Cognition 264 (2025) 106225

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Full Length Article

No privileged link between intentionality and causation: Generalizable effects of agency in language☆

Sehrang Joo [a], Sami R. Yousif [b], Fabienne Martin [c], Frank C. Keil [d], Joshua Knobe [e,*]

[a] Department of Psychology, Princeton University, United States of America
[b] Department of Psychology and Neuroscience, University of North Carolina, Chapel Hill, United States of America
[c] Institute for Language Sciences, Utrecht University, the Netherlands
[d] Department of Psychology, Yale University, United States of America
[e] Program in Cognitive Science and Department of Philosophy, Yale University, United States of America

ARTICLE INFO

Keywords:
Agency
Intentionality
Causation
Syntax
Semantics

ABSTRACT

People are more inclined to agree with certain causal statements when a person acts intentionally than when a person acts unintentionally or without agency. Most existing research has assumed that this effect is to be explained in terms of the operation of people's causal cognition. We propose a different explanation which involves a linguistic phenomenon involving the impact of agency on people's judgments about a broader class of sentences, including non-causal sentences. Study 1 shows that the effect arises for both causal and non-causal sentences. The remaining studies show that the effect arises only when the subject of the sentence is animate (Study 2), that the effect arises both for outcomes with negative valence and outcomes with neutral valence (Study 3) and that the effect is driven by whether or not a person exercises agentive control over her body, rather than whether or not she intends the particular outcome of her action (Study 4). We conclude with a formal linguistic theory that captures these effects.

Imagine a train platform with a line that people aren't supposed to cross. Tom deliberately steps over the line, and this ends up causing a train delay. In this case, it seems natural to say:

(1) Tom caused the train delay.

Now consider a slightly different case: Instead of intentionally crossing the line on the train platform, Tom blacks out and falls over it. Just as in the first scenario, this ultimately leads to a train delay. In this second case, would people still think it seemed natural to use sentence (1)?

Existing research in causal cognition has uncovered a surprising fact about people's judgments in cases like this one: Participants are less inclined to agree that a person caused an outcome when a person brings about the outcome through a behavior that doesn't involve an exercise of agency (Lombrozo, 2010; Rose, 2017). In other words, in scenarios like this one, participants tend to be more inclined to say that Tom caused the train delay when he intentionally crosses the line than when he simply blacks out and falls over the line. This finding has spurred a much larger research program, with numerous studies showing that participants' perceptions of what is going on within an agent's mind can impact their judgments about whether it is right to say that the agent caused some further outcome (Kirfel & Lagnado, 2021a, 2021b; Lagnado & Channon, 2008; Lombrozo, 2010; Phillips & Shaw, 2015; Rose, 2017; Schwenkler & Sytsma, 2020).

While several different theories attempt to explain this effect of perceived agency, they share a key assumption; the effect should be understood in terms of how perceived agency influences people's thinking about *causation* in particular. On these views, there is a clear motivation for why one might be interested in understanding the effect of perceived agency—in order to better understand causal cognition.

But consider a sentence that does not involve any claim about a person causing a further outcome, i.e., a sentence that is simply about the behavior the agent performed. For example, returning to our story about Tom and the train, consider the sentence:

☆ Link to data and materials: https://osf.io/7a6fq/?view_only=08f4cf81aa4647a898ae47789fac81ca.

* Corresponding author.
E-mail addresses: sehrang.joo@princeton.edu (S. Joo), sami.yousif@unc.edu (S.R. Yousif), f.e.martin@uu.nl (F. Martin), frank.keil@yale.edu (F.C. Keil), joshua.knobe@yale.edu (J. Knobe).

https://doi.org/10.1016/j.cognition.2025.106225
Received 23 November 2024; Received in revised form 16 June 2025; Accepted 17 June 2025
Available online 12 July 2025

(2) Tom crossed the line.

In contrast with *cause*, the verb "cross" is not what is called a *causative verb* (Levin, 1993, 1999), and accordingly, this sentence does not assert that Tom's behavior caused any further outcome. But might a manipulation of Tom's level of agency also affect evaluations of sentences like (2)? If we find the same effect for judgments of (2) as previous research has found for (1), perceived agency would be having some effect that extends *beyond* causal cognition as it is encoded in causal statements.

In other words, the effect of perceived agency on causal judgments may reflect a more general way in which people understand and talk about animate agents. In everyday conversation, we talk about people acting in many different ways—not always with causal statements. Might these effects of perceived agency arise for people's understanding of this much larger set of statements? If so, then understanding how reasoning about agency figures into people's evaluations of sentences like (1) and (2) would be of interest not only to psychologists working on causal cognition, but also to those interested more broadly in understanding agency and its role in language.

1. Agency in causation

The impact of agency on people's judgments about causal sentences is surprising in part because it points to a factor that one might not have expected to have any influence. Looking at a sentence like "Tom caused the train delay", one might expect that people's judgments about this sentence would be affected only by their understanding of Tom's behavior and the connection between this behavior and the train delay that eventually occurred. Existing research shows that this is not the case. People's judgments are also affected by their understanding of what was going on within Tom himself and, in particular, by the degree to which his behavior was the result of his own agency (e.g., Kirfel & Lagnado, 2021a; Lagnado & Channon, 2008; Phillips & Shaw, 2015; Schwenkler & Sytsma, 2020).

In perhaps the first demonstration of this effect, Lombrozo (2010) showed that agency has an impact on people's use of causal sentences in cases of double prevention. Consider a case in which a person prevents an event that would have in turn prevented some further outcome (had it actually occurred). Do people judge in such cases that the person caused the outcome? Lombrozo's studies showed that the answer depends on whether the person acted through her own agency. Participants were more inclined to say that the agent caused the outcome when the person performed a behavior by exercising her own agency (e.g., throwing a ball) than when the person performed a behavior without exercising her agency (e.g., dropping a ball).

Subsequent research has extended this effect to other kinds of scenarios, with different causal structures. For example, Rose (2017) presented participants with cases in which there was no double prevention, but the results nonetheless indicated that participants were more inclined to say that the agent caused the outcome when she exercised her own agency (intentionally pressing a button) than when she did not (having a stroke that results in her hand involuntarily hitting the button).

Existing theoretical work has explained this effect in terms of an impact of teleology or goal-directness on causal judgments (Lombrozo, 2010). The core idea is that when an agent is specifically trying to bring about an outcome, the relationship between the agent's behavior and the outcome is not as sensitive to background conditions. Consider first the case in which Tom crosses over the line entirely as an accident, and the result is an unexpected train delay. In this first case, if the background conditions had been slightly different (e.g., if the train had come just a few minutes later), then the outcome would not have arisen. Now, by contrast, consider a case in which Tom specifically crosses over the line in order to ensure that the train is delayed. In this latter type of case, we would not see the same sensitivity to background conditions. If it happened that the train came a few minutes later, or if background conditions differed in some other minor way, Tom would simply adjust his behavior to make sure that the train was still delayed. Research on causal judgments consistently points to an impact of robustness across background conditions on intuitions about causation (Hitchcock, 2012; Icard et al., 2017), so if the effect is fundamentally a matter of something about how people's causal judgments work, it could arise because agency leads to greater perceived robustness, which in turn leads to greater attribution of causation.

2. Agency in language

Thus far, we have been looking at the impact of agency on people's causal judgments. A question now arises as to whether there is also an impact of agency on judgments regarding sentences that do not directly involve causation.

To address this question, we first need to consider the different ways in which a sentence can express a causal claim. One way for a sentence to express a causal claim is to explicitly use a term like "cause," but most sentences that express causal claims do not involve the use of such terms. To illustrate, compare (3) with (4).

(3) a. Jamie caused the table to break.
b. Harry caused the butter to melt.
c. Samantha caused the door to open.

(4) a. Jamie broke the table.
b. Harry melted the butter.
c. Samantha opened the door.

The sentences in (4) do not explicitly use the term "cause", but all the same, they do seem to bear some important relation to the sentences in (3). Understanding precisely how the (4) sentences are similar to, but also different from, the (3) sentences is a topic of ongoing research (Levshina, 2022; Martin, 2018; Rose et al., 2021; Schwenkler & Sievers, 2022; Song & Wolff, 2005; Wolff, 2003), but even without a complete answer to that difficult question, we can note one respect in which they are clearly similar. Both types of sentences involve *causation*. For example, if you say that Jamie broke the table, you are clearly saying that Jamie caused the table to change state in some way. For this reason, verbs like "break" are referred to as *causative verbs*.

Importantly, not all transitive verbs are causative verbs. For some examples, consider (5).

(5) a. Jamie approached the table.
b. Harry touched the butter.
c. Samantha entered the room.

Research in linguistics has explored the ways in which sentences like these differ from the sentences in (4). If we say that Jamie "broke the table," we are saying that Jamie caused the table to change state, but if we say that Jamie "approached the table," we are not saying that Jamie caused the table to change state. In fact, we are not saying that Jamie's bodily movement caused any further outcome. For this reason, transitive verbs of this type are excluded from the class of causative verbs and will be called here "non-causative verbs".[1]

Thus far, we have been emphasizing that causative and non-causative transitive verbs are very different when it comes to causation, but there is another dimension on which they are fundamentally similar: namely, that they can both assign the agent role to the grammatical subject. Consider again the sentences "Jamie broke the table" and "Jamie touched the glass." Each of these sentences describes an event (a breaking event, a touching event), and each says that Jamie played a particular role in that event. The key point now is that although the events themselves are very different, the role that is assigned to the subject of the sentence in both cases is the same. In both cases, Jamie occupies the *agent role* in the event.

A long tradition of research in the study of language has explored this distinctive role (Cruse, 1973; DeLancey, 1984; Dowty, 1979, 1989, 1991; Fauconnier, 2012; Fillmore, 1967; Folli & Harley, 2008; Grimm, 2011; Levin & Rappaport Hovav, 1995; Massam, 2009; Rissman & Majid, 2019; Tollan, 2018; van Valin & Wilkins, 1996; Zúñiga & Kittilä, 2019, among many others). For instance, research shows that the assignment of thematic roles is closely tied to syntax. Suppose you see the sentence: "Tom walked right up to Magdalena and daxed her." Even though you don't know the meaning of the verb "dax", you can tell from the syntax which constituent of the sentence refers to the person who played the agent. Specifically, in sentences with this structure, the nominal phrase in subject position is typically understood as referring to the agent in most languages (Bickel, 2010; Bickel et al., 2015; Dryer, 2005; Fillmore, 1967; Greenberg, 1963; Rissman & Majid, 2019; Sauppe et al., 2023). In other words: if a person's name appears in subject position in sentences with this structure, the sentence is saying that the person played a distinctively agentive role in the event. Additional work in linguistics has led to the development of a variety of different, and often opposing, theories about how to understand the agent thematic role at a deeper level (e.g., Cruse, 1973; DeLancey, 1984; Dowty, 1979, 1991; Folli & Harley, 2008; Martin et al., 2025; Ramchand, 2008; Schlesinger, 1989; van Valin & Wilkins, 1996).

Work within psychology has explored people's ordinary way of reasoning about these roles—and suggests that reasoning about agents' mental states may also be involved in these judgments. For instance, consider the sentence "Jacob is dating Sabrina." This sentence describes a mutual interaction. To say that Jacob is dating Sabrina is also to say that Sabrina is dating Jacob. However, people more readily attribute intentionality to Jacob (occupying the agent role) than they do to Sabrina (occupying the theme role; Strickland et al., 2014). Unlike many of the effects we've discussed so far, effects like these are not specifically about causation. Instead, they seem to reflect a link between intentionality and the agent thematic role.

With this in the background, we can reconsider our sentence "Tom caused the train delay." Existing studies show that people's intuitions about whether it is right to use this sentence depend in part on facts about Tom's mental states. But what property of the sentence gives rise to this effect? Is it the fact that the sentence describes a *causal* relationship between Tom and the train delay, or is it the fact that this sentence assigns to Tom the *agent role*?

