Cause, 'Cause', and Norm[1]

John Schwenkler and Eric Sievers

1. Introduction

In an important passage in her seminal paper "Causality and Determination", G. E. M. Anscombe calls attention to the variety of words that are used to express causal concepts:

> The word 'cause' itself is highly general. […] I mean: the word 'cause' can be *added* to a language in which are already represented many causal concepts. A small selection: *scrape*, *push*, *wet*, *carry*, *eat*, *burn*, *knock over*, *keep off*, *squash*, *make* (e.g. noises, paper boats), *hurt*. (Anscombe 1981, p. 137)

The italicized words are quite diverse, but these "special causal verbs" (ibid.) are unified conceptually in that they all describe kinds of causation: that is, ways of *making* things happen, of *bringing about* events or changes of state in oneself or other things, of *creating* things or *destroying* them—most generally, of *acting on* things in such a way that in virtue of this activity, those things end up different than they otherwise would be.

be, **have**, do, say, go, can, **get**, would, **make**, know, will, think, **take**, see, come, could, want, look, **use**, find, **give**, tell, work, may, should, call, try, ask, need, feel, become, **leave**, **put**, mean, **keep**, **let**, **begin**, seem, **help**, talk, **turn**, **start**, might, **show**, hear, **play**, **run**, **move**, like, live, believe, **hold**, **bring**, happen, must, **write**, **provide**, **sit**, **stand**, lose, **pay**, meet, include, **continue**, **set**, learn, **change**, **lead**, understand, watch, follow, **stop**, **create**, speak, read, **allow**, **add**, **spend**, **grow**, **open**, **walk**, win, offer, remember, love, consider, appear, **buy**, wait, **serve**, die, **send**, expect, **build**, stay, fall, **cut**, reach, **kill**, remain

Table 1: The 100 most commonly used English verbs according to the Corpus of Contemporary American English. Those in **boldface** can be used to express causal concepts. From <http://www.wordfrequency.info>, retrieved 14. June, 2019.[2]

[1] David Rose was significantly involved in this project at an early stage, and we thank him for his valuable contributions. For discussion of this material, we thank Samantha Berthelette, Randy Clarke, Josh Knobe, Justin Sytsma, several members of the FSU Philosophical Psychology Group, and an audience at the 2020 meeting of the SSPP. We also thank Kevin Reuter for detailed comments on an earlier version of the chapter.

[2] There is no canonical list of which English verbs are causal, nor any uncontroversial criterion for determining which verbs are. (For an attempt at doing both of these things, see Schwenkler (in prep.).) As such, the decision of which verbs to highlight in Table 1 reflect our own considered judgments. To touch on a few of the controversial or borderline cases: we judge that *have* is used causally in a sentence like 'Beth had a milkshake' (i.e., she consumed it), *get* is used causally in 'Jim got the keys out of the drawer' (i.e., he removed them), *sit* is used causally in 'Annie sat the doll in the chair' (i.e., she made it to sit there), and

Special causal verbs like those in Anscombe's list make up a great part of our ordinary ways of describing causation. Indeed, among the hundred most commonly used verbs according to the Corpus of Contemporary English (see Table 1), about 40 can be used in ways that express causal concepts in the broad sense indicated above. If, for example, someone says that

(1) Tony *had* the cup of coffee that Beth *got* for him to drink, *using* to sweeten it some of the sugar that Keith had *put* in the cupboard.

—then she says a great deal about what in philosophy we might want to call Tony's, Beth's, and Keith's having *caused* various things to be the case, or to happen. Yet the sentence in (1) does this *without* using a "highly general" word like 'cause' or any of its variants. Indeed, the verb 'cause' is not itself very commonly used: it shows up at #135 on COCA's verb frequency list, behind all the special causal verbs identified in Table 1 as well as several more. When people talk about causing, it is usually these special causal verbs that they use.[3]

Further evidence suggests that when the verb 'cause' *is* used in ordinary speech and writing, this is very often in a way that expresses some "special" variety of causal relation rather than the general relation that is usually the interest of metaphysicians and philosophers of science. For example, Sytsma, Bluhm, Willemsen, and Reuter (2019) present evidence from corpus analysis supporting the hypothesis that the most frequent use of 'cause' is to assign *responsibility* to a person for something *bad* that happened—a pattern of use that is different from the one they observed for other causal verbs like 'create', 'generate', 'induce', 'lead to', 'make', and 'produce'. In addition, Wolff (2003) presents experimental evidence finding that people prefer to use periphrastic causal verbs like 'cause', 'enable', and 'prevent' over lexical causatives like 'break', 'pop', 'melt', and 'burn' when the agent and patient were separated by the operation of an intermediate entity, and vice versa: for example, the statement

(2) The girl caused the vase to break.

---

stand is used causally in 'Al stood the books on the shelf' (i.e., he made them to stand there). However, since the list in Table 1 is merely illustrative, nothing in our argument hinges on whether we are correct in every case, nor on deciding what exactly "correctness" in this context would be.

[3] Notice that this point holds even if many of the words in our list, like 'have', 'get', 'sit', and 'stand', aren't used causally very often (when they are used). It's enough that some of them, like 'put', 'build', 'cut', and 'kill', seem to have *only* a causal use: for the frequency data reveal that we use *each* of these words more often than we use 'cause', thus talking about causation using each of these special causal verbs more often than we talk about it using 'cause'.

is preferred over

(3) The girl broke the vase.

when the vase in question breaks after being hit by a ball that the girl bounces off her foot, whereas (3) is preferred to (2) when the girl throws the ball directly at the vase.[4]

Facts like these, about the ordinary usage of 'cause' and other causal verbs, bear directly on experimental work that uses verbal reports to study the psychology of causal thinking. Consider, for example, the well-known *pen vignette* presented in Knobe and Fraser (2008):

> The receptionist in the philosophy department keeps her desk stocked with pens. The administrative assistants are allowed to take the pens, but faculty members are supposed to buy their own.
>
> The administrative assistants typically do take the pens. Unfortunately, so do the faculty members. The receptionist has repeatedly emailed them reminders that only administrative assistants are allowed to take the pens.
>
> On Monday morning, one of the administrative assistants encounters Professor Smith walking past the receptionist's desk. Both take pens. Later that day, the receptionist needs to take an important message … but she has a problem. There are no pens left on her desk.

Having presented this vignette, Knobe and Fraser found that participants agreed much more with the statement that *Professor Smith* caused the receptionist's problem than the statement that *the administrative assistant* did so—a finding they take to support the hypothesis that "moral judgments … play a direct role in the process by which causal judgments are generated" (Knobe and Fraser 2008, p. 443). In light of the above, however, we should be concerned that it is only one *kind* of causal judgment—namely, the kind made with a statement of the form

(C1) *X* caused *Y*

—where *X* is an intentional agent and *Y* is an outcome with negative valence—that this experiment investigates directly. (Knobe and Fraser's other (2008) experiment has the same structure: it concerns a pair of siblings who log in at the same time to a computer,

[4] For similar findings, see Bar-Asher Siegal, Bassel, and Hagmayer (forthcoming, in prep.)

which then crashes, and participants then consider who *caused* the computer to crash.) And a review of the literature shows that this focus is hardly exceptional in recent experimental work on causal thinking. Consider, for example, the following list of sentences that were used as test statements or test questions in a selection of highly cited studies:

> "Lauren caused the system to crash." (Livengood, Sytsma and Rose, 2017)
>
> "The attending doctor's decision caused the patient's recovery." (Hitchcock & Knobe, 2009)
>
> "Did Sam cause the bottle to fall off the wall?" (Walsh and Sloman, 2011)
>
> "Turnbull caused Poole's death." (Alicke, Rose, and Bloom 2009)
>
> "Billy caused the motion detector to go off." (Kominsky and Phillips 2019)

