projects simultaneously.[38]

Let us begin with one motivation for following the FAD—conceptual analysis. Some philosophers have taken themselves to be analyzing the ordinary concept of causation. If one's goal is to analyze our pre-theoretic concept, then one is likely to find the FAD to be a natural constraint. Concepts must at some point and in some way give rise to behavior. Plausibly, ordinary causal attributions are an exercise of the ordinary concept of causation. If so, then one might think that the ordinary concept of causation may be analyzed by carefully examining ordinary causal attributions.

However, even philosophers, such as Lewis and Menzies, explicitly giving analyses of the ordinary concept of causation have offered theories that entail the compositionality constraint. How could they have gotten things so wrong? One possibility is that our empirical evidence is misleading. Alternatively, one might argue that concepts cannot be studied effectively by way of introspection. No armchair reflection reliably reveals what our concepts are really like. Better tools are called for.[39]

[38] Our conclusion is largely in agreement with methodological remarks by Armstrong (2004), Hall (2004), Collins, Hall, and Paul (2004), Paul and Hall (2013), and Bernstein (2017). Specifically, we agree that what we should say in light of observations of ordinary causal attributions should be sensitive to the philosophical project at stake. However, we have a different way of categorizing philosophical projects and different opinions with respect to how information about ordinary causal attributions engages those projects.

[39] For introductions to some tools to better investigate concepts, see Sytsma and Livengood (2015) and Sytsma et al. (2019).

We think skepticism about introspection is healthy, but abandoning the armchair altogether is not. A better answer begins by reflecting on the target of the analysis. Lewis (1973, 558-559) writes:

> 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 as the "causes," calling the rest mere "causal factors" or "causal conditions." Or we speak of the "decisive" or "real" or "principal" 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 principles of invidious discrimination. I am concerned with the prior question of what it is to be one of the causes (unselectively speaking). My analysis is meant to capture a broad and nondiscriminatory concept of causation.

Plausibly, Menzies was also interested in a broad, nondiscriminatory concept of causation. But is the nondiscriminatory concept of causation *our pre-theoretic concept?* One might suppose that we really share a nondiscriminatory concept that is later biased by principles of invidious discrimination or that we really share a concept of causation that has purely non-evaluative content but that application is later biased by evaluative considerations. We argued in Section 2.2.1, and in greater detail in Sytsma and Livengood (MS), that bias views are not adequate to the empirical data. And we agree with Schaffer (2005, 344) when he observes that some discriminatory character is "*an inseparable aspect of our causal concept*." The ordinary default concept is itself discriminatory and evaluative.

One might stop here and wonder whether we have any non-evaluative, egalitarian concept of causation at all. As Schaffer (2005, 344) puts it: "Perhaps the idea of a 'broad and nondiscriminatory concept' [of causation] is a *philosopher's myth*." If that's right, then conceptual analyses are perhaps better thought of as attempts to engineer a more theoretically useful concept of causation: one that might be only loosely connected with the ordinary default concept. However, the conceptual analyst might retrench in a different way by supposing that

there are multiple concepts of causation that have equal or near equal claim to the title of *the ordinary concept of causation*.

One might follow Danks, Rose, and Machery (2014) in pointing out that causal perception is almost certainly non-evaluative in character and that at least some causal reasoning also seems to be non-evaluative in character. Perhaps those cognitive activities involve a non-discriminatory, non-evaluative concept of causation and philosophical theories should be understood as targeting *that* concept. One might maintain that our studies only ever access the discriminatory, evaluative concept and still hold out hope that the compositionality constraint is true of the nondiscriminatory, non-evaluative concept. All of that may very well be correct. But if so, then we are owed an account of how we can isolate the nondiscriminatory, non-evaluative concept experimentally.[40] Moreover, we will need to experimentally check to see whether that concept satisfies the compositionality constraint. Until such research is conducted, philosophers aiming to analyze the ordinary concept of causation ought to be suspicious of the constraint.

Although many philosophers have maintained that conceptual analysis is central to philosophical inquiry, there are very few, if any, uncontroversial examples of successful analysis. However, *attempts* at conceptual analysis may be a valuable preliminary part of the related project of conceptual engineering. By exploring a wide range of cases, building theories, and

[40] An anonymous reviewer suggested that Danks et al. (arguably) provide a procedure for probing such a concept experimentally. While space prevents an extended discussion of their experimental work, we do not believe their probes do this. Across three experiments they tested "ratings of causal strength" between what they consider to be moralized and non-moralized scenarios, finding no effect for moralization. While these experiments are couched as testing judgments about causal strength, the questions posed to participants instructed: "Respond with -100 [or -10] if you think that [factor] always [prevents outcome]. Respond with +100 [or +10] if you think that [factor] always [produces outcome]. And respond 0 if you think [factor] is irrelevant for [outcome]." Judgments concerning "causal strength" described in this way look rather different from the types of judgments that conceptual analysts have typically called on in investigating the ordinary concept of causation, including that these judgments would appear to be more closely related to structural causation than actual causation. As such, conceptual analysts would owe us an account of why "causal strength" judgments should be thought to better isolate the concept of interest and reasons to believe that the intuitions called on in the literature reflect ordinary judgments concerning "causal strength," as well as evidence that such judgments abide by the compositionality constraint.