One possible way to address this question would be to ask whether there is also an effect of agency on intuitions about non-causative sentences. If we only find an effect of agency for causative sentences, we might think that the effect for causative sentences arises specifically because these sentences describe a causal relationship. By contrast, if we also find an effect for non-causative sentences, we would have at least some reason to think that the effect observed for causative sentences does not arise because of something involving the fact that these sentences are causative. Instead, the effect might be due to something far more general about the agent thematic role.

3. Understanding agents and agency

At the heart of this second approach is the notion of *agency*. Although our hypothesis is concerned specifically with a linguistic effect, this notion has also been explored in numerous other areas of cognitive science, and we will be drawing on that larger body of research here. In particular, we will be drawing on two key ideas that have been emphasized throughout existing research on agency.

The first idea is that people distinguish between *entities* that are agents, with the capacity for agency (e.g., human beings), and those that are not agents (e.g., rocks). A diverse body of work in cognitive science has shown that people are sensitive to various cues that imply the capacity for agency, even in cases of things which are otherwise not obviously agents (see Rose, 2022; see also the classic demonstrations by Heider & Simmel, 1944). Even young children notice subtle indications of agency, like self-propelled motion or the ability to create order (see, e.g., Johnson, 2000; Keil & Newman, 2015; Newman et al., 2010; Poulin-Dubois et al., 1996). For instance, young children perceive novel, autonomously moving 'blobs' as agents, but only when those blobs move towards a goal (e.g., Opfer, 2002), and even five-month-old infants distinguish between the behavior of agents and non-agents (e.g., Woodward, 1998).

The second idea is that people distinguish between *events* that arise through the exercise of full agency and those that do not. If we see a person being violently pushed to the ground, we might think that this person is clearly an agent, but we might also think that this specific movement of her body (being pushed to the ground) was not the result of an exercise of her agency. Our minds are also sensitive to this second distinction across many different contexts. For instance, perceptions of someone's relative exercise of agency influences what things people attend to in the first place (e.g., people will follow the gaze of an agent who intentionally looks away, but not the gaze of an agent whose gaze is merely 'deflected'; see Colombatto, Chen, & Scholl, 2020; see also Colombatto et al., 2019, Colombatto, van Buren, & Scholl, 2020, Colombatto et al., 2021), as well as how even very young children will react to an agent (e.g., Carpenter et al., 1998; Woo et al., 2017). Adults

[1] Among transitive verbs, causative and non-causative verbs significantly differ from each other in a range of interrelated syntactic and semantic properties (Levin & Rappaport Hovav, 1995; Levin, 1999; Rappaport Hovav & Levin, 1998, Alexiadou et al., 2015, Beavers & Koontz-Garboden, 2020, among many others). For instance, while the object of causative verbs is always an (affected) theme, the object of non-causative verbs can often have another semantic role than theme (such as Path in the case of *cross*), because the latter verbs do not express a change caused in the object. A consequence of this is that while near-synonyms of causative verbs typically are transitive themselves (see *melt/thaw*), near-synonyms of transitive non-causative verbs are often intransitive (see *cross/go across*, cf. Levin, 1999:5). Another point is that since causative verbs essentially describe changes, they can also be used to describe just a change of the theme and not its agent (e.g., we find *The door opened* next to *Tom opened the door*). Non-causative transitive verbs, on the other hand, describe a way of acting by an agent (and not a change of the object) and therefore do not allow the demotion of the agent (e.g., we do not find *The line crossed* next to *Tom crossed the line*). Next, while causative verbs are typically compatible with many types of subjects (individuals such as *Tom*, events or facts as in *The accident/this fact caused the delay*, etc.), non-causative verbs are much more restrictive in that they often require an individual-denoting subject (e.g *The accident crossed the line* is not a felicitous statement).

are also sensitive to relatively subtle variations in how much an action seems to reflect an agent's own agency—attributing more agency, for example, to a robot who cheats to win in rock-paper-scissors than to one who cheats to lose (see Litoiu et al., 2015).

The hypothesis we will be exploring here draws on both of these ideas. The hypothesis is that people tend to think that certain sentences do not sound right when (a) the subject of the sentence denotes an entity that they regard as an agent but (b) the verb phrase denotes an event that this entity did not bring about with agentive control.

If this hypothesis does turn out to be correct, we face further questions about how to spell it out in detail at the level of linguistic theory. To do so, we need an account of the syntax of these sentences, and of the semantics of the agent thematic role, and we need a detailed understanding of people's judgments regarding these sentences, such as whether the judgment that they should not be used in certain cases is best understood in terms of their truth conditions or in terms of pragmatic infelicity. In the General Discussion and the Appendix, we provide an account along these lines, drawing on technical tools from natural language semantics. However, we emphasize that the core claim of the present paper does not depend on the details of that account. Although linguists might disagree about precisely how to spell things out, the core hypothesis we will be testing in the present studies is simply this: The impact of agency on judgments of causative sentences does not arise because of something specific to causative sentences but rather because of something far more general about the agent thematic role.

4. Present studies

Across four experiments, we seek to understand the scope of the effect of perceived agency: When is it that people's judgments are and are not affected by how much agency was involved in the scenario?

Study 1 directly examines the influence of perceived agency on the evaluations of causal vs. non-causal sentences: Does perceived agency have the same effect when people are asked about sentences with non-causative verbs?

Studies 2–4 then explore several factors that may affect the role of perceived agency in judgments of both the causal and non-causal sentences. Study 2 examines the actions of *agents* (e.g., Tom) vs. the actions of *inanimates* (e.g., water from a rainstorm). Study 3 examines blameworthy actions of agents (e.g., causing a train delay) vs. harmless actions (e.g., causing a prize to drop down). Finally, Study 4 examines two different aspects of agency: acting *intentionally* (e.g., intending the specific outcome vs. not) and acting with *agentive control* (e.g., deliberately walking across a room vs. tripping and falling). In all studies, we explore people's evaluations of both causal and non-causal sentences.

5. Study 1

How much agency someone exercised in bringing about an outcome affects the extent to which people think they caused the relevant outcome. But is this effect limited to causal sentences? Here, we compare people's evaluations of causative sentences vs. sentences with non-causative verbs in the same scenarios.

5.1. Methods

Data, materials, and preregistration information for this experiment and all following can be found on the Open Science Framework (OSF) at https://osf.io/7a6fq/?view_only=08f4cf81aa4647a898ae47789fac81ca.

5.1.1. Participants

Four hundred adult participants completed a survey online through Prolific, (M_{age} = 28.2, SD_{age} = 8.5, 70.4 % white, 72.0 % female). All participants lived in the United States. Data from an additional 11 participants were collected but excluded for failing a comprehension check (see Procedure section).

5.1.2. Stimuli

Stimuli consisted of eight short vignettes (each participant saw a single vignette; see Procedure section). These vignettes described four possible scenarios in which a person, Tom, acted with either full agency or with very low agency. For example, in one scenario, participants were told that Tom is waiting for a train and that there is a yellow line on the platform that people aren't supposed to cross:

> *Tom is waiting for a train. In order to keep the passengers clearly out of the way from moving trains, nobody is supposed to cross a yellow line drawn on the platform.*

In the full agency condition, Tom then deliberately crosses over the line:

> *The train platform is very crowded today. Tom unexpectedly decides to cross the line to get in front of the crowd. He deliberately steps over the yellow line to stand in front of it.*

In the low agency condition, Tom passes out and falls over the line:

> *The train platform is very crowded today. In the heat, Tom unexpectedly blacks out and falls over the line.*

The same outcome then follows as a result:

> *Tom is now too close to the edge of the platform, and so the approaching train automatically initiates an emergency stop. Nobody is hurt, but this train and those following are delayed by several hours as a result of the incident.*

The full text of all of the vignettes is available in Table 1.

5.1.3. Procedure

Participants were randomly assigned to one of 16 conditions in a 4 (Scenario: hiking, train station, car, room) x 2 (Agency: full agency, low agency) x 2 (Statement type: causative, non-causative) between-subjects design. Participants were shown one of eight short vignettes about a scenario in which Tom acted with either full agency or with very low agency (see Stimuli section). They were asked to evaluate either a causative statement (e.g., "Tom caused the train delay") or a statement with a non-causative verb (e.g., "Tom crossed the line"). Across our four scenarios, these verbs included two verbs of contact "touch", "hit" and two path verbs "cross", and "enter." Participants were asked to respond to a 1–7 scale on the basis of whether this sentence was a "natural/valid way of describing the event."

Finally, participants were asked a comprehension question about whether Tom acted intentionally (e.g., "Tom intentionally crossed over the line") or with low agency (e.g., "Tom blacked out and fell over the line"). Participants who failed the comprehension check were excluded and replaced (see Participants section).

5.2. Results

Results are displayed in Fig. 1. Data were analyzed using R with the lme4 (Bates et al., 2015) and emmeans (Lenth, 2018) packages.

Data were fit to linear mixed-effects models, with *agency* and *statement type* (causative vs. non-causative) as fixed effects and vignette as a random effect (random intercepts only). There was a significant main effect of agency, $\chi^2(1) = 136.42$, $p < .001$, and a smaller main effect of statement type, $\chi^2(1) = 10.83$, $p = .001$. However, there was no significant interaction between agency and statement type, $\chi^2(1) = 0.52$, $p = .47$.

Our primary interest (preregistered) was in whether or not there was a significant effect of agency within each statement type. Using estimated marginal means, we found that participants were significantly more likely to endorse a causative sentence (e.g., "Tom caused the train

Table 1
Full text of all vignettes in Study 1. Each of the four columns shows the full text of one scenario. Participants were shown either the text from the full agency *or* low agency conditions, depending on which agency condition they were assigned. Then, for each vignette, participants were asked either about a causative sentence regarding that vignette (e.g., "Tom caused the train delay") or a non-causative sentence regarding that vignette (e.g., "Tom crossed the line").

	Tom crossed…	Tom touched…	Tom entered…	Tom hit…
	Tom is waiting for a train. In order to keep the passengers clearly out of the way from moving trains, nobody is supposed to cross a yellow line drawn on the platform. The train platform is very crowded tonight.	Tom is going hiking in a national park. In order to protect the natural habitat of this mountain range, nobody is supposed to touch anything that isn't on the hiking trail itself—animals, plants, and rocks are all supposed to be left alone. Tom makes it to the end of his hike, at the top of a cliff. Tom is enjoying the view off the mountain.	Tom works as a security guard in a museum. The security guards all have master keys, but they aren't supposed to enter the old storage room in the basement. In fact, nobody remembers what is in the storage room. Part of Tom's job is monitoring the stairs leading down to the basement.	Tom is driving in a race car race. One of the basic safety rules of the race is to avoid hitting the fence around the track—drivers who hit the fence are automatically disqualified. Tom is leading the race and is confident that he will win.
Full agency	*Tom wants to get in front of the crowd. He unexpectedly decides to cross the line to get in front of the crowd. He deliberately steps over the yellow line to stand in front of it. Tom is now too close to the edge of the platform, and so the approaching train automatically initiates an emergency stop.*	*Unexpectedly, he decides to throw a rock off the edge so that he can watch it fall. He deliberately steps off the hiking trail, picks up a large rock, and throws it off the cliff.*	*Tom becomes curious about the old room. One night, he unexpectedly decides to enter the room. He deliberately unlocks the room and goes in.*	*Then, he is passed by another driver who has been slowly gaining on him. Tom becomes angry. He unexpectedly decides to throw the competition. He deliberately drives his car into the fence around the track.*
Low agency	*In the heat, Tom unexpectedly blacks out and falls over the line. Tom is now too close to the edge of the platform, and so the approaching train automatically initiates an emergency stop.* Nobody is hurt, but this train and those following are delayed by several hours as a result of the incident.	*Unexpectedly, he ends up suffering a stroke. He passes out and falls off the hiking trail and against a large rock, which then falls off the cliff.* When the rock falls, it hits someone on the hiking trail below, injuring them.	*One night, he isn't feeling well and unexpectedly faints. As Tom passes out, he falls down the stairs, ultimately knocking the door to the storage room open and falling in.* This sets off a sensor just inside the room, triggering the building's system of alarms.	*Then, he unexpectedly suffers a seizure. His out-of-control car drives into the fence around the track.* The sudden stop against the metal creates sparks and sets off a small fire.

delay") when Tom acted with full agency ($M = 6.34$, $SD = 0.91$) vs. with very low agency ($M = 4.00$, $SD = 2.00$), $t(399) = 9.50$ $p < .001$. The same was true for their evaluations of non-causative sentences (e.g., "Tom crossed the line"): Participants rated these sentences as more natural when Tom acted with full agency ($M = 5.64$, $SD = 1.83$) vs. with very low agency ($M = 3.55$, $SD = 2.05$), $t(399) = 8.49$, $p < .001$.