Despite the many differences between these sentences, they all have in common the use of the word 'cause' as a way to link an agent with an outcome that can be construed as the result of what they did. And this limitation has not stopped philosophers from using the results of experiments like these to support conclusions about *causal judgments* or the ordinary *concept of causation*, and not merely the ordinary use *of the word 'cause'*: for example, alongside Knobe and Fraser's (2008, p. 443) conclusion that "moral judgments … play a direct role in the process by which causal judgments are generated", Hitchcock and Knobe (2009, p. 604) conclude from their findings that the "concept of actual causation enables us to pick out those factors that are particularly suitable as targets of intervention", while Livengood, Sytsma and Rose (2017, p. 292) take their findings to support the hypothesis that "folk causal attributions are inherently normative and are closely related to responsibility judgments". By contrast, the evidence surveyed above suggests that investigations that concern only what we may call *'cause'-judgments*, or judgments that are expressed with the ordinary use of the English word 'cause', may be revealing only of a partial and possibly unrepresentative range of causal thought and talk.

To address this lacuna, the present chapter presents a series of experiments that elicit causal judgments using statements of a different form than (C1) above. In particular, our interest is in exploring the extent to which the previously observed effects of normative considerations on causal judgments of the form (C1) extend as well to those of the form

(C2) *X V*-ed *Y*

—where '*V*' is a "special causal verb" like those in Anscombe's list. Our principal finding is that in many cases the effects do not extend in this way, and moreover that the cases where we *do* find the same pattern are those where the causal verb used has a negative valence of its own. We draw two main conclusions from this finding. First, it supports directly our contention that the almost exclusive focus on statements in the form of (C1) in the experimental study of causal judgments has led to findings that are *unrepresentative* of the full range of ordinary causal thinking, and provides a proof of concept as to how those judgments can be studied in their full variety. Second, as we discuss in more detail below, the results of our experiments provide significant indirect support for the contention that the effect of moral considerations on causal judgments in the form of (C1) reflect the fact that judgments in this form are most often used to assign *responsibility* for an event, and not just to describe the causal structure of what happened. It is not causal judgments in general that result from a process in which prior moral judgments play a role, but perhaps only those judgments that express a determination of moral responsibility.

2. Experiments

The principal aim of our experiments was to explore whether the above-discussed influence of norms on agreement with statements of the form

(C1) *X* caused *Y*

would also appear in connection with statements of the form

(C2) *X V*-ed *Y*

—where '*V*' is a causal verb other than the generic 'cause'. To explore this, we began with a well known case from the prior literature, then replicated and extended our findings using some cases of our own.

Experiment 1: The Pilfering Professor

Above we referenced Knobe and Fraser's (2008) *pen case*, which is one of the earliest demonstrations of the influence of norms on causal judgment. As we noted there, Knobe and Fraser's finding, confirmed in replications and extensions of the case (Hitchcock and

Knobe, 2009; Systma, Livengood and Rose, 2012), was that agreement with the statement that

(4) [Professor Smith / The administrative assistant] caused the problem.

is affected by whether or not the person described was *allowed* to take the receptionist's pens. Specifically, participants are more inclined to agree that someone caused the receptionist's problem if that person *violated a norm* by taking one of her pens.

It is arguable, however, that agreement with a judgment like (4) is less a way of describing what a person *did* or *brought about* than of *blaming* them for a bad outcome, or saying that they were *responsible for* this consequence of their disallowed action. As one of our experimental participants wrote to us following the conclusion of an experiment that used this case:

> An administrative assistant took a pen and a professor took a pen. […] I think the Professor Smith did help cause the problem since he wasn't supposed to take a pen (and apparently didn't return the one he used). But it also is at the fault of the Admin Assistant since they didn't return the pen they used.

Our participant gets something importantly right: if the professor and the administrative assistant both take pens from the receptionist's desk, then under *that* description what they do seems to be the same. Because of this, the statement that one or the other of them "caused the problem" seems in a way to reach beyond this description of the causal structure of the scenario, to a further judgment that applies a concept of blameworthiness or moral responsibility.

We should also notice, however, that the phrase 'took a pen' *itself* describes a way of causing something to happen. The linguist Beth Levin identifies this use of 'took' as belonging to the subclass of "Verbs of Possessional Deprivation" that she calls "*steal* verbs": as with verbs like *abduct*, *capture*, *confiscate*, *liberate*, *pilfer*, *snatch*, and so on, to say that someone *took* something is to "describe the removal of something from someone's possession" (Levin 1993, §10.5, pp. 128-129). It makes sense, then, to investigate whether the application of this causal description is likewise sensitive to whether the action it describes was a violation of a moral or statistical norm—i.e., whether the influence of norms on agreement with (4) extends as well to agreement with the simple statement that

(5) [Professor Smith / The administrative assistant] took a pen.

If it does not, then this will be our first piece of evidence that the effect of norms on causal judgments may be limited in the way we have hypothesized.

To do this, we began by replacing Knobe and Fraser's original vignette with the following:

> The receptionist in the philosophy department tries to keep her desk stocked with pens. [Administrative assistants / Faculty members] are allowed to use these pens, but [faculty members / administrative assistants] in the department are not allowed to use them. They are expected to buy their own.
>
> On Monday morning, Professor Smith needs a pen for grading exams, and decides to go to the receptionist's office to look for one. At the same time, an administrative assistant also decides to go to the receptionist's office to look for a pen. Later that day, the receptionist needs to sign a document ... but she has a problem. There are no pens left on her desk, and while she is trying to find one she misses an important call.

Here, the *permissible* version of the scenario is the one in which Professor Smith, whose action is described in both of the test statements presented below, was allowed to take the receptionist's pens, while the *impermissible* version of the scenario is the one in which he was not allowed to do this. Following the vignette each participant evaluated two comprehension statements and then indicated their agreement with the following test statements on a standard 7-point Likert scale:

Caused Professor Smith caused the problem.

Took Professor Smith took a pen from the receptionist's desk.

All statements were presented on a single page, with the comprehension statements coming first followed by the two test statements, which occurred in a randomized order. We predicted that, while we would replicate Knobe and Fraser's finding of a significant effect of the permissibility of Professor Smith's action on agreement with **Caused**, this effect would *not* be observed in connection with ratings of **Took**, even though this statement also involves a causal concept.

Three points should be emphasized about the structure of this very simple experiment. The first is that, unlike Knobe and Fraser's original, the text of our pen vignette does not say explicitly that either Professor Smith or the administrative assistant took one of the receptionist's pens from her desk, but only that each of them *went to look* for one in her office. The purpose of this was to ensure that judgments of who *took* the

pen, no less than judgments of who *caused* the problem, required participants to make an inference beyond the explicitly stated facts of the case. Our interest was in determining whether this causal inference would be affected in each case by the normative status of the agent's decision.

Our second point is concessive: it is clear that there are important differences between the structure of the causal process described in **Caused** and the structure of the causal process described in **Took**, including (i) that it is only in **Caused** that the agent is related explicitly to an independent event or state of affairs (here, the "problem" faced by the secretary) that might be viewed as the outcome of what they did, and (ii) that **Took**, unlike **Caused**, does not describe an outcome that results from the *joint* behavior of the two agents in the scenario. Together these points mean that, to the extent that the norm effect is supposed to arise just as a result of the role of normative considerations in making certain counterfactuals more salient than others (as in the account of Hitchcock and Knobe, 2009), it might therefore be expected not to extend to **Took**. However, we thought it important to treat this as an open question, since as we explain just below there was at least some prima facie reason to think that evaluations of **Took** might have been influenced by normative considerations as well. We will revisit this matter in more detail in introducing our third and fourth experiments, which were designed to address both of these concerns.

