

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

Sehrang Joo (sehrang.joo@yale.edu)^a
Sami R. Yousif (sami.yousif@yale.edu)^a
Fabienne Martin (fabienne.martin@hu-berlin.de)^c
Frank C. Keil (frank.keil@yale.edu)^a
Joshua Knobe (joshua.knobe@yale.edu)^b

^a Department of Psychology, Yale University

^b Program in Cognitive Science and Department of Philosophy, Yale University

^c ERC LeibnizDream Project, Humboldt-Universität zu Berlin

Abstract

Consider a causal claim like “Tom caused the train delay.” Previous research has shown that the extent to which Tom is seen to act intentionally (i.e., through his own agency) affects the extent to which people agree with this claim. But is this effect of perceived agency a unique phenomenon to causal judgments? Two experiments suggest this may *not* be the case. Study 1 finds that perceived agency affects people’s understanding of both causal and non-causal events. Study 2 then finds that while perceptions of agency were similarly involved in people’s understanding of causal and non-causal events, they affected *only* cases where these events were brought about by animate agents (e.g., people). These results thus suggest that perceptions of agency may have a much more general influence in how people understand events involving agents, and therefore in how they understand the sentences that describe them. We discuss implications for causal cognition, broader research in agency, and the intersections between both and linguistics.

Keywords: agency; intentionality; causation; syntax; semantics

Introduction

Imagine a train platform with a line that people aren’t supposed to cross; if they do, incoming trains will automatically stop. Suppose that Tom deliberately steps over the line to stand in front of it, and this ends up causing a train delay. In this case, it seems natural to say:

- (1) Tom caused the train delay.

Now consider, instead, this similar case: Instead of intentionally crossing the line on the train platform, Tom blacks out and falls over it. Just as in the first scenario, Tom is now too near the edge of the platform, and this ultimately leads to a train delay. In this case, however, (1) seems like a much less natural way to describe what has happened.

The extent to which people think that Tom (or any animate agent) caused a particular outcome depends in part on whether or not Tom was exercising his own agency in the way that he affected the outcome (see e.g., Kirfel & Lagnado, 2021a, 2021b; Lagnado & Channon, 2008; Lombrozo, 2010; Rose, 2017; Schwenkler & Sytsma, 2020). Here, it seems straightforward to say that “Tom caused train delay” only if Tom was acting intentionally when he ended up being involved in the train delay.

There are different theories as to what exactly explains this effect of perceived agency, but these theories share a key

perspective: We should seek to understand the role of perceived agency in shaping people’s thinking in order to better understand *causal* judgments. Accordingly, research has discovered many important dimensions to these effects on causal judgments: for instance that perceived agency affects evaluation of causal judgments that involve physical contact and those that don’t (Lombrozo, 2010), that intentional actions are judged to be both more causal and more blameworthy (Lagnado & Channon, 2008), and that reasoning about agents’ mental states may factor into how people identify relevant counterfactuals to causal statements (Kirfel & Lagnado, 2021a, 2021b). On these and related 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 imagine again the scenario in which Tom is waiting behind a line for a train. Consider now the following sentence:

- (2) Tom crossed over the line.

Unlike (1), (2) no longer has any information about causation (the path verb *cross* is not a causative verb; Levin, 1993). But might Tom’s level of agency also affect evaluations of sentences like (2)? If the relevant description of Tom’s actions—as either intentional or virtually without agency—affects people’s judgments of (2) as well as (1), this would suggest that perceived agency is having some effect that extends *beyond* causal cognition.

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 often talk about people acting in many different ways—not all of which are causal. Might these effects of perceived agency arise for people’s understanding of this much larger set of events? If so, then understanding how it is that 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.

Understanding agency and agents

Existing research on agency and causation provides important evidence about the precise way in which information about

agency affects causal judgments. In particular, this research has focused on cases that have two properties.

First, the causal claim is attributing causation to something that would normally be seen as an *agent*. For example, in (1), Tom is a human being, and hence a prototypical example of an animate agent. This distinguishes him from things like a rock or a chair that are clearly not animate and not agents in the same way. (For extensive discussion of how we understand the difference between agents and non-agents, see, e.g., Johnson, 2000; Keil & Newman, 2015; Leslie, 1984; Woodward, 1998.)