5.3. Discussion

Perceived agency has previously been found to influence people's causal judgments, suggesting that reasoning about how much agency was involved is part of how people understand what qualifies as a cause of a given outcome. Yet here we find that this phenomenon may actually be far more general than causal cognition. Whether Tom acted intentionally or with low agency affected not only the extent to which people endorsed causal sentences, but also the extent to which they endorsed sentences with non-causative verbs (i.e., sentences with path or contact verbs like "cross" or "touch"). These results suggest that there may be a more general story as to how it is that perceptions of agency are involved in people's understanding of sentences about agents' actions—even beyond their causal judgments.

6. Study 2

In the previous study, participants evaluated sentences that were all about a particular kind of entity: a person, Tom, who we would typically think of as an animate agent and who we would therefore expect to act by exercising agency. When this agent did *not* act with their typical level of agency (e.g., when they had a stroke and passed out), participants were less likely to agree with both causal and non-causal sentences about them.

But consider the following sentences:

(4) a. The water caused the train delay.
b. The water crossed the line.

Here, the water occupies the same role in the sentence as Tom might, crossing a line or causing a train delay. Unlike Tom, however, the water is inanimate and thus *cannot* act with a high level of agency. In other words, the water more resembles Tom when he is incapacitated in that it has a low level of agency. But do people treat these sentences about inanimate agents in the same way that they do sentences about animate agents?

There are two possibilities, with very different implications for how to best explain the findings from Study 1. One possibility is that people are just as reluctant to agree with sentences about inanimate agents (such as water) as they are to agree with sentences about incapacitated animate agents (such as Tom when he has suffered a stroke). On this view, people are simply less likely to agree with sentences in which the subject exercises a low level of agency. Tom acts with low agency due to being incapacitated, and the water acts with low agency in virtue of being inanimate. On this view, people may equate these cases because they involve similarly low levels of agency.

Another possibility is that people are perfectly willing to agree with sentences about inanimate agents, even though they are acting with low agency. On this view, the relevant difference between the scenarios where Tom acts with high vs. low agency is that people are less likely to agree with sentences in which the subject is not exercising its full capacity to be agentive. This view differs in a subtle but important way from the first. Here, sentences involving an incapacitated Tom seem less natural because we *expect* Tom to act with more agency, but he does not. However, inanimate things, like water, are not expected to act with high levels of agency. Thus, this view would predict that sentences involving inanimate agents acting with little agency (as they typically do) will be judged similarly to those involving animate agents acting with full agency (as they typically do).

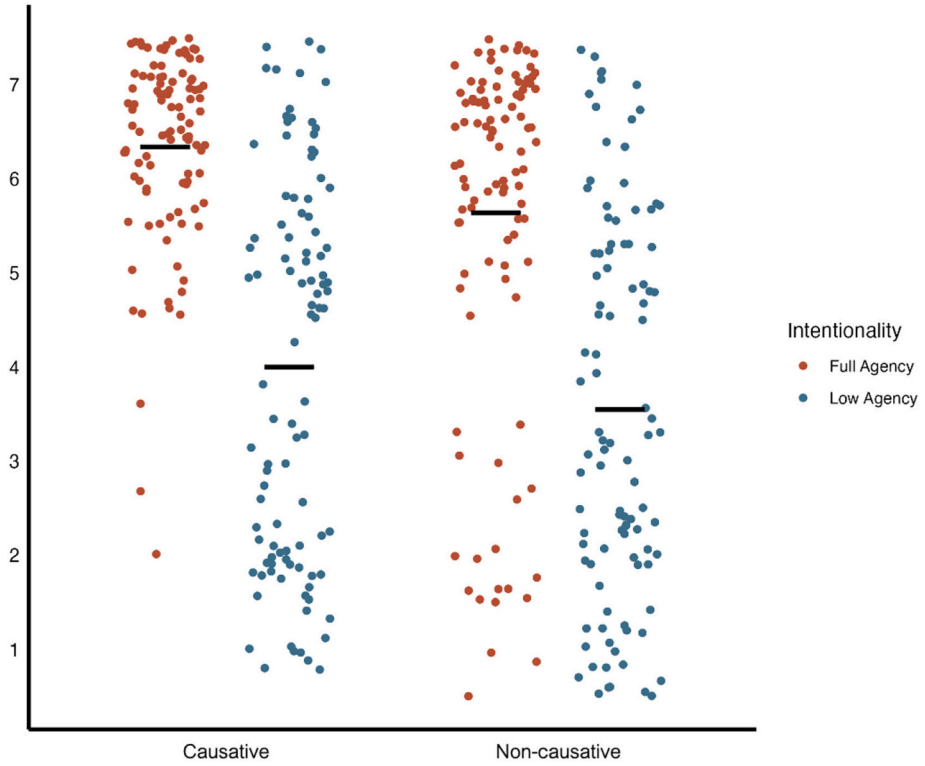

Fig. 1. The results by condition in Study 1. Participants were asked to evaluate the extent to which the relevant (causative or non-causative) sentence was a "natural/valid way describing the event." Jittered points show the responses of individual participants. Black lines represent group means.

We introduce sentences with inanimate subjects in order to distinguish between these possible explanations. We ask: Is the inanimate agent treated like Tom when he is exercising his full agency (in that both reflect an agent acting with their full capacity to be agentive), or like Tom when he is incapacitated (in that both reflect an agent who is acting with very low agency)?

6.1. Methods

All elements of the experimental design were identical to those of Study 1, except as stated below.

6.1.1. Participants

600 new participants completed a survey online through Prolific, (M_{age} = 30.3, SD_{age} = 9.1, 67.8 % white, 61.5 % female). This sample size was chosen in order to have the same number of participants per condition as in Study 1. Data from an additional 35 participants were collected but excluded for failing a comprehension check.

6.1.2. Stimuli

Participants were shown one of twelve short vignettes. These covered the same four scenarios (hiking, train station, car, room) as in Study 1. For each scenario, participants were assigned to one of three agency conditions, resulting in vignettes about either (1) a person, Tom, acting with full agency, (2) a person, Tom, acting with low agency, or (3) an inanimate entity (e.g., water from a storm) acting the way inanimates do (i.e., with low agency). Both of the conditions involving an animate agent (i.e., Tom) were closely adapted from the vignettes in Study 1; the only changes were in order to be consistent with the inanimate condition. Consistent with Study 1, participants in all conditions were given the same initial context about norms that were in place in the scenario (e.g., that there was a line people aren't supposed to cross).

In the inanimate condition, participants were told that something acted in the same way that Tom did in the other conditions (e.g., crossing a line). For example, in one vignette, participants were told that water from a storm crossed the line and caused a train delay:

> *One day, there is an unexpectedly strong storm in the area. Rain floods the train station. It covers the platform, over the yellow line. The water is so heavy near the edge of the platform that it triggers the approaching train to initiate an emergency stop. Nobody is hurt, but this train and those following are delayed by several hours as a result of the incident.*

The full text of all of the vignettes is available on our OSF page.

6.1.3. Procedure

Participants were randomly assigned to one of 24 conditions in a 4 (Scenario: hiking, train station, car, room) x 3 (Agency: person with full agency, person low agency, inanimate) x 2 (Statement type: causative, non-causative) between-subjects design.

6.2. Results

Results are displayed in Fig. 2.

Data were fit to linear mixed-effects models, with agency and statement type as fixed effects and vignette as a random effect (random intercepts only). As found in Study 1, there was a significant main effect of agency, $\chi^2(1) = 114.9$, $p < .001$, and a much smaller effect of statement type, $\chi^2(2) = 5.12$, $p = .02$. There was again no significant interaction between agency and statement type, $\chi^2(1) = 0.72$, $p = .70$.

Our main interest (preregistered) was not in the main effect of agency, but in the specific pairwise comparisons between the agency conditions. Agency affected participants' evaluations of sentences about Tom, such that sentences describing Tom's actions were more valid when Tom acted intentionally ($M = 5.76$, $SD = 1.52$) than when he acted with low agency ($M = 3.98$, $SD = 2.01$), $t(601) = 10.37$ $p < .001$. In contrast, agency did not affect participants' evaluations of sentences about inanimate entities in the same way. Participants were significantly more likely to endorse a sentence like "The water caused the train delay"

(even though the water also acted with a very low degree of agency) than they were to endorse the equivalent sentence about Tom acting with very low agency ($M = 5.51$, $SD = 1.63$), $t(601) = 8.89$, $p < .001$. In fact, participants' evaluations of sentences about inanimate entities were not significantly different from their evaluations of sentences about Tom acting intentionally, $t(601) = 1.49$, $p = .30$.

6.3. Discussion

Here, we find that, in participants' evaluation of sentences about ordinary events, inanimate agents (e.g., water) are evaluated not like an incapacitated animate agent (e.g., Tom after a stroke) but instead like an ordinary animate agent (e.g., Tom deliberately crossing a line). In other words, people seem to be evaluating these sentences based not on the absolute agency of the subject, but rather on its level of agency *relative to a typical level of agency*. People agree with the sentence "The water caused the train delay," even though the water cannot and did not act with high levels of agency. However, the water did act in its full *capacity* for agency, and thus with its typically expected level of agency.

This finding clarifies why it is that people are reluctant to agree with the sentence about Tom acting when he is incapacitated. People do not find it natural to say that "Tom caused the train delay" when he passed out not just because he was acting with very little agency, but because he is an agent that we typically expect to act with a much higher level of agency.

In addition to our key finding about inanimate agents, we also again found that these effects of agency are not limited to instances of causation. All observed effects of agency were consistent across the causal and non-causal sentences.

7. Study 3

In all of the examples examined so far, Tom acts in a way that leads to a negative outcome and so may be worthy of blame. For instance, when Tom crosses the line, the trains are forced to stop, causing a delay. Might the negative valence of the outcome in these cases be affecting participants' judgments?

One possibility is that the impact of agency observed in these studies will arise only when the outcome has a negative valence. After all, studies consistently find an impact of moral considerations on causal judgments (Alicke, 1992; Hitchcock & Knobe, 2009), and recent work has led to the development of numerous different theories designed to explain that effect (Alicke et al., 2011; Driver, 2008; Halpern & Hitchcock, 2015; Icard et al., 2017; Quillien, 2020; Samland & Waldmann, 2016). It might be argued that some of these theories would also predict an effect on non-causal judgments such as the ones we have been exploring here. If so, the phenomenon we have been exploring might turn out to be simply one instance of a broader phenomenon involving the impact of moral considerations. When the outcome has a negative valence, people may think it is morally wrong for Tom to bring about that outcome through an exercise of agency, but that it is not morally wrong for Tom to bring about the outcome just by having a stroke—and this moral difference may lead to the effect on people's evaluations of the relevant sentences.

However, another possibility is that the effect will arise even in cases that do not involve outcomes with a negative valence. Perhaps perceived agency will affect people's judgments of these sentences regardless, even if Tom is not violating any norms and the outcomes are completely harmless. If this is the case, then the effect of perceived agency could not be attributed to an impact of moral considerations.