Finally, and related to our second point, it is important to the structure of this experiment that the event described in **Took**—Professor Smith's taking a pen from the receptionist's desk—*is itself* either a violation of a norm, or not a violation of a norm, depending on the version of the vignette in question. This is what made it prima facie possible that our participants' agreement with **Took** would have been subject to normative influence: for while there is *a* sense in which the professor will have done the same thing whether or not he was allowed to do what he did, it is possible that the tendency to describe him as *taking* a pen will be affected by whether doing so was a violation of a norm. As we shall see, however, there was in fact no such effect.

One hundred twenty-six people (71% male, Average Age=31.6 years) participated in this experiment, of whom fifty-five were excluded for failing a comprehension check. This experiment, like all the other experiments in this study, was administered on Qualtrics, and all participants were recruited from Amazon's Mechanical Turk platform. A 2x2 mixed ANOVA with Statement (**Caused**, **Took**) as a within-subjects factor and Condition (Permissible, Impermissible) as a between-subjects factor revealed a significant main effect of Statement, $F(1,68)=30.66$, $p<.001$, a significant within-subjects interaction between Statement and Norm, $F(1,68)=9.19$, $p=.003$, and a main between-subjects effect of Condition, $F(1,68)=10.772$, $p=.002$. Subsequent analyses revealed a significant effect of condition on ratings of **Caused**, $F(1,68)=14.21$, $p=<.001$, $d=.90$, as participants agreed more that Professor Smith had caused the problem in the

impermissible condition, where he was not allowed to take the pen (*M*=5.32, *SD*=1.51), than in the *permissible* condition, where he was allowed to take it (*M*=3.70, *SD*=2.08). However, there was no such effect of whether the professor's behavior was permissible or impermissible on ratings of **Took**, $F(1,68)=.31$, p=.58, d=.13, as participants agreed equally that Professor Smith had taken a pen whether he was allowed to take one (*M*=5.97, *SD*=1.12) or not (*M*=5.82, *SD*=1.14). These findings are illustrated in Figure 1.

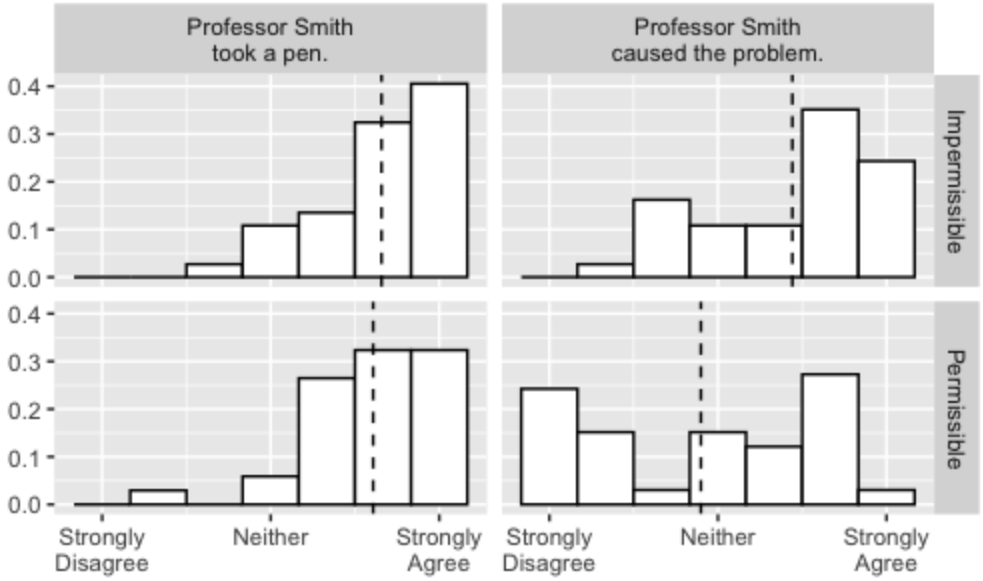

Figure 1: Results of Experiment 1. Bars indicate the proportion of participants who chose each response option in the corresponding condition, and dashed lines indicate mean response.

This experiment replicated the previously observed effect of normative considerations on agreement with **Caused** statements for our modified version of the Pen case. People agreed more that Professor Smith had *caused the problem* when he was not allowed to take pens from the receptionist's desk than when he was allowed to take them. However, we found that this effect of normative considerations on causal judgments was limited: it did not extend to ratings of **Took**, a statement that describes a causal process using a "special causal verb" in the form of (C2). Our subsequent experiments sought to extend this pattern in connection with a number of different ways of expressing causal concepts.

Experiment 2: Mrs. Smith's Missing Keys

Our second experiment sought to replicate the findings from Experiment 1 using a novel vignette and a different pair of causal verbs. Of particular interest was the potential influence of norms on periphrastic causative constructions with the following form:

(C3) *X* made *Y V*

—where *X* and *Y* are agents, and *V* describes an action that *Y* performs due to *X*'s influence. Part of what makes the construction in (C3) interesting is that, as noted above (see Table 1), 'make' occurs much more frequently than 'cause' in everyday speech and writing.[5]

Three hundred and fourteen individuals (49% male, Average Age=37 years) participated in this experiment. Ninety-eight were excluded for failing a comprehension check. Participants began by reading one of two versions of the following vignette:

> On Tuesday evening, Mr. Smith borrowed his wife's car for a trip to the grocery store. The next day, Mrs. Smith couldn't find her keys when it was time to go to work. After 10 minutes of looking, she found them in the dresser drawer –
>
> *Permissible variant:* even though she had always told Mr. Smith not to leave things there.
> *Impermissible variant:* just where Mr. Smith had told her he was going to leave them.
>
> Mrs. Smith ended up late to an important appointment.

After reading the vignette, participants rated three comprehension statements and then rated their agreement with one of the following test statements on a standard 7-point Likert scale:

Cause Mr. Smith caused Mrs. Smith to be late to her appointment.

Made Mr. Smith made Mrs. Smith late to her appointment.

Put Mr. Smith put the keys in the dresser drawer.

[5] We should note that 'make' is used most often as a lexical causative, in sentences like "We made a mistake" and "Sally made dinner". However, corpus data from iWeb (http://www.english-corpora.org/iweb/, retrieved 17. July, 2019) show that the construction 'made *X V*' is still used over three times more frequently than 'caused *X* to *V*' (32,962 occurrences vs. 9,272).

In contrast with our first experiment, Experiment 2 used an exclusively between-subjects design: the three test questions were randomized so that each participant only saw one of them. Otherwise, the three points that we emphasized in connection with Experiment 1 remain relevant here: while **Caused** and **Made** both relate Mr. Smith's action counterfactually to his wife's eventual lateness, the description of his action supplied by **Put** does not, and moreover this last description does not concern an outcome that results *jointly* from the behavior of two agents. However, the statement in **Put** does concern an action that either is (in the permissible variant) or is not (in the impermissible variant) a violation of a norm, under the description in question. Finally, as with our first experiment, neither version of our vignette says explicitly where Mr. Smith put the keys, which means that participants were required to make an inference from the information supplied to them in order to evaluate the description of his action in **Put**.