Second, the agent is behaving in a way that does not involve *the exercise of agency*. Agents most typically act in a way that is again distinct from how things like rocks typically act; they act by exercising their own agency. Imagining Tom going about his daily life, for instance, likely involves imagining a wealth of ways that he exercises his own agency—e.g., by making choices and acting deliberately. (For extensive discussion of what it means to act with varying degrees of agency, see, e.g., Knobe, 2003; Murray & Lombrozo, 2017; Quillien & German, 2021.)

What is striking, then, about the case where Tom blacks out is that Tom is *not* acting with full agency. Instead, Tom is acting in a way that we wouldn't normally expect from an animate agent.

A question now arises about what people's intuitions will be in cases where causation is being attributed to an inanimate object. Consider a scenario in which water floods the train station:

(3) The water crossed the line.

In many ways, (2) and (3) are quite similar: The same event has occurred, and again with little to no agency being involved. But is (3) in fact *odd* in the same way? While the water clearly must have acted with virtually no agency, this is now completely in line with how we would expect water to normally act. Because the water isn't an animate agent, its low level of agency is not out of the ordinary. In that case, might (3) be an acceptable way to describe the water's behavior, even despite its similarities with (2)?

Previous research in linguistics and cognitive science has identified ways that reasoning about agents (e.g., Rissman & Majid, 2019) and their agency (e.g., Childers & Echols, 2004; Naigles, 1990; Strickland, Fisher, Keil, & Knobe, 2014) is embedded in the language we use to describe events in which such agents are involved. For example, young children assume that a novel subject (but not a novel word in other grammatical roles) is likely an animate agent (e.g., a person like Tom) acting intentionally; Childers & Echols, 2004). In other words, when approaching a sentence that seems to be about an agent, people already have some intuitions about what must likely be involved in order for this sentence to be true—often, that the agent must have acted in a specifically agential or intentional way.

One natural hypothesis would therefore be that the effect

of perceived agency on people's understanding of sentences like (1) are part of these broader ways in which people reason about animate agents. If this hypothesis is on the right track, then we should be explaining the effect of perceived agency *not* specifically through causal cognition, but through different types of theories: namely, the sorts of theories linguists have developed for understanding these sentences.

Current Studies

Across two 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?

In Study 1, we examine the influence of perceived agency on people's evaluations of causal (e.g., (1)) vs. non-causal sentences (e.g., (2)): Does perceived agency have the same effect when people are asked about actions that have nothing to do with causation? Then, in Study 2, we examine the influence of perceived agency on people's evaluations of sentences about animate agents (e.g., (1)) vs. about non-animate entities (e.g., (3)): Is the influence of perceived agency specific to how people understand animate agents?

We discuss the implications of these experiments for causal cognition and for broader questions of agency in language. Finally, we provide a sketch of a linguistic explanation for how agency may be involved in people's thinking across these different sentences.

Study 1

How much agency someone exercised in bringing about an outcome (e.g., whether or not they acted intentionally) 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 causal and non-causal sentences in the same scenarios.

Method

Four hundred adult participants completed a survey online through Prolific. All methods and analyses of this experiment were preregistered; preregistrations for this and following experiments can be found on our OSF page: https://osf.io/teyz2/?view_only=a42583c273e54ba88ae31b493ce489ce. Data from an additional 11 participants were collected but excluded for failing a comprehension check.

Participants were shown one of four short vignettes about a person, Tom, acting either intentionally or with very low agency. For example, in one vignette, 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. In the full agency condition, Tom then deliberately crosses over the line:

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:

In the heat, Tom unexpectedly blacks out and falls over the line.

The same adverse outcome then ensues as a result (e.g., Tom being too close to the edge of the platform ends up resulting in a train delay). See our OSF page for full stimuli of both experiments.