Here, we test whether valence influences the relation between agency and causation by having participants view vignettes with neutral as well as negative outcomes.

7.1. Method

All elements of the experimental design were identical to those of previous experiments except as stated below.

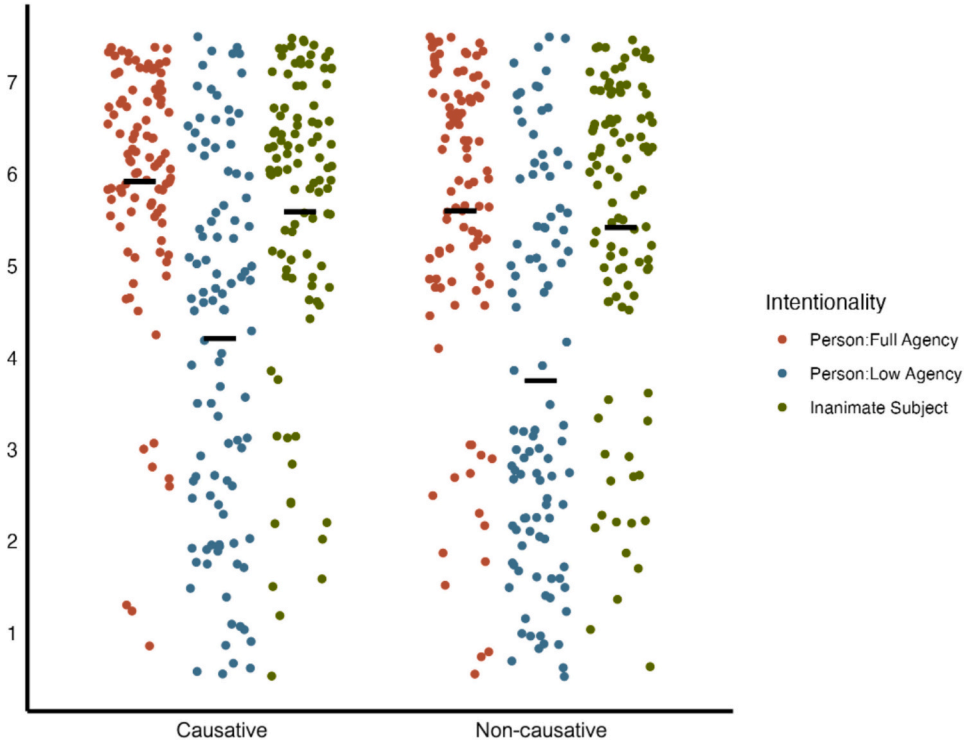

Fig. 2. The results by condition in Study 2. Participants were asked to evaluate the extent to which the relevant (causal or non-causal) sentence was a "natural/valid way describing the event." Jittered points show the responses of individual participants. Black lines represent group means.

7.1.1. Participants

800 new participants completed a survey online through Prolific, (M_{age} = 36.2, SD_{age} = 13.5, 72.2 % white, 66.7 % female). This sample size was chosen in order to have the same number of participants per condition as in previous experiments. Data from an additional 9 participants were collected but excluded for failing a comprehension check.

7.1.2. Stimuli

Participants were shown one of sixteen short vignettes. These covered the same four scenarios (hiking, train station, car, room) and the same agency conditions (full agency, low agency) as in Study 1. Participants were also assigned to one of two valence conditions, such that Tom's actions had either (1) a negative valence (as in prior experiments, breaking a norm and causing a negative outcome; e.g., crossing a line that isn't supposed to be crossed and causing a train delay), or (2) a neutral valence (following norms and causing an innocuous outcome). For example, in the neutral case of Tom crossing a line, participants were told that Tom was playing a carnival game:

> *Tom is at a carnival, playing games. In the game he is playing now, people try to jump as far as they can to cross over different colored lines drawn on the ground for different prizes. When participants cross the lines, different desserts drop down on pies they can keep.*

Then, as in the negative valence cases, Tom crossed the line either by acting with full agency:

> *Tom is looking at the different possible lines, when he decides to jump over the yellow line. He deliberately jumps over the yellow line and lands in front of it. This automatically initiates a whipped cream faucet. The faucet turns on and a stream of whipped cream pours down onto a pie.*

or by acting with minimal agency:

> *Tom is looking at the different possible lines, when he suffers a heat stroke. He passes out and falls over the nearby yellow line. This automatically initiates a whipped cream faucet. The faucet turns on and a stream of whipped cream pours down onto a pie.*

The full text of all of the vignettes is available on our OSF page.

7.1.3. Procedure

Participants were randomly assigned to one of 32 conditions in a 4 (Scenario: hiking, train station, car, room) x 2 (Agency: full agency, low agency) x 2 (Valence: negative, neutral) x 2 (Statement type: causative, non-causative) between-subjects design.

7.2. Results

Results are displayed in Fig. 3. Data were fit to linear mixed-effects models, with agency, valence, and statement type as fixed effects and vignette as a random effect (random intercepts only). As in previous experiments, there was a significant main effect of agency, $\chi^2(2) = 162.48$, $p < .001$, and a smaller effect of statement type, $\chi^2(2) = 36.57$, $p < .001$. However, there was no significant main effect of valence, $p = .12$. There was also a significant interaction between agency and valence, $\chi^2(2) = 15.62$, $p < .001$, and a significant three-way interaction, $\chi^2(1) = 3.86$, $p = .050$.

To better understand these interactions, we next looked at specific pairwise comparisons between agency conditions, within each valence condition. (Because there were no significant interactions with statement type, we did not look at pairwise comparisons also within each statement type; this analysis plan was preregistered.) In both the negative and neutral valence conditions, participants rated the causative and non-causative sentences to be more natural when Tom acted with full vs. very low agency, all $ts > 7.11$, $p < .001$. However, there was a larger effect in the negative conditions (Full agency: $M = 6.14$, $SD = 1.49$; Low agency: $M = 4.02$, $SD = 2.10$) than in the neutral conditions (Full agency: $M = 5.52$, $SD = 1.62$; Low agency: $M = 4.26$, $SD = 2.02$).

7.3. Discussion

Here, we find that the basic effect of perceived agency persists regardless of the valence of the scenarios in question. Whether Tom's actions could be seen as blameworthy, people were more inclined to endorse sentences describing Tom acting with full agency than sentences describing him acting with reduced agency. This effect was once again also consistent regardless of whether or not the sentence described Tom as causing some further outcome.

We also found an interaction such that there was a greater effect of agency when Tom's actions resulted in a negative outcome (vs. a neutral outcome). This interaction suggests that in addition to the effect we have been focusing on here, there is also an effect of moral considerations. When the outcome is negative, Tom's behavior is seen as more of a norm violation when he acts intentionally than when he acts with very low agency, and this difference appears to be triggering the moral effect that has already been explored in many previous studies.

In sum, although there does appear to be some effect of negative valence, the results indicate that there is also an effect of agency that arises even in the absence of negative valence.

8. Study 4

Across three experiments, we find that there are generalizable effects of agency—such that perceptions of agency influence how people evaluate both causal and non-causal statements. In short, when animate agents act with reduced agency, people are less inclined to agree with the class of action sentences we are looking at (built with causative or non-causative verbs). But what does it mean to say that an agent acts with reduced agency? So far, we have referred to an agent acting with full or with low agency, but have not specified what aspects of agency have been reduced.

One possible approach would be to try to explain this notion by drawing on ideas from cognitive science research on the influence of agents' mental states on different judgments. For example, if we are trying to understand the role of agency in judgments about the sentence "Tom caused the train delay", we might focus on the impact of thinking that Tom *intended* to bring about the train delay, or that he *knew* that he would bring about the train delay. Existing research in cognitive science finds that mental state judgments like these impact people's cognition in multiple different domains (e.g., Bloom, 1996; Cushman & Young, 2011; Noyes & Dunham, 2017), and it may well be that the impact of agency on the use of causal sentences should be understood in much the same way.

A second and very different approach would be to try to explain this notion by drawing on ideas from the linguistics literature about what it means to exercise agency. As this literature has emphasized, it is possible for a sentence to describe something that a person did through her own agency even if that sentence is not describing something that the agent intended to do. For example, consider the sentence: "The child accidentally ate something poisonous" (Kittilä, 2005). Here, the child did not intend to eat something poisonous. Yet at the same time, it is also clear that the child's behavior was the result of an exercise of her own agency, i.e., that the actual bodily movements she performed were under her agentive control, rather than arising because she had a stroke or because someone else had grabbed her body and moved her limbs.

In this experiment, we independently manipulate intention and agentive control to assess the impact of each on the effect we have thus far been exploring. For example, suppose that a racetrack has been set up in such a way that if a person crosses over a particular line, confetti will automatically fall. Now suppose that Tom crosses over the line, and confetti falls. We can now manipulate Tom's mental state regarding this outcome and also independently manipulate whether Tom exercises agentive control over his own bodily movements. The key question is whether either or both of these factors will impact people's judgments about the causal and non-causal sentences.

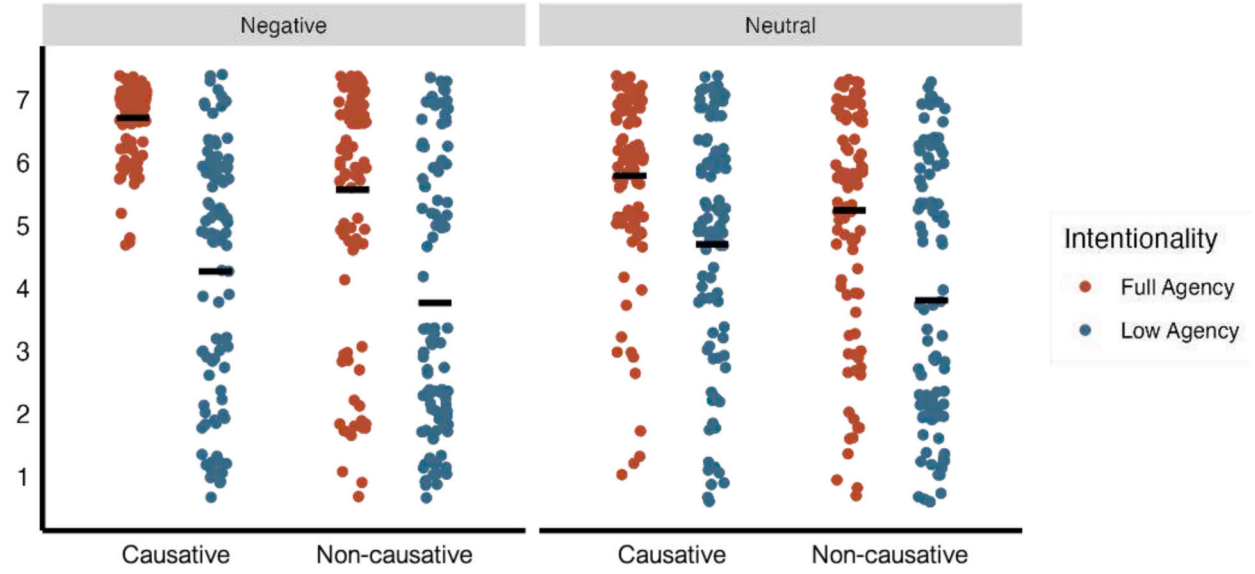

Fig. 3. The results by condition in Study 3. Participants were asked to evaluate the extent to which the relevant (causal or non-causal) sentence was a "natural/valid way describing the event." Jittered points show the responses of individual participants. Black lines represent group means.

Because we are interested in agency broadly, here our manipulations are ones that involve temporary loss of control (e.g., stumbling or tripping) rather than total incapacitation (e.g., passing out from a stroke).

8.1. Methods

All elements of the experimental design were identical to those of previous experiments except as stated below.