In line with prior work, we hypothesized that participants would agree more strongly with **Caused** when they read the *impermissible* version of the vignette, in which the keys were found where Mr. Smith was not supposed to put them, than when they read the *permissible* version, where the keys were found where he said. We also predicted that, similar to Experiment 1, agreement with **Put** would be unaffected by the difference between the two vignettes. Finally, we expected the pattern of agreement with **Made** to parallel the agreement with **Caused** rather than **Put**, since to say that Mr. Smith *made his wife late* is thereby to blame him, or hold him responsible, for this outcome of what he did. This prediction reflects the intuition that, as with **Caused**, the most natural reading of **Made** in this context is as a statement that assigns responsibility for a bad outcome.

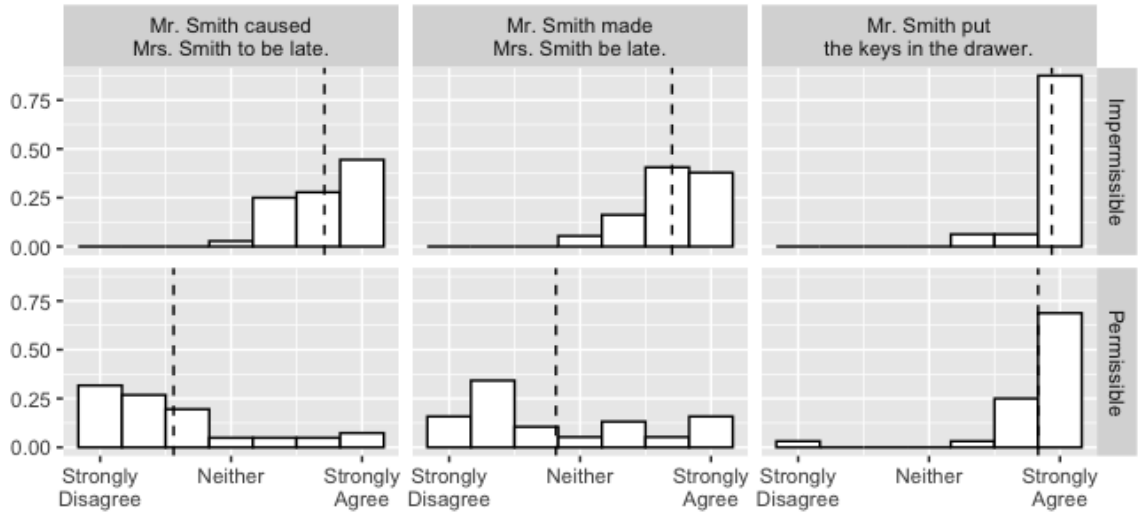

Figure 2: Results of Experiment 2. Bars indicate the proportion of participants who chose each response option in the corresponding condition, and dashed lines indicate mean response.

Once again, the results of this experiment were in line with our predictions. A 2x3 ANOVA revealed main effects of Condition (Permissible, Impermissible), $F(1,210)=141.09$, $p<.001$, and Statement (**Caused**, **Made**, **Put**), $F(2,210)=55.06$, $p<.001$, as well as a significant interaction between Statement and Condition, $F(2,210)=23.62$, $p<.001$. Subsequent analyses showed that the effect of Condition was significant only on ratings of **Caused**, $F(1,76)=10.85$, $p<.001$, $d=2.43$, and **Made**, $F(1,73)=50.21$, $p<.001$, $d=1.65$, and not on ratings of **Put**, $F(1,61)=.99$, $p=.323$, $d=.24$. That is, participants agreed more that Mr. Smith had *caused* Mrs. Smith to be late when the keys were found in an impermissible location ($M=6.14$, $SD=.90$) rather than a permissible one ($M=2.64$, $SD=1.83$), and likewise they agreed more that Mr. Smith had *made* Mrs. Smith be late when the keys were found in an impermissible location ($M=6.11$, $SD=.88$) rather than a permissible one ($M=3.44$, $SD=2.11$), but they agreed equally that Mr. Smith had *put* the keys in the dresser drawer whether doing so was permissible ($M=6.68$, $SD=.54$) or impermissible ($M=6.81$, $SD=.54$). These results are illustrated in Figure 3.

These findings indicate that agreement with sentences of the form '*X* made *Y V*' can be influenced by normative considerations in the same way as agreement with those of the form '*X* caused *Y*'. However, we also found that judgments of whether Mr. Smith *put* something somewhere were not influenced by normative considerations in this way. Moreover, unlike Experiment 1 the participants in this experiment saw only one of the three test questions, and so the difference between the pattern of ratings for **Caused** and **Made** compared to **Put** cannot have been due to their making implicit comparisons between them—that is, it is not as if participants could have been treating **Caused** as the "true" causal judgment, so that the norm effect was swallowed up by it. Indeed, we observed the same effect of norm on ratings of **Made**, but not on ratings of **Put**, even as each participant was presented with only one of these statements.

Experiment 3: Alex's Double Dose

We have suggested that our participants' greater willingness to say that Professor Smith *caused* the receptionist's problem, and that Mr. Smith *caused* or *made* Mrs. Smith (to be) late to her appointment, is due to the fact that these uses of 'caused' and 'made' have the function of *blaming* a person for doing something, or assigning responsibility for a negative outcome. However, another possible explanation of the difference between these findings and the pattern of results we observed in connection with the verbs 'take' and 'put' is that the latter verbs are both *lexical* causatives, while 'cause' and (as we used it in Experiment 2) 'make' both appear in *periphrastic* causal constructions. Our third experiment addressed this possibility by exploring the potential effects of normative

considerations on agreement with statements containing a lexical causative that *is* used frequently to express blame or assign responsibility, namely the verb *kill*.

One hundred and thirty-eight participants (71% male, Average Age=33 years) participated in this experiment. Thirty-four were excluded for failing a comprehension check. Each participant began by reading the following vignette, which draws on some elements of a similar vignette presented in Phillips and Kominsky (2016):

> Alex and Benni are conducting a science experiment on a mouse.
>
> [Alex/Benni] is supposed to inject the mouse with a solution containing 10 milliliters of a poisonous chemical compound called AX 300. AX 300 is poisonous to mice in 10 milliliter doses, though it is lethal only in doses of 20 milliliters or more.
>
> Meanwhile, [Benni/Alex] is supposed to inject the mouse with a solution containing 10 milliliters of VT 4000, a chemical compound that will counter the effects of AX 300.
>
> But [Benni/Alex] secretly decides to load her syringe with a 10 milliliter dose of AX 300, instead. The double dose proves to be lethal, and the mouse dies immediately.

Following the vignette participants indicated their agreement with each of three test statements regarding Alex's actions, presented in random order, on a standard 7-point Likert scale:

Caused Alex caused the mouse to die.

Killed Alex killed the mouse.

Injected Alex injected the mouse with AX 300.

Here, the *impermissible* condition is that in which Alex, who is described in all three statements, is the one who violates a norm by loading her syringe with AX 300 rather than the antidote, while the *permissible* condition is that in which Alex is supposed to inject the mouse with AX 300 and it is Benni who violates the norm instead. We predicted that, since in this context saying that Alex *killed* the mouse is a way of blaming Alex or attributing responsibility to her for the mouse's death, participants would agree more with both **Killed** and **Caused** when Alex's actions were impermissible than when

Alex does what she is supposed to do. Further, we predicted that we would not observe this pattern for agreement with **Injected**, since this description of what happened is not so normatively laden.

As with our first and second experiments, the description of Alex's action in **Injected** differed from the description in **Caused** in that it does not relate her action explicitly to a further outcome that depends counterfactually on it, nor does it describe something that was a joint effect of her action and Benni's. However, the description given in **Killed** was arguably different in this regard, given that 'kill' plausibly contains the sense 'cause to die' (though for skepticism concerning this analysis, see Fodor 1970 and Wierzbicka 1975). Moreover, as with the statements **Took** (in Experiment 1) and **Put** (in Experiment 2), the judgment that Alex *injected* the mouse with poison is itself a description of her as doing something that either did or did not violate a norm. Finally, as with our first two experiments, the vignette used in Experiment 3 does not say explicitly whether Alex had injected the mouse with AX 300, killed the mouse, or caused it to die, and so participants needed to make an inference from the information supplied in order to evaluate each of the three statements.