challenging them with counter-examples, would-be conceptual analysts provide a lot of material for developing conceptual tools to solve specific problems. Reflecting on patterns of ordinary causal attribution might be quite valuable to a conceptual engineer, but insofar as one is interested in building useful conceptual tools, one need not be *constrained* by them. Instead, conceptual engineers are constrained by the specific problems they face. But when we have specific problems in mind, our theories are constrained by measures of performance with respect to those problems. The clearest and most interesting example of this approach to actual causation that we know of is set out in Halpern (2016).

In addition to conceptual analysis and conceptual engineering, one might be interested in descriptive psychology. We take descriptive psychology to be a philosophical enterprise that may include conceptual analysis as a part but is typically broader in scope. For example, one might describe the cognitive processes leading from environmental stimuli to causal attributions without ever characterizing any concepts involved. A descriptive psychologist might doubt that humans have any concepts at all! The philosophical value of descriptive psychology is grounded in the project of understanding ourselves—our reasoning abilities, our language, and our forms of life. The studies described in this paper are obviously only a single thread in a great tapestry of such inquiry. We think that something like the FAD is a natural commitment in pursuing descriptive psychology. However, we also think that there are projects in descriptive psychology that need not endorse it. For example, one might be interested in what sorts of factors influence ordinary causal attributions without ever trying to produce a theory that predicts exactly what the attributions will be in any given case.

Finally, we consider the project of realist metaphysics. When thinking about the point of theorizing about actual causation, one might be tempted to say, “I’m trying to figure out what

*causation* is, of course!" The thought is that we study the causal relation in just the same way that we study electrons, temperature, or time. When we theorize about actual causation, we are theorizing about a relation in the world. Investigation proceeds by first fixing a target—perhaps by pointing to obvious examples—and then studying how that thing works. We may be surprised by what we learn about actual causation, just as we have been surprised by what we have learned about electrons, temperature, and time, and mature metaphysical theories of actual causation may turn out to be very different from our ordinary conception. Here, then, the FAD is reasonable only insofar as one is worried about having inadvertently changed the subject, as philosophers working on free will are sometimes accused of doing, or insofar as one antecedently thinks that ordinary causal attributions are likely to be *correct*.

Summing up, we have identified what we take to be a plausible constraint on descriptive theories of actual causation. We have provided evidence that ordinary causal attributions do not always satisfy the constraint. We then stepped back to observe that different philosophical projects are constrained by ordinary causal attributions to different degrees. We conclude with a bit of modest methodological advice. Philosophers can and should pursue multiple projects at once. But we should be clear about what our projects are and keep the constraints on our theorizing clearly in view.

**References**


Alicke, M. (1992). Culpable Causation. *Journal of Personality and Social Psychology*, 36: 368–378.

Armstrong, D. (2004). Going Through the Open Door Again: Counterfactual versus Singularist Theories of Causation. In Collins, Hall, and Paul (eds.) *Causation and Counterfactuals*. MIT Press, 445-457.

Bernstein, S. (2017). Intuitions and the Metaphysics of Causation. In Rose (ed.), *Advances in Experimental Metaphysics*.

Blanchard, T. and Schaffer, J. (2017). Cause without Default. In Beebee, Hitchcock, and Price (eds.), *Making a Difference*. Oxford University Press.

Bloom, P. (2007). Religion is Natural. *Developmental Science*, 10, 147–151.

Chockler, H. and Halpern, J. Y. (2004). Responsibility and Blame: A structural-Model Approach. *Journal of Artificial Intelligence Research*, 93-115.

Collins, J., Hall, N., and Paul, L., eds. (2004). *Causation and Counterfactuals*. Cambridge: MIT Press.

Danks, D., Rose, D., & Machery, E. (2014). Demoralizing causation. *Philosophical Studies*, *171*(2), 251-277.

Dowe, P. (1995). Causality and Conserved Quantities: A Reply to Salmon. *Philosophy of Science*, *62*(2), 321-333.

Dowe, P. (2000). *Physical Causation*. Cambridge: Cambridge University Press.

Dowe, P. (2008). Causal Processes. In Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. URL = <https://plato.stanford.edu/archives/fall2008/entries/causation-process/>.

Ehring, D. (1997). *Causation and Persistence*. Oxford: Oxford University Press.

Ehring, D. (2004). Counterfactual Theories, Preemption and Persistence. In Dowe and Noordhof (eds.) *Cause and Chance*. London: Routledge, 58-76.

Glymour, C., Danks, D., Glymour, B., Eberhardt, F., Ramsey, J., Scheines, R., ... and Zhang, J. (2010). Actual Causation: A Stone Soup Essay. *Synthese*, *175*(2), 169-192.