8.1.1. Participants

800 new participants completed a survey online through Prolific, (M_{age} = 36.7, SD_{age} = 12.5, 71.2 % white, 57.8 % female). This sample size was chosen in order to have the same number of participants per condition as in previous experiments. Data from an additional 8 participants were collected but excluded for failing one or more comprehension checks (see Procedure section).

8.1.2. Stimuli

Participants were shown one of sixteen short vignettes. These covered four possible scenarios, which were designed to involve the same actions described by the non-causative verbs in previous experiments (i.e., "touch", "hit", "cross", "enter"). All scenarios had a neutral valence (i.e., had no norm violations or negative outcomes). For example, in one scenario, Tom was described at a park where races often finish:

Tom is spending time in a park where local races often finish. There is a set area that serves as a finish line for races, and when people cross the line, confetti will automatically fall.

Within each scenario, Tom was described as either knowing or being ignorant about the line and the consequences of crossing it:

Tom is very familiar with the park and today's race. He knows where the finish line is, and what happens when it is crossed.

or

Tom has never been to the park and doesn't know about the race. He has no idea where the finish line is, or what happens when it is crossed.

Tom was also described as either having agentive control or not:

Tom is running across the park to greet a friend. In doing so, he runs over the race's finish line, and the confetti falls.

or

Tom is walking around the park when his foot gets caught on a tree root, and he trips. As he falls, he stumbles over the race's finish line, and the confetti falls.

The full text of all of the vignettes is available on our OSF page.

8.1.3. Procedure

Participants were randomly assigned to one of 32 conditions in a 4 (Scenario: hiking, train station, car, room) x 2 (Knowledge: knowledge, ignorance) x 2 (Control: control, no control) x 2 (Statement type: causative, non-causative) between-subjects design.

At the end of the task, participants were asked two comprehension check questions (modeled after comprehension questions from previous experiments about whether or not Tom acted with agency). The first question checked to see whether participants understood that Tom acted with / without agentive control (e.g., whether it was the case that Tom tripped and fell); the second checked whether participants understood that Tom acted with / without knowledge (e.g, whether it was the case that Tom knew there was a race). Participants who failed either comprehension check were excluded and replaced (see Participants section).

8.2. Results

Results are displayed in Fig. 4.

Data were fit to linear mixed-effects models, with agentive control, knowledge, and statement type as fixed effects and vignette as a random effect (random intercepts only). As in previous experiments, there was a significant main effect of agentive control, $\chi^2(2) = 52.24$, $p < .001$, and a smaller effect of statement type, $\chi^2(2) = 13.77$, $p < .001$. However, there was no significant main effect of knowledge, $p = .30$. There was also a significant interaction between agentive control and statement type, $\chi^2(2) = 11.18$, $p = .004$, between knowledge and statement type, $\chi^2(2) = 6.43$, $p = .04$, between agentive control and knowledge, $\chi^2(2) = 6.13$, $p = .04$, as well as a significant three-way interaction, $\chi^2(1) = 16.13$, $p = .005$. Below, we decompose these interactions individually.

To further explore these interactions, we conducted pairwise comparisons using emmeans, looking at the impact of control separately within each level of the knowledge and statement variables. There was a significant effect of control within each pair, but the effect size differed

across pairs. Specifically, in the conditions where the agent had knowledge and participants received the causative statement, there was a smaller difference between the condition with control ($M = 5.6$, $SD = 1.6$) and the condition without control ($M = 5.1$, $SD = 1.8$), $p < .05$. The other three pairs showed a larger effect. For the conditions with no knowledge and a causative statement, control ($M = 5.7$, $SD = 1.3$) received higher ratings than no control ($M = 5.1$, $SD = 1.8$), $p < .01$. For the conditions with knowledge and a non-causative statement, control ($M = 5.9$, $SD = 1.3$) received higher ratings than no control ($M = 4.3$, $SD = 2.00$), $p < .001$. For the conditions with no knowledge and a non-causative statement, control ($M = 5.2$, $SD = 1.6$) received higher ratings than no control ($M = 4.4$, $SD = 2.0$), $p = .001$.

8.3. Discussion

This final experiment yielded two key findings. First, whether or not Tom had agentive control over his own movements consistently impacted participants' judgments regarding his actions—both in causal and non-causal sentences, and both when Tom knew about all relevant dimensions to his actions and when he was ignorant. In other words, even when Tom had absolutely no idea what he was doing, participants were more inclined to say that our sentences sounded right when his bodily movements were under his own agentive control. Moreover, the cases of reduced agency involved ordinary ways that one might temporarily lose full control of their own movements (e.g., stumbling or tripping) in contrast to the total incapacitation (e.g., passing out from a stroke) examined in Studies 1–3. Even in the case of these subtler differences in agentive control, we observed a consistent effect on how people evaluated sentences about Tom's actions.

Second, not only do we find a main effect of agentive control, we also find no main effect of whether the agent knew what he was doing. In other words, there was no overall tendency such that participants were more inclined to say that our sentences sounded right when Tom knew what he was doing than when he did not. We were quite struck by this null effect, and we explore it further in the General Discussion.

9. General discussion

People do not find it natural to say that a person caused an outcome when that person did not exercise agentive control. One highly plausible approach to explaining this effect—taken by much prior work—would be to say that it is due to something about causal cognition in particular (i.e., that agency is somehow related to how humans understand causation). Another very different approach would be to say that it reflects a more general phenomenon, perhaps involving the linguistic properties of a broader class of sentences in which a person appears in subject position.

Four studies explored these two possible approaches. Study 1 showed that the effect arises not only for causal sentences but also for sentences with non-causative verbs. Study 2 showed that the effect arises only for sentences with animate subjects. Study 3 showed that the effect arises regardless of whether the outcome is something morally bad or something neutral in the valence. Study 4 showed that the effect is driven not by perceptions of the mental states that the agent has towards the outcome but rather by whether the agent's behavior reflected a controlled exercise of agency.

Taken together, these studies appear to provide evidence against the view that these effects can be explained by causal cognition in particular. Instead, they appear to provide support for an explanation based on the linguistic properties of a broad class of sentences. In what follows, we explore the implications of these findings both for work in causal cognition and for work on agency in language.

10. Agency and causation

The present studies find that there is an effect of agency not only for sentences that involve causing some further outcome, such as “Tom caused the train delay,” but also for non-causative sentences, such as “Tom crossed the line.” Although it would be possible in principle to suggest that this is all just a coincidence (i.e., that the effect for causative sentences and the effect for non-causative sentences arise for unrelated reasons), we find no interaction between agency and statement type. Therefore, the more parsimonious explanation is that there is a single underlying effect here that arises for both causative and non-causative sentences. A question now arises as to what this result indicates regarding the relationship between the impact of agency and people‘s way of thinking about causation specifically.

First off, the results certainly do not show that the effect is completely unrelated to people's capacity for causal cognition, broadly construed. One might think that people's basic capacity for understanding agency is best understood in terms of a generative causal model (Baker et al., 2017). Similarly, it has been suggested that people's ordinary distinction between physical movements that reflect agency (e.g., intentionally moving one's arm) and physical movements that do not

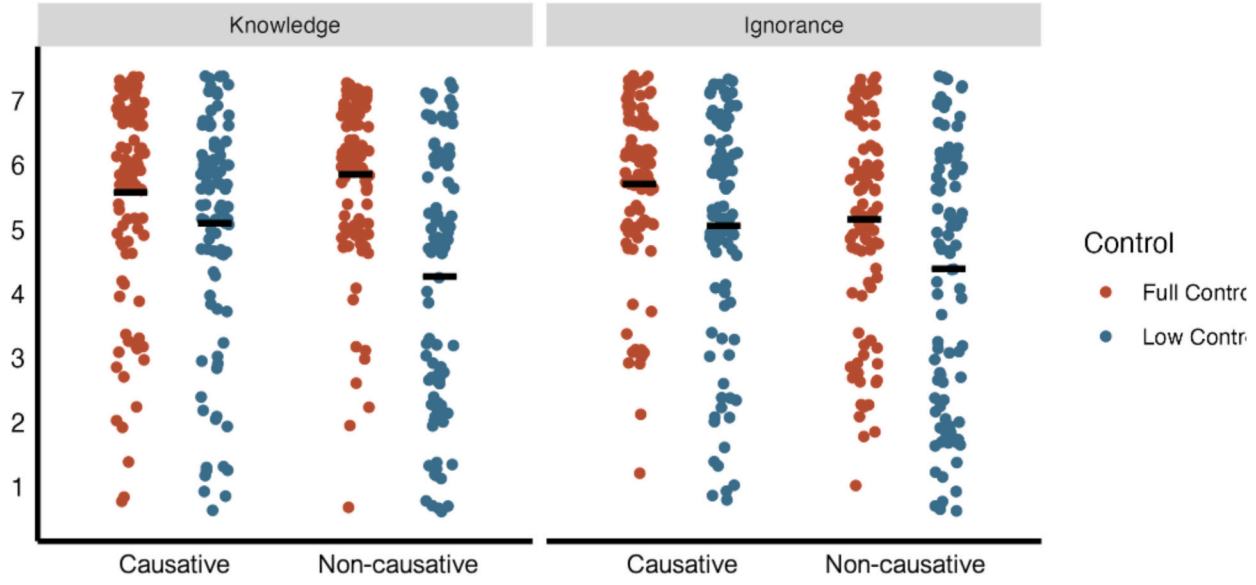

Fig. 4. The results by condition in Study 4. Jittered points show the responses of individual participants. Black lines represent group means.

reflect agency (e.g., arm movements that arise only because one has just had a stroke) should be understood in terms of counterfactuals in a way that can be derived from more general theories of causal cognition (Quillien & German, 2021). Nothing in the present studies provides evidence against any of these claims.

Still, the present studies do show that the effect of agency is not limited to causal statements. A sentence of the form "Tom caused…" has certain special properties that are not shared by certain other classes of sentences. It seems to be saying that what Tom did caused some further outcome, one that goes beyond Tom's behavior itself. There has been a great deal of research on sentences of this type, both linguistically and psychologically (e.g., Glass, 2023; O'Neill et al., 2022; Quillien, 2020). The key point now is that the impact of agency is not limited to sentences of that distinctive type. It seems to be a matter of something far more general.

Finally, a question arises as to whether there are any effects of mental states on causal statements for which the opposite conclusion holds – effects that genuinely are specific to causal statements in particular. When it comes to the effect we have been exploring in the present studies, we find evidence that the same result obtains when one switches over to non-causative sentences, but are there perhaps other effects that would disappear if one switched to non-causative sentences?

One especially promising place to look for such an effect would be in the studies from Kirfel and Lagnado (2021a). In these studies, the agent is performing a behavior through an exercise of her own agency in all conditions, but people's causal judgments end up changing depending on whether the agent knows that this behavior will bring about a particular further outcome. It seems unlikely that the framework we have been developing here could explain this effect. Moreover, the results appear to be very opposite of the null effect we find within the causal sentence conditions in the present Study 4. How then are these results to be explained?

The most salient difference between the Kirfel and Lagnado studies and the present Study 4 lies in the valence of the outcomes. Our Study 4 used outcomes with a neutral valence, whereas the Kirfel and Lagnado studies used outcomes with a negative valence. Thus, the effect observed in those studies might potentially be explained in terms of the impact of moral considerations on causal judgments. Previous research has found that moral considerations can impact causal judgments (Alicke et al., 2011; Driver, 2008; Halpern & Hitchcock, 2015; Icard et al., 2017; Quillien, 2020; Samland & Waldmann, 2016). Research also finds a complex relationship between attributions of ignorance and moral judgment (e.g., Murray et al., 2019; Murray et al., 2023; Young & Saxe, 2011). One possibility would be that these are cases in which ignorance does have an impact on moral judgment and that it is this impact that explains the effect for causal judgment.