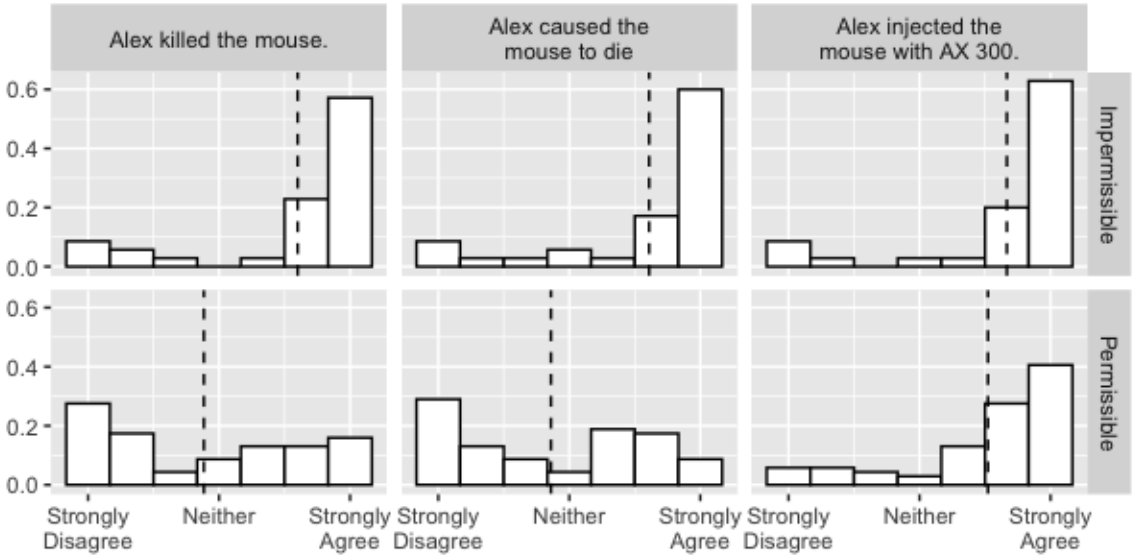

Figure 3: Results of Experiment 3. Bars indicate the proportion of participants who chose each response option in the corresponding condition, and dashed lines indicate mean response.

The results of Experiment 3 confirmed all three of our predictions. A 2x3 ANOVA revealed main effects of Condition, $F(1,102)=19.84$, $p<.001$, and Statement, $F(2,204)=36.24$, $p<.001$, as well as a significant interaction between Statement and Condition, $F(2,204)=13.72$, $p<.001$. Subsequent analyses showed Condition had a highly significant effect on ratings of **Caused**, $F(1,102)=26.4$, $p<.001$, $d=1.06$, and **Killed**, $F(1,102)=22.15$, $p<.001$, $d=.98$, while there was no significant effect of Condition on

ratings of **Injected**, $F(1,102)=1.313$, $p=.254$, $d=.24$. That is, participants agreed more that Alex had *caused* the mouse to die when she acted impermissibly ($M=5.83$, $SD=1.95$) than when she acted permissibly ($M=3.58$, $SD=2.19$) in loading her syringe with AX 300, and likewise they agreed more that Alex had *killed* the mouse when she was not supposed to inject the mouse with AX 300 ($M=5.80$, $SD=2.00$) than when she was supposed to do this ($M=3.65$, $SD=2.29$), but they agreed about equally that Alex had *injected* the mouse with poison whether doing so was permissible ($M=6.00$, $SD=1.86$) or impermissible ($M=5.56$, $SD=1.81$). These results are illustrated in Figure 3.

These findings challenge the possible interpretation of our previous findings as resting just on the difference between lexical and periphrastic causative constructions, and provide further support for our contention that the tendency to identify a norm-violating agent as the one who *caused* something bad to happen or *made* someone do an undesired thing reflects the tendency to blame that person for what they did, or hold them morally responsible for it. Whether Alex violated a norm in loading her syringe with AX 300 had a significant effect on participants' willingness to say that she *killed* the mouse and she *caused* it to die, but it did not have such an effect on their judgment of whether she *injected the mouse* with AX 300—since this description of what Alex did applied equally whether or not she violated a norm in doing it.

Experiment 4: The Shape Playground

One barrier to using vignettes as stimuli in eliciting judgments of actual causation is that so much of our ordinary language uses causal verbs simply in the description of what happens. As we have noted, the vignettes that we used in Experiments 1-3 all get around this difficulty by avoiding any explicit description of the causal processes we were interested in having participants form judgments about: thus our version of the pen vignette differs from Knobe and Fraser's (2008) original in that nowhere in our vignette does it say that Professor Smith took a pen, and the vignettes used in Experiments 2 and 3 also left implicit what Mr. Smith had done with the keys and Alex had done to the mouse.

However, another way of inviting participants to draw inferences about causal processes from stimuli that do not represent those processes explicitly is by using animations, rather than text-based stimuli, as experimental prompts. A large body of research has shown that participants readily infer (or even perceive: cf. Anscombe 1981, pp. 136-138; Siegel 2009; Helton 2018) causal interactions and agential behavior on the basis of 2D stimuli consisting in nothing more than the movements of simple shapes (for reviews, see Scholl and Tremoulet, 2000; Danks, 2009). Paradigms that employ such materials provide a way of exploring causal judgments that avoids the potential pitfall described just above, since in them the causal structure of the scenario under

consideration is not *described* to participants at all, and they must instead *infer* its causal structure on the basis of what they see.[6]