Participants were then asked to evaluate either a causal 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 vignettes, these non-causative verbs included “touch”, “hit”, “cross”, and “enter”; note that all can have an agentive subject, as shown by their compatibility with agent-oriented adverbials such as *carefully*. They 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 without agency (e.g., “Tom blacked out and fell over the line”). Participants who failed the comprehension check were excluded and replaced.

Results and Discussion

Results are displayed in Figure 1. 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). There was a significant main effect of agency, $X^2(1) = 136.4, p < .001$, and a smaller main effect of statement type, $X^2(1) = 10.83, p = .001$. However, there was no significant interaction between agency and statement type, $X^2(1) = .52, p = .47$.

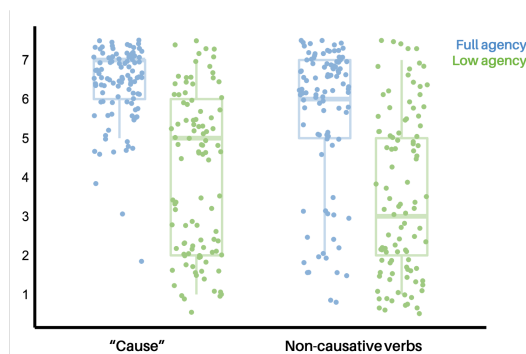


Figure 1: Results from Study 1.

Our primary interest was in whether or not there was a significant effect of agency within each statement type. Participants were significantly more likely to endorse a causal sentence (e.g., “Tom caused the train delay”) when Tom acted intentionally ($M=6.34, SD=.91$) 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-causal sentences (e.g., “Tom

crossed the line”): Participants rated these sentences as more natural when Tom acted intentionally ($M=5.64, SD=1.83$) vs. with very low agency ($M=3.55, SD=2.05$), $t(399)=8.49, p < .001$.

Perceived agency has previously been found to influence people’s causal judgments (see e.g., Kirfel & Lagnado, 2021b; Lombrozo, 2010; Rose, 2017; Schwenkler & Sytsma, 2020), suggesting that reasoning about how much agency was involved is part of how people understand what qualifies as cause of a given outcome. Yet here we find that this phenomenon—the effect of perceived agency—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 that did *not* involve causation at all (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.

Study 2

If perceptions of agency are not specifically involved in people’s understanding of causation, then what may explain when perceived agency factors into people’s understanding of these sentences?

One possibility focuses on the fact that all of the sentences in Study 1 were about Tom—a person who does normally exercise their own agency and who we would typically think of as an animate agent. Could these effects be explained by something about how we think about animate agents? To address this, we introduce sentences with inanimate subjects:

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

(4a) resembles similar causal statements that are also about an animate agent, i.e., Tom (see (2)).¹ But unlike Tom, the water is an inanimate entity that normally lacks the agency characteristic of human agents. Does this lack of agency also influence people’s judgments of relevant sentences?

Method

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

600 new participants completed a survey online through Prolific. 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.

¹With some of our verbs, inanimate subjects can be associated with another thematic role than Agent (e.g., with the role Theme for path verbs). But inanimate subjects *can* also be construed as agents across all verbs, as shown e.g. by the *do*-test (*What the water did was cause the train delay/cross over the line*). See Cruse (1973) and Fauconnier (2012) against the inclusion of animacy among core features of the role Agent.

Participants were shown one of six short vignettes, now about either (1) a person, Tom, acting intentionally, (2) a person, Tom, acting with very low agency, or (3) an inanimate entity (e.g., water from a storm) acting the way inanimates do. Both of the conditions involving a person were closely adapted from the vignettes in Study 1; the only changes were in order to be consistent with the inanimate condition. In all conditions, participants 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 (i.e., crossing a line). For example, in the 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.

Results and Discussion

Results are displayed in Figure 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, $X^2(2) = 114.9$, $p < .001$, and a much smaller effect of statement type, $X^2(2) = 5.12$, $p = .02$. There was again no significant interaction between agency and statement type, $X^2(2) = .72$, $p = .70$.