In sum, the present studies suggest that some of the effects of agency on causal judgment observed in previous research might not be specific to causal judgments. For each of those separate effects, a new question arises as to whether the effect is specific to causal judgments or not. In some cases, we might find that the effect is due to something more general that applies also to non-causative sentences, whereas in others, we might find that it is specific to causative sentences. When it does appear that the effect is specific to causative sentences, we would face a question as to what explains the effect. As with the work of Kirfel and Lagnado, one possible answer would involve the well-established impact of moral considerations on causal judgments.

11. Agency in language

The present findings also have implications for the study of agency in language. Existing research in this area has developed complex frameworks for understanding the role that agency plays in people's judgments about certain sentences (Folli & Harley, 2008; Kittilä, 2005; van Valin & Wilkins, 1996; Wolff et al., 2009). The present studies then reveal some potentially interesting new patterns in people's judgments that such frameworks will need to explain.

Specifically, we find that many participants think it sounds wrong to use certain sentences in cases where the agent denoted by the subject of the sentence did not perform the action denoted by the verb phrase with agentive control. We then find that this effect arises for both causal and non-causal sentences (Study 1), that it only arises when the subject is animate (Study 2), that it arises regardless of the valence of the outcome (Study 3), and that it depends only on whether the agent denoted by the subject performed the action denoted by the verb phrase through an exercise of her agentive control, not whether she did so knowingly (Study 4).

We now offer a more specific linguistic hypothesis that spells out how these effects arise. The full hypothesis, drawing on technical tools from semantics and pragmatics, appears in the Appendix. In this section, we provide a brief non-technical overview of how the hypothesis works.

At the core is the idea that there is an important similarity between causal sentences like "Tom caused the train delay" and sentences with non-causative verbs like "Tom touched the rock." The impact of agency is then to be explained in terms of the aspect of the structure of these sentences that is shared across causal and non-causal sentences. Very broadly speaking, the claim is that a sentence like (5a) has a meaning that can be paraphrased with (5b).

(5) a. Tom caused the train delay.

b. There was an event that is a causing of the train delay, and Tom occupies the agent role in that event.

In other words, the semantics of this sentence includes at least two distinct elements that might be worth exploring: the notion of *causing* and the notion of occupying the *agent role* in an event. The idea is that the impact of agency on intuitions about these sentences arises not from anything about causation but rather from something about the agent role.

The next key claim is that 'agent' is ambiguous between two different meanings. One is a simple 'effector' meaning (an entity that *does* something); the other is an 'in-control agent' meaning (an entity that does something by exerting agentive control). These two meanings are ordered in such a way that one of them entails the other. An in-control agent of an event will always also be an effector of this event, while the reverse is not necessarily true.

People tend to prefer the stronger meaning, but what the stronger meaning is depends on the linguistic context. When an inanimate like 'water' occupies the agent role, people will only consider the 'simple effector' meaning of 'agent', and will then accept sentences such as 'The water touched the rock', even if the water does not exert any control on its behavior. But now suppose that Tom has a stroke and his finger touches the rock, and consider people's intuitions about the sentence "Tom touched the rock." In that case, Tom is a human being, but he only is an effector of this event. On the view we develop, in a linguistic context as the one we have in "Tom touched the rock" (what is called an 'upward-entailing environment'), the 'in-control' meaning of 'agent' is stronger, and is for this reason preferred. So the sentence, although technically true, should sound pragmatically infelicitous.

Let us now turn to the negative version of this sentence: "Tom didn't touch the rock". In this linguistic context (a 'downward-entailing' environment), it is now the effector meaning which is stronger, and therefore selected. Let us assume again that Tom made the right sort of movement to be the agent of a touching event, but only because he has had a stroke. In this context, Tom as effector touched the rock. What's more, he is a mere effector. This means that "Tom (as a mere effector) didn't touch the rock" is false. Thus the theory predicts that people should tend to think that this sentence is not appropriate either in the given context.

In other words, people should think that there is something misleading about saying "Tom touched the rock" in situations of that type, because this sentence is preferentially interpreted as meaning "Tom (*as an in-control agent*) touched the rock", but they should also think that it is wrong in the same situation to say "Tom didn't touch the

rock", because this sentence is preferentially interpreted as meaning "Tom (*as a mere effector*) didn't touch the rock".

Putting everything together, we can now explain what this hypothesis says about sentences like "Tom caused the train delay." On the account we have been developing, this sentence means, roughly, that there was an event that was a causing of the train delay, and Tom occupied the agent role in this event. If Tom's bodily movements are the result of a stroke, this does not make the event itself be any less of a causing of the train delay. Rather, the important point is that Tom does not occupy a certain role in this event, namely, the role of an in-control agent. Thus, people will feel on the whole that even if the sentence is true, it is not a felicitous way to describe the event that occurred.

12. Agency beyond language

We have been focusing on the idea of the agent role as a way of understanding certain patterns in people's use of language, but a question now arises as to whether our findings are specific to language or whether they reflect something more general about people's cognition. Quite independently of the fact that people use language, it seems that people's cognition involves representations of *events* (e.g., Lee et al., 2024; Yates et al., 2023; Zacks & Swallow, 2007). These representations may then involve the entities playing specific roles in events (e.g., Hafri et al., 2013; Ünal et al., 2024). Thus, if we obtain certain findings about thematic roles in natural language, one intriguing hypothesis would be that these findings actually point to something very general about how people understand roles in events.

In the linguistic theory we introduce in the Appendix, a core claim is that there is a particular role such that (a) human beings are seen as playing this role when they act in the normal way, (b) inanimate objects are seen as playing this role when they move or affect things in the normal way, but (c) human beings are not seen as playing this role when they undergo bodily movements that do not involve the exercise of their own agency (e.g., because they suffered a stroke). A key question for further work will be whether this pattern reveals something specifically about natural language or whether it reveals something broader about people's representations of events.

Existing research provides some intriguing hints about the degree to which these phenomena might or might not extend beyond language. Studies show that people automatically categorize entities as occupying the agent role in events, even in a task for which this role would not be directly relevant (Hafri et al., 2018). Indeed, roles like agent and patient are said to be represented even in visual perception (Hafri & Firestone, 2021). And a broad review finds that although people may not make use of certain sorts of thematic roles (goals, recipients, etc.) outside of language, people do seem to use categories objects as occupying the agent role in events, even in non-linguistic cognition (Rissman & Majid, 2019).

Of course, this existing work does not provide evidence either way about whether the findings of the present studies would extend beyond language to people's ordinary event representations. However, this work does suggest that there is a real question here that would be worth investigating. Independent of the use of language, people do seem to represent entities as occupying roles in events, and thus it remains an open question whether the specific findings of the present studies would arise for those representations.

13. Conclusion and further directions

Existing research finds an effect such that judgments about sentences that say that a person caused an outcome can be impacted by the degree to which the person exercised agency. The present studies provide evidence that this effect is not specific to causal judgments but instead arises for a far broader class of different judgments.

The results thereby suggest that the original effect is pointing to something broader, and hence potentially deeper and more important, than we might initially have expected. Simply, the effect does not appear to be something that arises only when people are thinking about causation. It seems instead to be something that applies quite generally to people's way of thinking about events. Given this, any explanation will necessarily have to connect with some fundamental features of the way people ordinarily understand events. For example, at the heart of the explanation we have developed here are the ideas that (a) people have a very general way of representing different behaviors as displaying different degrees of agency and (b) people can use these representations to assess the degree to which a person plays the agent role in an event.

In the present inquiry, we have merely scratched the surface of these deeper issues, investigating them only insofar as they are relevant to the specific linguistic effect we have been exploring. One exciting avenue for further research would be to take up these issues as phenomena worth exploring in their own right. That is, even independent of the specific linguistic effect we have been trying to explain here, further research could continue to explore questions about how people assess the degree to which a person exercises agency in performing a behavior and about how people determine whether someone plays the agent role in an event.

CRediT authorship contribution statement

Sehrang Joo: Writing – original draft, Methodology, Investigation, Conceptualization. **Sami R. Yousif:** Writing – original draft, Visualization, Methodology, Investigation, Conceptualization. **Fabienne Martin:** Writing – original draft, Investigation, Formal analysis, Conceptualization. **Frank C. Keil:** Writing – review & editing. **Joshua Knobe:** Writing – original draft, Supervision, Methodology, Investigation, Conceptualization.

Acknowledgement

We are grateful to Christopher Piñón, Florian Schäfer, Diego Feinmann, Sam Carter, Malka Rappaport Hovav and Giorgos Spathas as well as our anonymous referees for valuable comments and suggestions, and to Ben Sluckin and Yining Nie for their judgments on English data.

Appendix A. Appendix

A.1. Preliminaries

In the General Discussion, we proposed that the role of agent has a strong and a weaker meaning. In this Appendix, we sketch one technical way to implement this idea.[2]

Our first assumption is that there are two relations between events and individuals behind the concept of agent, 'effector' and 'in-control (participant)':

(1) a. $\lambda x \lambda e$.**effector**(e, x)
(`x is an effector of e').
b. $\lambda x \lambda e$.**in-control**(e, x).
(`x is in control of e').

Being the effector in an event e simply means doing something in e. For inanimate entities (at least those that are not instruments, like in Study 2), just being an effector will be sufficient to count as the strongest possible agent, since the other ways to be agentive are clearly irrelevant for inanimates' agent-membership. But for animates, things are different. If a person does something, she already counts as an agent in some weak sense of the term, but more than doing something is typically required of a person to count as an agent sensu stricto. We propose that

[2] See Martin (2023) for a more detailed version of the semantic analysis summarized in this Appendix.

in order to count as an agent in this stronger sense, a person should not only do something, but also exert agentive control, e.g., walk across a line rather than stumbling across it.

That agentive control may be central for the concept of agency in language has been argued for in linguistic research on the role of agent. For instance, for Dowty (1979:118) or Jacobs (2011), the concept of agent has more to do with the notion of agentive control than with the notions of intentionality or volition. To express this, we introduce **in-control** (see (1b)), a specific notion of agent that defines 'in-control' agency (expressed for instance by control morphologies in Salish languages, see Davis & Matthewson, 2009, Jacobs, 2011), as opposed to 'out-of control' agency (expressed for instance by out-of-control morphologies such as the ability/involuntary action form in Salish or Tagalog, see Davis et al., 2009, Jacobs, 2011, Alonso-Ovalle & Hsieh, 2021).

The two relations in (1a/b) are ordered by strength: 'in-control (participant)' entails 'effector' but not vice versa, as stated in (2).

(2) a. $\forall x \forall e(\textbf{in-control}(e, x) \rightarrow \textbf{effector}(e, x))$
(`If x is in control of e, then x is an effector of e').
b. $\exists x \exists e(\textbf{effector}(e, x) \wedge \neg \textbf{in-control}(e, x))$.
(`There is an effector that isn't in control').

In the linguistic literature, agentive control has been characterized in several related ways: ability of an entity to 'initiate and carry out an event' (Davis, 2000), 'to influence the outcome of an event' (Davis & Matthewson, 2009), or to function with 'usual average capacities in keep things under control' (Thompson, 1985). A non-human entity might potentially be seen as fulfilling these descriptions as long as it directs its behavior towards a certain goal (think about ants or alarm-clocks).[3] We therefore assume that 'in-control (participant)' entails 'goal-oriented', while 'effector' does not, as formulated in (3).

(3) a. $\forall x(\exists e(\textbf{in-control}(e, x)) \rightarrow \textbf{goal-oriented}(x))$
(`If x is in control of e, then x is goal-oriented').
b. $\exists x \exists e(\textbf{effector}(e, x) \wedge \neg \textbf{goal-oriented}(x))$.
(`There is an effector that isn't goal-oriented').

In-control agency is therefore not reachable for inanimate entities such as water or stones (except if they are personified, or perhaps if some person uses them as instruments). This corresponds to the intuition that inanimate agents are characterized by a very low degree of agency.