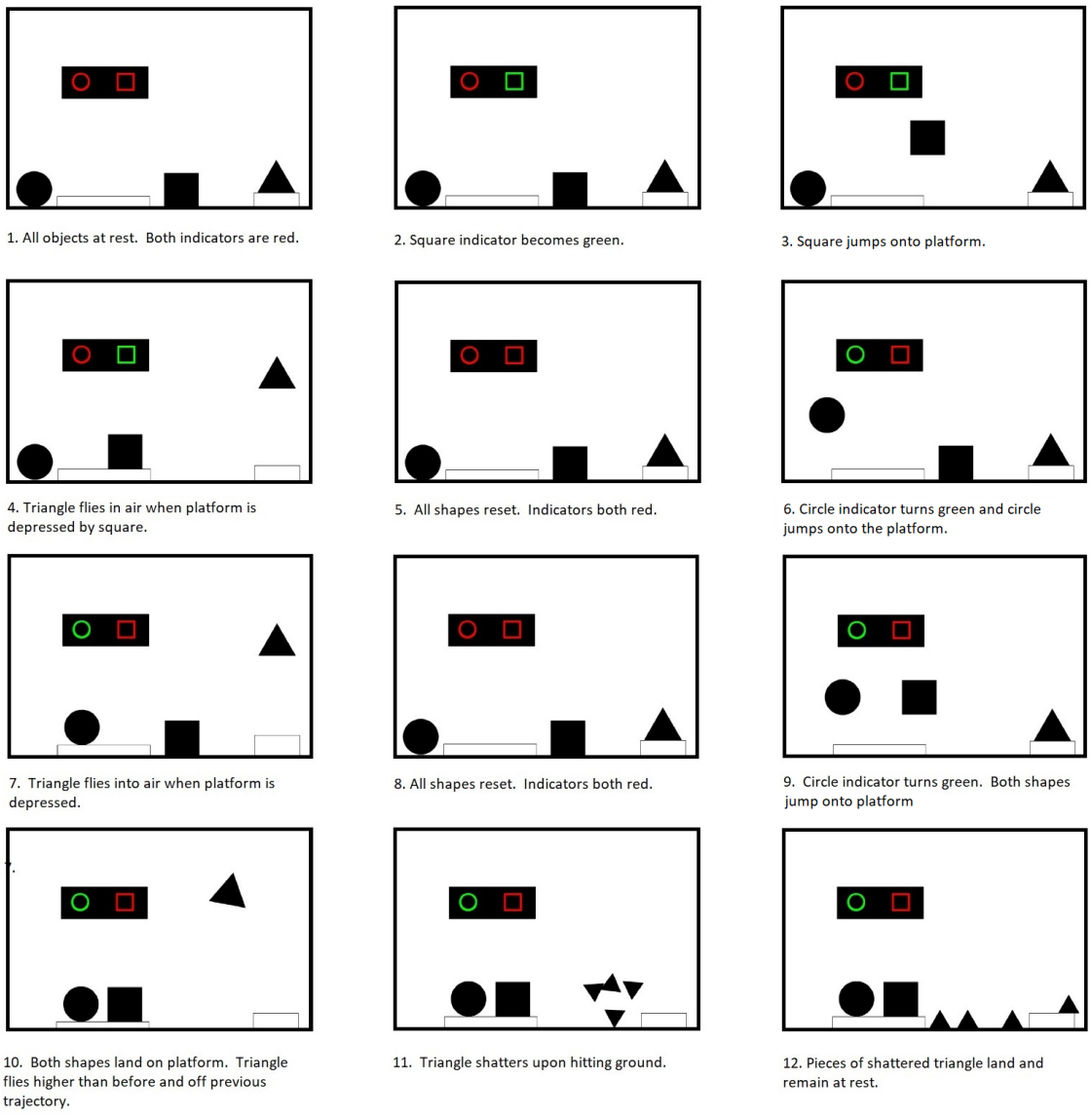

Figure 4: Illustration of the animation used in Experiment 4. The pattern in slides 2-8 was repeated three times before the final event shown in slides 9-12 began. Shown here is the *impermissible* condition, in which only the circle was "supposed" to jump onto the platform at the end, as indicated by the fact that only its indicator light turns green. By contrast, in the *permissible* condition the indicator light for both shapes turned green at the point shown in slide 9, and remained so until the end of the animation.

[6] For other recent work that utilizes this approach, see Wolff (2003; 2007) and Wolff and Thorstad (2016).

Our final experiment attempted to extend and replicate our earlier findings in a way that builds on this research, by displaying a simple animation and inviting participants to make causal judgments about it. Three hundred and nineteen participants (61% male, Average Age=34 years), of whom one hundred and fifty-six were excluded for failing a comprehension check, watched one of two 30-second animations that depicted the motion of several geometric shapes. These animations are illustrated in Figure 4, and can be viewed in full at http://www.schwenkler.org/cause-norm.html. At the beginning of each animation (Figure 4, Slide 1) a circle and a square were located next to a rectangular box, above which was a rectangular shape containing outlined images of a square and a circle. Each time one of the outlined images changed from red to green, the corresponding shape rose up and landed on the box—according to the pattern "circle, square, circle, square, …" (Figure 4, Slides 2-3 and 6). Each time a shape jumped on the box, a triangle that was on a different box rose up into the air as the box rose up underneath it (Figure 4, Slides 4 and 7). Finally, after three rounds of taking turns jumping on the platform, the circle and the square both jumped on the platform at once, and the triangle rose higher into the air than usual and then fell to the ground, shattering into pieces.

These events proceeded according to one of two conditions. In the *impermissible condition*, as illustrated in Slides 9-12 of Figure 4, before the circle and the square jumped simultaneously onto the box the signal above it illuminated only the circular shape, indicating that the square had jumped out of turn. By contrast, in the *permissible condition* the signal illuminated both shapes at once, indicating that both of them were allowed to jump.[7] Following a comprehension question designed to ensure that they had picked up on the rule that the jumping shapes were supposed to follow, each participant responded to the following three test statements, presented in random order, by choosing which of the phrases 'square', 'circle', or 'circle and the square' best completed each one:

Broke At the end, the _____ broke the triangle.

Bounced At the end, the _____ bounced the triangle into the air.

Caused At the end, the _____ caused the triangle to break.

We predicted that, in line with the findings of our earlier experiments, participants would tend to choose 'square' to complete both **Caused** and **Broke** more frequently in the permissible condition than the impermissible condition, and that we would not observe this pattern in connection with **Bounced**. (Recall that it is the square that jumps "out of

[7] We thank Samantha Berthelette for giving us the idea to convey the normative structure of the scenario in this way.

turn" in the impermissible condition, but follows the direction of the signal in the permissible one.) And our analyses confirmed these predictions, as planned comparisons found a significant effect of Condition on the percentage of participants who chose 'square' to complete **Caused** (t=6.42, p<.001) and **Broke** (t=5.47, p<.001), but not on the percentage of participants who chose "square" to complete **Bounced** (t=1.48, p=.140). That is, participants judged consistently that *both* shapes had bounced and broken the triangle, and caused the triangle to break, in the permissible condition, when both shapes obeyed the signal in jumping together onto the platform. By contrast, in the impermissible condition, in which the square's final jump was out of turn, a similar proportion of participants judged that both shapes had *bounced* the triangle into the air, while significantly larger proportions judged that the square had *broken* the triangle and *caused* it to break. The full results of this experiment are presented in Table 2 and illustrated in Figure 5.

Statement	*Condition*	*"circle"*	*"square"*	*"circle and the square"*
Broke	Impermissible	2.78%	31.94%	65.28%
	Permissible	1.10%	1.10%	97.80%
Bounced	Impermissible	5.56%	15.28%	79.17%
	Permissible	0%	7.69%	92.31%
Caused	Impermissible	4.17%	38.89%	56.94%
	Permissible	0%	1.10%	98.90%

Table 2: Results of Experiment 4. Percentages indicate the proportion of participants who chose each response option.

The structure of this experiment differed in an important way from the structure of our first three, and this difference helps to address the concessive point that we raised in our discussion of Experiment 1. As with those experiments, the description of either shape as having *bounced* the triangle into the air required participants to make an inference that went beyond what was immediately depicted in the experiment, and concerned an action that either was or was not a violation of a norm, under the description given. However, unlike the description of Professor Smith as having *taken* a pen, of Mr. Smith as having *put* the keys in the dresser drawer, and of Alex as having *injected* the mouse with AX 300, the description supplied in **Bounced** is easily understood as describing an event (the triangle's going into the air) that can be construed as depending counterfactually on the activity of two separate agents (the circle and the

square). Because of this, the lack of a significant effect of norm on ratings of this statement was especially revealing: it suggests that our earlier findings were not just due to the distinctive causal structure of the events described in **Take** (in Experiment 1), **Put** (in Experiment 2) and **Injected** (in Experiment 3), since the event described in **Bounced** had just the causal structure described in those judgments that have been found in the past, and in the other experiments in the present study, to be affected by moral norms. Because of this, we take the insensitivity of these statements to normative considerations as evidence that participants were treating them as doing a quite different thing from the judgments expressed in our several **Caused** statements and in **Made**, **Killed**, and **Broke**: rather than expressing a judgment of moral responsibility, their role was to describe the causal structure of the scenario that participants had just been shown.