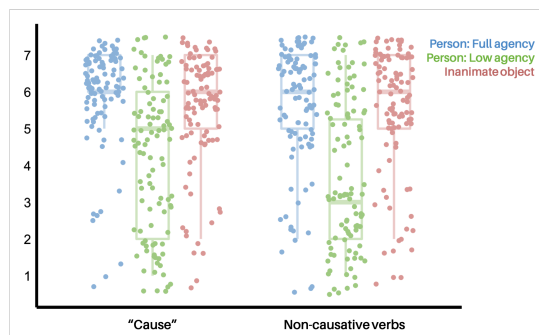


Figure 2: Results from Study 2.

Our main interest 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 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, $t(601) = 1.49$, $p = .30$.

In Study 2, we find that people endorse sentences involving inanimate subjects (e.g., water) just as much as they endorse sentences about animate agents acting intentionally (e.g., Tom deliberately crossing the line), and, again, in the same way across causal and non-causal sentences. Thus the critical factor to understanding the effect of perceived agency seems to lie in the distinction between animate agents and inanimates—and not in the distinction between causal and non-causal events.

General Discussion

Across two experiments, we find that perceptions of agency affect more than people's understanding of causation. Instead, perceptions of agency affected judgments of both causal and non-causal sentences (Study 1). This effect persisted so long as these sentences are describing the actions of animate agents (as opposed to inanimate things; Study 2). We thus suggest that the influence of perceived agency is a more general phenomenon, one which can be applied to a much broader class of statements concerning animate agents.

Agency and Causation

The effect of perceived agency on people's causal judgments has typically been thought to demonstrate a critical, and perhaps surprising, fact about how people reason about *causation* (see, e.g., Kirfel & Lagnado, 2021a). We suggest, however, that this finding from causal cognition is much more general to how people reason and talk about agents. These results thus have important implications, both for the study of causation and for the study of these effects of perceived agency.

First, knowing that agency doesn't uniquely influence causation has implications for how future research may approach the effect of perceived agency. Best understanding these effects now need not involve investigating specific aspects of causal cognition. For example, if we conclude that causal cognition involves counterfactual thinking in a way that certain other kinds of cognition do not, our explanation of the effect of agency should presumably not focus on counterfactual thinking in particular. Instead, the explanation will have to focus on processes that arise for both causal and non-causal judgments.

In finding this explanation, a key first step will be to investigate the boundary conditions of the effect of perceived agency itself, across the broad range of causal and non-causal events where it seems to apply. For example, here we contrast a case where Tom acts fully intentionally and a case

where he falls unconscious (i.e., acting with as little agency as possible)—and in all cases, these actions lead to a negative outcome (e.g., an injury, a train delay, etc.). Are these effects sensitive to other ways Tom could act with less than full agency, such as being pushed by another? And might they arise also in the cases that have no moral valence at all? Future research may address these questions in order to better understand the effects of perceived agency across these varied cases.

Second, these results open up an intriguing new possibility regarding the implications of agency effects for the study of causal cognition. If these effects were specific to causal judgments, then it seems that a satisfactory theory of causal cognition would have to provide an explanation of them. Thus, if we developed a theory of causal cognition that completely failed to predict them, we would have a reason to assume that there was something mistaken or incomplete in that theory. By contrast, if we find that these effects arise for a much broader class of judgments, it would be at least plausible that the explanation of these effects involves a process that lies *outside* the domain of causal cognition. We might therefore consider the possibility that even a complete theory of causal cognition would not include an explanation for these effects, and that they may instead be explained in terms of something else entirely. Below, we will be sketching an explanation along precisely those lines.

In short, if the effect of perceived agency on sentences like “Tom caused the train delay” is best understood as a result of a general tendency for perceptions of agency to affect how people understand events brought about by agents—then future research aiming to better understand either causal judgments or these particular effects need not necessarily be constrained to account for the other.

Agency in Language

Thus far, we have been providing evidence for a general view about the role of agency in people’s evaluation of sentences like “Tom caused the train delay.” We now offer a more specific hypothesis that spells out precisely how this role arises from the syntax and semantics of those sentences. At the core of this hypothesis is the idea that there is an important similarity between the structure of these causal sentences and the structure of non-causal sentences like “Tom crossed over the line.” The impact of agency is then to be explained in terms of the aspect of the structure of these sentences that is shared with the non-causal sentences. In this section, we provide a non-technical overview of the hypothesis; for the technical details, see the Appendix.