A.2. Semantics

The question is how these two notions (effector and in-control) characterize the meaning of agent Voice, which is the syntactic piece that introduces an external argument and associates the agent role with it (Kratzer, 1996). The agent Voice phrase is typically not pronounced; it is the part of the sentence that makes it clear that the subject of the sentence is the agent of the event. Agent Voice is part of the semantics of both causal and non-causal sentences like *Tom caused the train delay* or *Tom touched the rock*, where it plays exactly the same role and leads to an effect of agency in exactly the same way.

We propose that $\text{Voice}_{\text{agent}}$ is ambiguous between these two notions. More concretely, $\text{Voice}_{\text{agent}}$ is polysemous, and represented as the relational variable ***agent***, which is valued either as **effector** or as **in-control**:

(4) $\text{Voice}_{\text{agent}} \rightsquigarrow \textbf{\textit{agent}}$, where $\textbf{\textit{agent}} \in \{\lambda x \lambda e.\textbf{effector}(e, x), \lambda x \lambda e.\textbf{in-control}(e, x)\}$.

As an illustration of how an agentive sentence is semantically represented, consider the example in (5a), which has the syntax in (5b) with the agent Voice head, and receives the semantic analysis (ignoring tense) in (5c).

(5) a. Tom touched the stone.
b. [Tom [$\text{Voice}_{\text{agent}}$ [touched the stone]]].
c. $\exists e(\textbf{\textit{agent}}(e, \textbf{tom}) \wedge \textbf{touch-the-stone}(e))$.
(`There is an event e in which Tom as agent touches the stone').

Since ***agent*** is a relational variable that may be valued as **effector** or as **in-control**, the formula in (6c) corresponds to one of the two propositional meanings listed in (6).

(6) (6) a. $\exists e(\textbf{effector}(e, \textbf{tom}) \wedge \textbf{touch-the-stone}(e))$
b. $\exists e(\textbf{in-control}(e, \textbf{tom}) \wedge \textbf{touch-the-stone}(e))$.

We propose that the choice between these two propositions is determined by an application of the Strongest Meaning Hypothesis (Dalrymple et al., 1998:193), which we formulate for the present case as follows:

(7) **Strongest Meaning Hypothesis**: A sentence S with agent Voice can be used felicitously in a context c that supplies non-linguistic information I relevant to agent Voice's interpretation, and in this case, the use of S in c expresses the logically strongest proposition in A_c:

$A_c = \{p \mid p$ is consistent with I and p is an interpretation of S obtained by interpreting agent Voice as one of the two relations in (4)$\}$.

Assuming the Strongest Meaning Hypothesis, in the case of the sentence in (5a), the set A_c would contain the two propositions in (6). Since 'in-control (participant)' entails 'effector' (recall (2a)) and the linguistic context of the sentence in (5a) is upward-entailing, the proposition in (6b) is the strongest meaning, entailing the proposition in (6a).

The default preference, in an upward-entailing context, for the in-control meaning with animate subjects, accounts for why in Studies 1–4, many participants found sentences with an animate subject less natural/appropriate in the low agency condition, where Tom is an out-of-control agent. But the availability of the effector use of 'agent' explains why these sentences are nevertheless accepted in the same scenario by many participants. For instance, (5a) can also be interpreted as the proposition in (6a), where agent Voice is rendered as 'effector'.

In an upward-entailing context, the preference for the in-control meaning arises only in the absence of contrary information from the linguistic context. Thus, if the direct context explicitly indicates that the subject's referent is *not* an in-control agent, the set A_c would contain the propositional meaning (6a) only. We therefore expect the interpreter not to consider the meaning (6b), which is normally chosen in upward-entailing environments. For instance, take sentences like "After he completely lost control of his body, Tom touched the stone" or "Tom accidentally caused the train delay". In such sentences, the *after*-clause or the modifier *accidentally* are in conflict with the in-control meaning of 'agent'. Our prediction for such sentences is that participants would choose the effector meaning of 'agent', and therefore rate such sentences highly even in the low agency condition.

In a linguistic context that is downward-entailing, the choice of effector is the stronger meaning for agent Voice, as shown, for instance, by the negated version of (5a) in (8a), whose semantics and (surface) syntax are given in (8b/c):

(8) a. Tom didn't touch the stone.
b. [Tom [$\text{Voice}_{\text{agent}}$ [didn't [touch the stone]]]].
c. $\neg \exists e(\textbf{agent}(e, \textbf{tom}) \wedge \textbf{touch-the-stone}(e))$.
(There isn't an event e in which Tom as agent touches the stone).

[3] Instrumental inanimate subjects are often agentive in more respects than non-instrumental inanimate subjects (Alexiadou & Schäfer, 2006; Schlesinger, 1989). For instance, an instrument often 'goes proxy' for an intentional, in-control agent (compare, e.g., *My alarm clock woke me up* with *The storm woke me up:* the alarm-clock is agentive in more respects than the storm is, as what the alarm-clock did reflects the agentive control and intention of its user).

In this case, the set A_c would contain the two propositions in (9), where the proposition in (9a) with 'effector' is now logically stronger than the proposition in (9b) with 'in-control (participant)'.

(9) a. $\neg\exists e$(**effector**(e, **tom**) $\wedge$ **touch-the-stone**(e))
b. $\neg\exists e$(**in-control**(e, **tom**) $\wedge$ **touch-the-stone**(e)).

The reasoning is as follows: if it's not the case that Tom as effector touches the stone, then it's also not the case that he as in-control participant touches the stone, but not vice versa (because if it's not the case that Tom as in-control participant touches the stone, then it's still possible that he as effector touches the stone). The existence of an event of the type in question is denied under both meanings (9a/b).[4] Thus, participants would likely all converge in the view that sentences like *Tom didn't touch the stone* are not natural or appropriate in a context where Tom touched the stone, independently of whether they select the stronger meaning in (9a), with the effector meaning of 'agent', or the weaker meaning in (9b), with the in-control meaning of 'agent'.

In this analysis, notions like foreknowledge or intention do not play any role in the semantics of $\text{Voice}_{\text{agent}}$.[5] Foresight and intention might very well be dimensions characterizing intentional agents, but not agents simpliciter as introduced by $\text{Voice}_{\text{agent}}$. This is completely compatible with the very plausible assumption that by default, a hearer tries to make the subject's referent the strongest possible agent that is compatible with the facts and the context, that is, tends to conceive this entity as an intentional agent when possible (see, e.g., van Valin & Wilkins, 1996). But this enrichment is of pragmatic nature and goes beyond the strongest meaning of $\text{Voice}_{\text{agent}}$.

Data availability

Data, materials, and preregistration information can be found on the Open Science Framework (OSF) at https://osf.io/7a6fq/?view_only=08f4cf81aa4647a898ae47789fac81ca

References

Alexiadou, A., Anagnostopoulou, E., & Schäfer, F. (2015). *External arguments in transitivity alternations: A layering approach*. Oxford: Oxford University Press.

Alexiadou, A., & Schäfer, F. (2006). Instrument subjects are agents or causers. In D. Baumer, D. Montero, & M. Scanlon (Eds.), *Proceedings of the west coast conference on formal linguistics (WCCFL) 25* (pp. 40–48). Somerville, MA: Cascadilla Proceedings Project.

Alicke, M. D. (1992). Culpable causation. *Journal of Personality and Social Psychology, 63*, 368–378.

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

Alonso-Ovalle, L., & Hsieh, H. (2021). Causes and expectations: On the interpretation of the Tagalog ability/involuntary action form. *Journal of Semantics, 38*, 441–472.

Baker, C. L., Jara-Ettinger, J., Saxe, R., & Tenenbaum, J. B. (2017). Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour, 1*, 0064.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., … Bolker, M. B. (2015). Package 'lme4'. *Convergence, 12*, 2.

Beaver, D., & Clark, B. (2009). *Sense and sensitivity: How focus determines meaning*. Malden: John Wiley & Sons.

Beavers, J., & Koontz-Garboden, A. (2020). *The roots of verbal meaning*. Oxford: Oxford University Press.

Bickel, B. (2010). Grammatical relations typology. In J. Song (Ed.), *The Oxford handbook of language typology* (pp. 399–414). Oxford: Oxford University Press.

Bickel, B., Witzlack-Makarevich, A., Choudhary, K. K., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2015). The neurophysiology of language processing shapes the evolution of grammar: Evidence from case marking. *PLoS One, 10*(8), Article e0132819.

Bloom, P. (1996). Intention, history, and artifact concepts. *Cognition, 60*, 1–29.

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen-through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior and Development, 21*, 315–330.

Colombatto, C., Chen, Y. C., & Scholl, B. J. (2020). Gaze deflection reveals how gaze cueing is tuned to extract the mind behind the eyes. *Proceedings of the National Academy of Sciences, 117*(33), 19825–19829.

Colombatto, C., Van Buren, B., & Scholl, B. J. (2019). Intentionally distracting: Working memory is disrupted by the perception of other agents attending to you—Even without eye-gaze cues. *Psychonomic Bulletin & Review, 26*, 951–957.

Colombatto, C., van Buren, B., & Scholl, B. J. (2020). Gazing without eyes: A "stare-in-the-crowd" effect induced by simple geometric shapes. *Perception, 49*, 782–792.

Colombatto, C., van Buren, B., & Scholl, B. J. (2021). Hidden intentions: Visual awareness prioritizes perceived attention even without eyes or faces. *Cognition, 217*, Article 104901.

Cruse, D. A. (1973). Some thoughts on agentivity. *Journal of Linguistics, 9*, 11–23.

Cushman, F., & Young, L. (2011). Patterns of moral judgment derive from nonmoral psychological representations. *Cognitive Science, 35*, 1052–1075.

Dalrymple, M., Kanazawa, M., Kim, Y., Mchombo, S., & Peters, S. (1998). Reciprocal expressions and the concept of reciprocity. *Linguistics and Philosophy, 21*, 159–210.

Davis, H. (2000). Salish evidence on the causative-inchoative alternation. In W. Dressler, O. Pfeiffer, M. Pöchtrager, & J. Rennison (Eds.), *Morphological analysis in comparison* (pp. 25–60). Benjamins: Amsterdam & Philadelphia.

Davis, H., & Matthewson, L. (2009). Issues in Salish syntax and semantics. *Lang & Ling Compass, 3*(4), 1097–1166.

Davis, H., Matthewson, L., & Rullmann, H. (2009). 'Out of control' marking as circumstantial modality in St'át'imcets. In L. Hogeweg, A. Malchukov, & H. de Hoop (Eds.), *Cross-linguistic semantics of tense, aspect and modality* (pp. 205–244). Benjamins: Amsterdam & Philadelphia.

DeLancey, S. (1984). Notes on agentivity and causation. *Studies in Language, 8*, 181–213.

Delancey, S. (1985). On active typology and the nature of agentivity. In F. Plank (Ed.), *Relational typology* (pp. 47–60). De Gruyter: Berlin/New York.

Dowty, D. (1979). *Word meaning and Montague grammar*. Dordrecht: Reidel.

Dowty, D. (1989). On the semantic content of the notion of 'thematic role'. In *Properties, types and meaning* (pp. 69–129). Dordrecht: Springer.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language, 67*, 547–619.

Driver, J. (2008). Attributions of causation and moral responsibility. In W. Sinnott-Armstrong (Ed.), *Moral psychology, Vol. 2. The cognitive science of morality: Intuition and diversity* (pp. 423–439). MIT Press.

Dryer, M. (2005). Order of subject, object and verb. In M. Dryer, & M. Haspelmath (Eds.), *The world atlas of language structures online* (pp. 330–333).

Fauconnier, S. (2012). *Constructional effects of involuntary and inanimate agents: A cross-linguistic study*. PhD Thesis. Universiteit Leuven.

Fillmore, C. J. (1967). The case for case. In E. Bach, & R. T. Harms (Eds.), *Universals in linguistic theory*. Holt, Rinehart and Winston.

Folli, R., & Harley, H. (2008). Teleology and animacy in external arguments. *Lingua, 118*, 190–202.