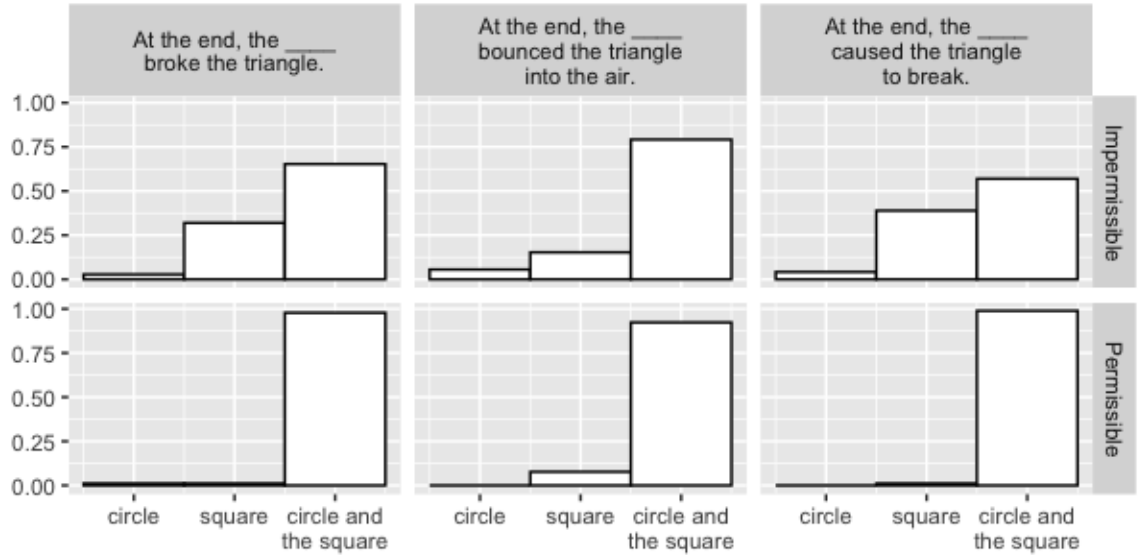

Figure 5: Results of Experiment 4. Bars indicate the proportion of participants who chose each response option in the corresponding condition.

These findings replicate and extend the results of our first three experiments, while also illustrating the value of causal perception paradigms for the philosophical study of causal judgments. When the square's final jump was permissible, response patterns were similar across **Caused**, **Broke**, and **Bounced**, as participants tended to identify both the square and the circle as having done the thing in question. By contrast, in the norm-violating condition participants selected "square" at similar rates for **Caused** and **Broke** but were less likely to do so for **Bounced**. Indeed, participants selected "square" at similar rates to complete **Bounced** regardless of whether a norm was violated. This contrasts with their judgments of which of the shapes *broke* the triangle and *caused* the triangle to break.

3. Discussion

What is the structure of ordinary causal thinking? Following Knobe (2010), we may distinguish accounts of human causal cognition that depict it as broadly *scientific* in character from those that depict it as *moralized* through and through. According to accounts of the first kind, human beings aim to make sense of the world in roughly the same way that, at least in the ideal, scientists reach conclusions about their domains of expertise: they draw on "observation, experimentation, [and] the construction of explanatory theories" (Knobe, 2010, p. 315) in filling out their understanding of how the world works. And Knobe finds this account to be undermined by evidence showing the influence of moral norms on causal judgments: the process of causal cognition, he argues, is "suffused through and through with moral considerations" (ibid., p. 317) in a way that reveals that this process does not even *aim* at developing the kind of objective, purely evidence-based model of causal structure that these accounts imagine.

The evidence that we have presented in this paper gives reason to think that this very strong conclusion is overstated. The extensive influence of normative considerations on what we have called *'cause'-judgments*, or judgments that are ordinarily expressed with the use of the English verb 'cause', is not an influence on *causal judgment in general*—i.e., on the much broader family of judgments that are expressed with the use of special causal verbs like 'take', 'put', 'inject', 'bounce', and so on. Many of these latter judgments, we have shown, are totally unaffected by whether the action they describe was a violation of a moral norm—and to the extent that they are so affected, it seems to be because they involve causal concepts, such as the one expressed by the use of 'kill' in our third experiment and 'break' in our fourth one, that possess their own moral valence.

How should we understand these findings, and what is the model of causal cognition that they support? We contend that these findings can be explained by reflecting on two quite different causal tasks that were faced by the participants in our four experiments. One of these tasks was to *describe* the causal structure of a scenario, to identify *what happened* in a way that brings it under various specific causal concepts expressed by words like those in Anscombe's list. Doing this requires collecting, from observation and also from inference, various facts relating to what happened, and putting these facts together in order to generate a description of what took place. And our chief finding was these descriptions tend to be unaffected by the normative valence of the events they concern—so that, for example, the question whether Professor Smith *took a pen* doesn't seem depend at all on whether doing so was a violation of a moral norm.

Along with this descriptive task, the other task our participants faced was that of forming a conclusion about how the facts of the case they considered related to a further question, namely that of which of the parties in the scenario was *responsible* for a

negative outcome of it. Across all of our experiments we found that it is when causal concepts are used in this latter way—to say, for instance, whether or not it was Professor Smith who *caused the problem* that the secretary eventually faced, or whether Mr. Smith *made his wife late* by doing what he undeniably did, viz., putting her keys in the dresser drawer, or whether Alex *killed the mouse* when she injected it with a poisonous substance—that the judgments they figure in are subject to normative influence. It is when words like 'cause', 'make', and 'kill' are used to express what we may call *responsibility concepts*, or concepts of who is responsible and therefore blameworthy for a certain negative outcome, that the use of these words is affected by moral considerations.

The difference between the two tasks that were faced by our participants can be analogized to the difference between two tasks facing judges and jurors in a court of law: on the one hand, that of establishing what are the facts that are relevant to determining the responsibility of a given party for a certain outcome; and on the other, that of reaching this further determination of guilt or innocence.[8] H. L. A. Hart and Tony Honoré appeal to this distinction in their visionary work *Causation in the Law*, where they outline what they call an "apparent paradox" that turns on how we understand the relationship between causation and responsibility:

> Much modern thought on causation in the law rests on the contention that the statement that someone has caused harm either means no more than that the harm would not have happened without ('but for') his action or where (as in normal legal usage and in all ordinary speech) it apparently means more than this, it is a disguised way of asserting the 'normative' judgment that he is responsible, i.e. that it is proper or just to blame or punish him or make him pay. On this view to say that a person caused harm is not really, though ostensibly it is, to give a *ground* or *reason* for holding him responsible in the first [that is, 'but-for'] sense; for we are only in a position to say that he has caused harm when we have decided that he is responsible. (Hart and Honoré, 1959, p. 66)

[8] Indeed, it is precisely in order to separate these two causal notions that The Restatements (Second) of Torts draws a distinction between what it calls "philosophic cause" and "substantial cause" (Restatement, Third, of Torts: § 431 cmt. a). The former term implies an extended counterfactual causal concept in which background conditions and events play causal roles in bringing about an event, while the latter term uses cause "in the popular sense, in which there always lurks the idea of responsibility" (ibid.). And the most recent treatment of causation in The Restatements (Third) of Torts preserves this traditional distinction with updated language by distinguishing "factual cause" from "harm within the scope of liability (which has historically been called 'proximate cause')" (Restatement, Third, Torts: Liability for Physical and Emotional Harm § 6 cmt. b at 67-68). In this case, the latter is used to distinguish the causal agents which the court may hold liable from those which are too attenuated to fall "within the scope of liability". In other words, what has traditionally been called proximate cause is a concept used to assign responsibility.