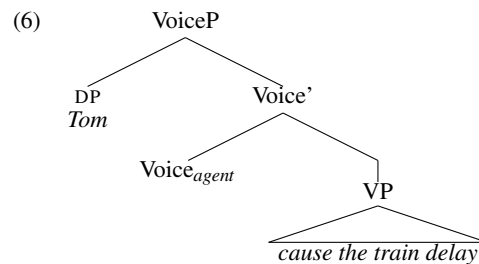
Very broadly speaking, the hypothesis is that a sentence like (5a) has a meaning that can be paraphrased with (5b).

- (5) a. Tom caused the train delay.
b. There was an event which is a causing of the train delay, and Tom is the agent of that event.

Thus, if we want to understand the semantics of a causal

sentence like (5a), one thing we will need to understand is what it means for an event to be a **causing** event, but on the present hypothesis, that aspect of the semantics is not what explains the effect of agency. Instead, that effect arises from a completely different aspect of the semantics of this sentence: namely, from the idea of being the **agent** of an event.

The tree in (6) gives a syntactic structure for this sentence. On this proposal, causation appears in the verb phrase (“cause the train delay”), and a full account of the semantics of that verb phrase would have to involve a deeper account of the semantics of causation (which will presumably involve using ideas from the literature on causal modeling; see, e.g., Kirfel & Lagnado, 2021b). But the role of agency in the sentence is not arising from the verb phrase. Instead, it is arising from a completely different part of the sentence—the agent voice phrase (Kratzer, 1996). Thus, an account of the role of agency in this sentence will not involve further exploring the semantics of the verb cause but instead further exploring the semantics of the agent voice phrase.



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. This agent voice phrase also appears in non-causal sentences like (2) (*Tom crossed over the line*), where it plays exactly the same role and leads to an effect of agency in exactly the same way.

What, then, explains why people treat sentences like “Tom caused the train delay” differently from sentences like “The water caused the train delay”? Our hypothesis is that the agent voice phrase works a lot like a gradable adjective, such as *big*. When people are trying to determine whether an object is big, they do so by checking to see whether the size of the object surpasses a particular threshold, but the threshold obviously depends on what type of object it is. For a planet to be big, it has to surpass the size one might expect for a planet, whereas for a pen to be big, it only has to surpass the size one might expect for a pen (Kennedy, 2013). With that in mind, consider what happens when the agent voice phrase is applied to a human being like Tom vs. to an inanimate object like the water. The hypothesis is that people will say that Tom falls below the threshold to count as a normal agent of an event when his behavior lacks certain qualities that one might expect from the movements of a human being (intention, knowledge, etc.), but they will not use that same standard when faced with a case involving an inanimate object. Instead, when considering the movements of an inanimate object like water, they will ask whether it lacks some of

the qualities one might expect from the movements of water (e.g. effectivity).

Thus our results suggest the following for understanding the way in which perceived agency enters into people's evaluations of sentences about animate agents: First, it matters what kind of subject is involved, and whether it is animate. Second, for those things which are animate (e.g., people), it then matters whether or not they are *acting with enough agency* in this particular event. Finally, and most critically, neither part of reasoning about agents or agency requires anything specific to causation—and so we suggest that the effect of perceived agency is located outside of the verb *cause*.

Conclusion

How is it that reasoning about agency influences how we understand and describe events in the world? Here, we suggest that this question is general to causal and non-causal events, and that the answer may therefore lie beyond causal cognition. The influence of perceptions of agency on our understanding of the world may be a much broader phenomenon, one with rich potential for research across both psychology and linguistics.