Glass, L. (2023). Using the Anna Karenina principle to explain why cause favors negative-sentiment complements. *Semantics and Pragmatics, 16*(6).

Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. *Universals of Language, 2*, 73–113.

Grimm, S. (2011). Semantics of case. *Morphology, 21*, 515–544.

Hafri, A., & Firestone, C. (2021). The perception of relations. *Trends in Cognitive Sciences, 25*(6), 475–492.

Hafri, A., Papafragou, A., & Trueswell, J. C. (2013). Getting the gist of events: Recognition of two-participant actions from brief displays. *Journal of Experimental Psychology: General, 142*(3), 880–905.

Hafri, A., Trueswell, J. C., & Strickland, B. (2018). Encoding of event roles from visual scenes is rapid, spontaneous, and interacts with higher-level visual processing. *Cognition, 175*, 36–52.

Halpern, J. Y., & Hitchcock, C. (2015). Graded causation and defaults. *The British Journal for the Philosophy of Science., 66*, 413–457.

[4] In principle, negation can also associate with a more specific constituent of a sentence, via focus; this is what is called 'constituent negation'. For instance, *TOM didn't touch the stone, MARY did* (with a focus on 'Tom') takes for granted that there is an event e in which someone as agent touches the stone, and asserts that it is not the case that there is an event e in which Tom as agent touches the stone (see Beaver & Clark, 2009 among others). But constituent negation is only possible when the constituent is overtly pronounced. As we saw earlier, this is not the case of Voice in English, which is silent. For this reason, in *Tom didn't touch the stone*, negation cannot associate with agent Voice in order to negate the stronger 'in-control' meaning and implicate the weaker 'effector' meaning. Another problem is that thematic information not overtly expressed in the morphology is *presupposed*. Thus, to deny thematic information, we would need *metalinguistic* negation. But let's assume, for the sake of the argument, that it is in principle possible to associate negation with agent Voice. In this scenario, the negative sentence *Tom didn't touch the stone* could be used to presuppose that Tom touched the stone as agent, and assert that it is not the case that he did it as an in-control agent. But then, another problem would arise: it would still be very awkward to choose the negative sentence to express this thought rather than the positive version *Tom touched the stone*, for this much simpler option can very well express this very same thought under the weaker, effector meaning of 'agent'.

[5] But these notions are likely to be relevant for certain overt agent-marking morphologies in some languages–such as a specific verbal prefix in Ahcenese (Legate, 2014), or some perfective markers in Tibetan dialects like Lhasa or Newari, which similarly indicate that the event is under the conscious and intentional control of the subject (Delancey, 1985, 52).

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology, 57*, 243–259.
Hitchcock, C. (2012). Portable causal dependence: A tale of consilience. *Philosophy of Science, 79*, 942–951.
Hitchcock, C., & Knobe, J. (2009). Cause and norm. *The Journal of Philosophy, 106*, 587–612.
Icard, T. F., Kominsky, J. F., & Knobe, J. (2017). Normality and actual causal strength. *Cognition, 161*, 80–93.
Jacobs, P. (2011). *Control in Skwxwú7mesh*. PhD Dissertation,. University of British Columbia.
Johnson, S. C. (2000). The recognition of mentalistic agents in infancy. *Trends in Cognitive Sciences, 4*, 22–28.
Keil, F. C., & Newman, G. E. (2015). Order, order everywhere, and only an agent to think: The cognitive compulsion to infer intentional agents. *Mind & Language, 30*, 117–139.
Kirfel, L., & Lagnado, D. (2021a). Causal judgments about atypical actions are influenced by agents' epistemic states. *Cognition, 212*, Article 104721.
Kirfel, L., & Lagnado, D. (2021b). *Changing minds—Epistemic interventions in causal reasoning*. PsyArXiv.
Kittilä, S. (2005). Remarks on involuntary agent constructions. *Word, 56*, 381–419.
Kratzer, A. (1996). Severing the external argument from its verb. In J. Rooryck, & L. Zaring (Eds.), *Phrase structure and the lexicon* (pp. 109–137). Berlin & New York: Springer.
Lagnado, D. A., & Channon, S. (2008). Judgments of cause and blame: The effects of intentionality and foreseeability. *Cognition, 108*, 754–770.
Lee, S. H. Y., Ji, Y., & Papafragou, A. (2024). Signatures of individuation across objects and events. *Journal of Experimental Psychology. General, 153*(8), 1997–2012.
Legate, J. (2014). *Voice and v: Lessons from Acehnese*. Cambridge: MA, MIT Press.
Lenth, R. (2018). Emmeans: Estimated marginal means, aka least-squares means. R Package Version 1.1 https://CRAN.R-project.org/package=emmeans.
Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. Chicago: University of Chicago press.
Levin, B. (1999). Objecthood: An event structure perspective. In *Proceedings of the Chicago Linguistic Society Meeting 35, Part 1, Main session* (pp. 223–247).
Levin, B., & Rappaport Hovav, M. R. (1995). *Unaccusativity: At the syntax-lexical semantics interface* (Vol. 26). Cambridge, MA: MIT Press.
Levshina, N. (2022). *Communicative efficiency*. Cambridge: Cambridge University Press.
Litoiu, A., Ullman, D., Kim, J., & Scassellati, B. (2015, March). Evidence that robots trigger a cheating detector in humans. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (pp. 165–172).
Lombrozo, T. (2010). Causal–explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions. *Cognitive Psychology, 61*, 303–332.
Martin, F. (2018). Time in probabilistic causation: Direct vs. indirect uses of lexical causative verbs. In U. Sauerland, & S. Solt (Eds.), *60–61. Proceedings of Sinn und Bedeutung 22* (pp. 107–124). ZAS Papers in Linguistics.
Martin, F. (2023). *Scaling agents via dimensions*. Unpublished manuscript. Utrecht University.
Martin, F., Schäfer, F., & Piñón, C. (2025). Transitives with inchoative semantics. *Glossa: Journal of General Linguistics, 10*(1).
Massam, D. (2009). The structure of (un)ergatives. In S. Chung, D. Finer, I. Paul, & E. Potsdam (Eds.), *16. Proceedings of AFLA* (pp. 125–135).
Murray, S., Krasich, K., Irving, Z., Nadelhoffer, T., & De Brigard, F. (2023). Mental control and attributions of blame for negligent wrongdoing. *Journal of Experimental Psychology. General, 152*, 120–138.
Murray, S., Murray, E., Stewart, G. W., Sinnott-Armstrong, W., & De Brigard, F. (2019). Responsibility for forgetting. *Philosophical Studies, 176*, 1177–1201.
Newman, G. E., Keil, F. C., Kuhlmeier, V. A., & Wynn, K. (2010). Early understandings of the link between agents and order. *Proceedings of the National Academy of Sciences, 107*, 17140–17145.
Noyes, A., & Dunham, Y. (2017). Mutual intentions as a causal framework for social groups. *Cognition, 162*, 133–142.
O'Neill, K., Henne, P., Bello, P., Pearson, J., & De Brigard, F. (2022). Confidence and gradation in causal judgment. *Cognition, 223*, Article 105036.
Opfer, J. E. (2002). Identifying living and sentient kinds from dynamic information: The case of goal-directed versus aimless autonomous movement in conceptual change. *Cognition, 86*, 97–122.
Phillips, J., & Shaw, A. (2015). Manipulating morality: Third-party intentions alter moral judgments by changing causal reasoning. *Cognitive Science, 39*, 1320–1347.
Poulin-Dubois, D., Lepage, A., & Ferland, D. (1996). Infants' concept of animacy. *Cognitive Development, 11*, 19–36.
Quillien, T. (2020). When do we think that X caused Y? *Cognition, 205*, Article 104410.
Quillien, T., & German, T. C. (2021). A simple definition of 'intentionally'. *Cognition, 214*, Article 104806.
Ramchand, G. (2008). *Verb meaning and the lexicon: A first-phase syntax*. Cambridge: MA, Cambridge University Press.
Rappaport Hovav, M., & Levin, B. (1998). Building verb meanings. In M. Butt, & W. Geuder (Eds.), *The projection of arguments: Lexical and compositional factors* (pp. 97–134). CSLI Publications: Stanford, CA.
Rissman, L., & Majid, A. (2019). Thematic roles: Core knowledge or linguistic construct? *Psychonomic Bulletin & Review, 26*, 1850–1869.
Rose, D. (2017). Folk intuitions of actual causation: A two-pronged debunking explanation. *Philosophical Studies, 174*, 1323–1361.
Rose, D. (2022). Mentalizing objects. In *, 4. Oxford Studies in Experimental Philosophy* (p. 182).
Rose, D., Sievers, E., & Nichols, S. (2021). Cause and burn. *Cognition, 207*.
Samland, J., & Waldmann, M. R. (2016). How prescriptive norms influence causal inferences. *Cognition, 156*, 164–176.
Sauppe, S., Næss, Å., Roversi, G., Meyer, M., Bornkessel-Schlesewsky, I., & Bickel, B. (2023). An agent-first preference in a patient-first language during sentence comprehension. *Cognitive Science, 47*(9).
Schlesinger, I. (1989). Instruments as agents: On the nature of semantic relations. *Journal of Linguistics, 25*, 189–210.
Schwenkler, J., & Sievers, E. (2022). Cause, "cause" and norm. In P. Willemsen, & A. Wiegmann (Eds.), *Advances in experimental philosophy of causation* (pp. 123–144). Bloomsbury Publishing.
Schwenkler, J., & Sytsma, J. (2020). *Reversing the norm effect on causal attributions*. Unpublished manuscript. Florida State University http://philsci-archive.pitt.edu/18220/.
Song, G., & Wolff, P. (2005). Linking perceptual properties to linguistic expressions of causation. In M. Achard, & S. Kemmer (Eds.), *Language, culture, and mind* (pp. 237–250). Stanford: CSLI Publications.
Strickland, B., Fisher, M., Keil, F., & Knobe, J. (2014). Syntax and intentionality: An automatic link between language and theory-of-mind. *Cognition, 133*, 249–261.
Thompson, L. (1985). Control in Salish grammar. In F. Plank (Ed.), *Relational typology* (pp. 391–428). Berlin: Mouton de Gruyter.
Tollan, R. (2018). Unergatives are different: Two types of transitivity in Samoan. *Glossa: A Journal of General Linguistics, 3*(1).
Ünal, E., Wilson, F., Trueswell, J., & Papafragou, A. (2024). Asymmetries in encoding event roles: Evidence from language and cognition. *Cognition, 250*, Article 105868.
van Valin, R., & Wilkins, D. (1996). The case for "effector": Case roles, agents, and agency revisited. In M. Shibatani, & S. Thompson (Eds.), *Grammatical constructions* (pp. 289–322). Oxford University Press: Oxford.
Wolff, P. (2003). Direct causation in the linguistic coding and individuation of causal events. *Cognition, 88*, 1–48.
Wolff, P., Jeon, G. H., & Li, Y. (2009). Causers in English, Korean, and Chinese and the individuation of events. *Language and Cognition, 1*, 167–196.
Woo, B. M., Steckler, C. M., Le, D. T., & Hamlin, J. K. (2017). Social evaluation of intentional, truly accidental, and negligently accidental helpers and harmers by 10-month-old infants. *Cognition, 168*, 154–163.
Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition, 69*, 1–34.
Yates, T. S., Sherman, B. E., & Yousif, S. R. (2023). More than a moment: What does it mean to call something an 'event'? *Psychonomic Bulletin & Review, 30*(6), 2067–2082.
Young, L., & Saxe, R. (2011). When ignorance is no excuse: Different roles for intent across moral domains. *Cognition, 120*, 202–214.
Zacks, J. M., & Swallow, K. M. (2007). Event segmentation. *Current Directions in Psychological Science, 16*, 80–84.
Zúñiga, F., & Kittilä, S. (2019). *Grammatical voice*. Cambridge, MA: Cambridge University Press.