Glynn, L. (2011). A Probabilistic Analysis of Causation. *British Journal for the Philosophy of Science*, 62, 343-392.

Hall, N. (2004). Two Concepts of Causation. In Collins, Hall, and Paul (eds.) *Causation and Counterfactuals*. MIT Press, 225-276.

Halpern, J. Y. (2016). *Actual Causality*. MIT Press.

Halpern, J. Y. and Hitchcock, C. (2015). Graded Causation and Defaults. *The British Journal for the Philosophy of Science*, *66*(2), 413-457.

Halpern, J. Y. and Pearl, J. (2005). Causes and Explanations: A Structural-Model Approach. Part I: Causes. *The British journal for the philosophy of science*, *56*(4), 843-887.

Hilton, D. and Slugoski, B. (1986). Knowledge-Based Causal Attribution: The Abnormal Conditions Focus Model. *Psychological Review*, *93*(1), 75-88.

Hitchcock, C. (2001). The Intransitivity of Causation Revealed in Equations and Graphs. *The Journal of Philosophy*, *98*(6), 273-299.

Hitchcock, C. (2007a). Prevention, Preemption, and the Principle of Sufficient Reason. *The Philosophical Review*, *116*(4), 495-532.

Hitchcock, C. (2007b). Three Concepts of Causation. *Philosophy Compass* 2/3, 508-516.

Hitchcock, C. and J. Knobe (2009). Cause and Norm. *Journal of Philosophy* 106(11), 587-612.

Knobe, J. (2006). *Folk Psychology, Folk Morality*. Dissertation.

Kominsky, J. F., Phillips, J., Gerstenberg, T., Lagnado, D., and Knobe, J. (2015). Causal Superseding. *Cognition*, *137*, 196-209.

Lagnado, D., Gerstenberg, T., and Zultan, R. (2013). Causal Responsibility and Counterfactuals. *Cognition*, *37*, 1036-1073.

Lewis, D. (1973). Causation. *The Journal of Philosophy, 70*(17), 556-567.

Lewis, D. (1986). Postscripts to Causation. In *Philosophical Papers: Volume 2*. Oxford University Press, 172-213.

Lewis, D. (2000). Causation as Influence. *The Journal of Philosophy*, *97*(4), 182-197.

Livengood, J. (2013). "Actual Causation and Simple Voting Scenarios." *Noûs*, 47(2): 316–345.

Livengood, J., J. Sytsma, and D. Rose (2017). "Following the FAD: Folk Attributions and Theories of Actual Causation." *Review of Philosophy and Psychology*, 8(2): 274–294.

McDermott, M. (1995). Redundant Causation. *British Journal for the Philosophy of Science*, 46, 523-544.

McGrath, S. (2005). Causation by Omission: A Dilemma. *Philosophical Studies, 123*(1-2), 125-148.

McGraw, K., and Wong, S. (1992). A Common Language Effect Size Statistic. *Psychological bulletin*, *111*(2), 361.

Menzies, P. (1989). Probabilistic Causation and Causal Process: A Critique of Lewis. *Philosophy of Science*, 56(4), 642-663.

Paul, L., and Hall, N. (2013). *Causation: A User's Guide*. Oxford: Oxford University Press.

Reichenbach, H. (1956). *The Direction of Time*. Los Angeles: University of California Press.

Rogmann, J. (2013). Ordinal Dominance Statistics (orddom): An R Project for Statistical Computing package to compute ordinal, nonparametric alternatives to mean comparison (Version 3.1). Available online from the CRAN website http://cran.r-project.org/.

Rose, D. (2015). Persistence Though Function Preservation. *Synthese*, 192, 97-146.

Rose, D. and Schaffer, J. (2017). Folk mereology is teleological. *Noûs*, *51*(2), 238-270.

Salmon, W. (1994). Causality without Counterfactuals. *Philosophy of Science* 61(2), 297-312.

Salmon, W. (1997). Causality and Explanation: A Reply to Two Critiques. *Philosophy of Science* 64(3), 461-477.

Schaffer, J. (2005). Contrastive causation. *Philosophical Review*, *114*(3), 327-358.

Sytsma, J. (MS). "The Extent of Causal Superseding."

Sytsma, J., R. Bluhm, P. Willemsen, and K. Reuter (2019). "Causal Attributions and Corpus Analysis." In E. Fischer and M. Curtis (eds.), *Methodological Advances in Experimental Philosophy*, Bloomsbury Press.

Sytsma, J. and J. Livengood (2015). *The Theory and Practice of Experimental Philosophy*. Broadview Press.

Sytsma, J. and J. Livengood (MS). "Intervention, Bias, Responsibility… and the Trolley Problem."

Sytsma, J., J. Livengood, and D. Rose (2012). "Two Types of Typicality: Rethinking the Role of Statistical Typicality in Ordinary Causal Attributions." *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43: 814–820.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford University Press.