Hart and Honoré's two senses of 'cause' correspond to the two kinds of causal judgment that were elicited from participants in our experiments: their concept of "but for" causation, or what in the Restatements (see fn. 8 above) is called "philosophic" or "factual" causation, is the concept involved in what above we called *describing* the causal structure of a scenario, while their "normative" causal concept is, as they suggest, the one involved in identifying the party that was morally responsible for an outcome—i.e., the concept that in the Restatements is described as that of "substantial cause" or "harm within the scope of liability". Our argument here has turned on the simple observation that many of our ordinary causal judgments fall into the first of these categories, and that these descriptions are largely presupposed by 'cause'-judgments that express a determination of moral responsibility. And the evidence we have presented here strongly suggests that ordinary causal judgment is influenced by normative considerations in the way proposed by Knobe only to the extent that it expresses a conclusion of this latter sort.

While these parallel notions of causation are conceptually distinct, people don't always keep them separate in practice. It is for this reason that contemporary law establishes procedural safeguards to prevent jurors from conflating these two lines of inquiry. Such safeguards aim to ensure that complex legal judgments mirror the way in which our common sense causal reasoning unfolds. For example, the relevant portions of Rule 701 of the Federal Rules of Evidence limit witness testimony to statements which are "rationally based on the witness's perception" and "helpful to clearly understanding the witness's testimony or to determining a fact in issue" (Fed. R. Evid. 701). In other words, a witness cannot be asked to form a conclusion, but can only assist the court in establishing facts relevant to the case. For this reason, judges and jury are often referred to as the "fact finder" during this stage of a trial. Forming a conclusion is the duty of the jury at the end of the trial, *after* the facts "in issue" have been established.

Judgments of moral or legal responsibility, including those implying the "normative" or "substantial" concept of causation that the ordinary use of 'cause' most often expresses, are *conclusions* in the sense just defined. And we cannot form conclusions until we get our facts straight. In this way, causal action-descriptions are structurally prior, and conceptually quite different, from causal conclusions. Hart and Honoré note this difference when they point out the distinction between simple causal action statements such as 'He pushed it' and 'He broke it' with what they call "the most obvious and fundamental case of all for the attribution of responsibility: the case in which we can simply say 'He did it'" (1959, p. 73). Our data, however, strongly suggest that it is not necessarily anything inherent in the content of verbs like 'push', 'break', and 'do' that accounts for this distinction. As our fourth experiment reveals, the claim that so-and-so *broke* such-and-such can, in the right context, function to express a judgment of moral responsibility that is affected by normative considerations in just the same way as 'So-and-so did it' or 'So-and-so caused such-and-such to break'. And our second and

third experiments showed that something similar can happen with the ordinary use of 'make' and 'kill': our participants' use of these words was equivalent to their use of 'cause', and gave them a way of saying something quite different than what they said using verbs like 'put' and 'inject'. It is a task for further research to explore the factors that influence whether the use of various verbs is liable to be treated as doing one or the other of these things—whether describing facts about the causal relationships between various entities, events, and states of affairs, or conveying normatively valenced conclusions that assign responsibility for what happened.

References

Alicke, M. D. (1992). Culpable causation. *Journal of Personality and Social Psychology* 63(3), 368.

Alicke, M. D., Rose, D., & Bloom, D. (2011). Causation, norm violation, and culpable control. *The Journal of Philosophy* 108(12), 670-696.

American Law Institute. (1965). Restatement of the Law Second, Torts. St. Paul: American Law Institute Publishers.

American Law Institute. (2010). Restatement of the Law Third, Torts, Liability for Physical and Emotional Harm. St. Paul: American Law Institute Publishers.

Bar-Asher Siegal, E., Bassel, N. and Hagmayer, Y. (Forthcoming). Causal selection -- the linguistic take. *Experiments in Linguistic Meaning* 1.

Bar-Asher Siegal, E., Bassel, N. and Hagmayer, Y. (In prep. B). The role of agency in causal selection.

Danks, D. (2009). The psychology of causal perception and reasoning. In Helen Beebee, Christopher Hitchcock, and Peter Menzies (Eds.), *The Oxford Handbook of Causation* (447-470). New York: Oxford University Press.

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

Fodor, J. (1970). Three reasons for not deriving "kill" from "cause to die". *Linguistic Inquiry* 1(4), 429-438.

Godfrey-Smith, P. (2009). Causal pluralism. In Helen Beebee, Christopher Hitchcock, and Peter Menzies (Eds.), *The Oxford Handbook of Causation* (326-337). Oxford University Press.

Hall, N. (2004). Two concepts of causation. In J. Collins, N. Hall and L. A. Paul (Eds.), *Causation and Counterfactuals* (225-276). Cambridge: The MIT Press.

Hart, H. L. A. & Honoré, T. (1959). *Causation in the Law*. Oxford: Oxford University Press.

Helton, G. (2018). Visually perceiving the intentions of others. *Philosophical Quarterly* 68(271), 243-264.

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

Kominsky, J. F. and Phillips, J. (2019). Immoral Professors and Malfunctioning Tools: Counterfactual Relevance Accounts Explain the Effect of Norm Violations on Causal Selection. *Cognitive Science* 43: e12792. doi:10.1111/cogs.12792

Knobe, J., & Fraser, B. (2008). *Causal judgment and moral judgment: Two experiments.* In W. Sinnott-Armstrong (Ed.), *Moral psychology, Vol. 2. The cognitive science of morality: Intuition and diversity* (p. 441–447). Cambridge: The MIT Press.

Knobe, J., Prasada, S., & Newman, G. E. (2013). Dual character concepts and the normative dimension of conceptual representation. *Cognition* 127(2), 242-257.

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. Chicago: University of Chicago Press.

Livengood, J. M., Sytsma, J., & Rose, D. (2017). Following the FAD: Folk Attributions and Theories of Actual Causation. *Review of Philosophy and Psychology* 8(2), 273-294. https://doi.org/10.1007/s13164-016-0316-1.

Phillips, J. S., & Kominsky, J. F. (2016). Causation and norms of proper functioning: Counterfactuals are (still) relevant. https://doi.org/10.31234/osf.io/dwajn

Kominsky, J. F. and Phillips, J. (2019), Immoral Professors and Malfunctioning Tools: Counterfactual Relevance Accounts Explain the Effect of Norm Violations on Causal Selection. *Cognitive Science* 43: e12792. doi:10.1111/cogs.12792

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Science* 4(8), 299-309.

Schwenkler, J. (In prep.). The varieties of causation.

Siegel, S. (2009). The visual experience of causation. *Philosophical Quarterly* 59(236), 519-540.

Sytsma, J., Livengood, J., & Rose, D. (2012). Two types of typicality: Rethinking the role of statistical typicality in ordinary causal attributions. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 43(4), 814-820.

Sytsma, J., Bluhm, R., Willemsen, P., & Reuter, K. (2019). Causal Attributions and Corpus Analysis. In Eugen Fischer & Mark Curtis (ed.), *Methodological Advances in Experimental Philosophy*. Bloomsbury Press.

Walsh, C. R., & Sloman, S. A. (2011). The meaning of cause and prevent: The role of causal mechanism. *Mind & Language* 26(1), 21-52.

Wierzbicka, A. (1975). Why "kill" does not mean "cause to die": The semantics of action sentences. *Foundations of Language* 13(4), 491-528.

Wolff, P. (2003). Direct causation in the linguistic encoding and individuation of causal events. *Cognition* 88, 1-48.

Wolff, P. (2007). Representing causation. *Journal of Experimental Psychology: General* 136(1), 82-111.

Wolff, P., & Thorstad, R. (2016). Force dynamics. In M. Waldmann (ed.), *The Oxford Handbook of Causal Reasoning* (147-167). New York: Oxford University Press.