Appendix

In the discussion, we proposed that the agent Voice has something to do with degrees. In this appendix, we sketch one technical way to introduce gradability in the semantics of agency, that we take to be a multidimensional concept (Kamp, 1975; Sassoon, 2013). We start with the idea that agents can be ordered according to their closeness to prototypical agents, by counting the key dimensions of agents they satisfy. Thus for instance, if a person crosses the line knowing perfectly this will cause the train delay, she is a full agent satisfying all typical agentive dimensions (intention, knowledge, control, etc). By contrast, if a person accidentally manages to solve a problem, she is not a full agent anymore, but still more agentive than if she inadvertently broke a vase.

We first introduce a (second-order) predicate dimension in order to make reference to a dimension of 'agent':

$$(7) \quad \lambda R.\text{dimension}(R, \lambda x \lambda e.\text{agent}(e, x))$$

(R is a dimension of 'agent')

Next, we adopt a principle identifying critical dimensions of 'agent' (the exact nature of these dimensions does not matter for the analysis; those in (8) are given for the example). Any agent is characterized by at least one dimension of 'agent', that is, is an agent to at least some extent.

$$(8) \quad \forall R(\text{dimension}(R, \lambda x \lambda e.\text{agent}(e, x)) \leftrightarrow$$

$$R = \lambda x \lambda e.\text{intention}(e, x) \vee R = \lambda x \lambda e.\text{control}(e, x) \vee$$

$$R = \lambda x \lambda e.\text{foreknowledge}(e, x) \vee$$

$$R = \lambda x \lambda e.\text{effectivity}(e, x))$$

(The dimensions of 'agent')

In order to be able to specify the number of dimensions of 'agent' present in a given instance, we introduce

a function cardinality for counting the elements of a set: $\lambda \mathcal{R}.\text{cardinality}(\mathcal{R})$. Next, we introduce a function *agential* which is a function from events and individuals to degrees: $\lambda x \lambda e.\text{agential}(e, x)$ (of type $\langle e, \langle s, d \rangle \rangle$). The value of this function for an event e and an individual x is identical to the number of dimensions of 'agent' for e and x (see the counting-dimension function encoded by multidimensional predicates in Sassoon and Fadlon (2017)):

$$(9) \quad \forall e \forall x (\text{agential}(e, x) = d \leftrightarrow$$

$$\text{cardinality}(\lambda R.\text{dimension}(R, \lambda x' \lambda e'.\text{agent}(e', x')) \wedge$$

$$R(e, x)) = d)$$

('agential' for e and x is d iff the number of dimensions of 'agent' for e and x is d)

Since for a choice of e and x , 'agential' could yield zero (in which case x wouldn't be at all agential in e), we define a version of *agential*, that we call *agential*⁺, that is restricted to values of at least 1 for d :

$$(10) \quad \forall e \forall x (\text{agential}^+(e, x) = d \leftrightarrow \text{agential}(e, x) = d \wedge d \geq 1)$$

('agential⁺' for e and x is equivalent to 'agential' for e and x with a degree of at least 1)

Since *agential* and *agential*⁺ are gradable, statements of comparison such as "*agential*⁺(e, tom) > *agential*⁺(e', tom)" are meaningful, for values of e, x, e' , and x' . (In prose, "Tom is more agential⁺ in e than Tom is in e' ".)

Next, we introduce a predicate *agential*st that restricts *agential*⁺ to degrees that are at least as high as the standard degree in some context c :

$$(11) \quad \forall e \forall x (\text{agential}^{\text{st}}(e, x) = d \leftrightarrow \text{agential}^+(e, x) = d \wedge d \geq$$

$$s_c(\text{agential}^+))$$

('agentialst' for e and x is equivalent to 'agential⁺' for e and x with a degree that is at least as high as $s_c(\text{agential}^+)$)

In (11), $s_c(\text{agential}^+)$ denotes the standard degree in context c for 'agential⁺'. Obviously, this standard degree will be different with animate vs. inanimate subjects (since for inanimates, satisfying the single dimension of effectivity will as a rule suffice to make a perfect agent).

Once we have *agential*⁺ and *agential*st at our disposal, they can serve as an alternative analysis of *Voice_{agent}*, encoding a disjunction between these two functions (cf. (12)):

$$(12) \quad \text{Voice}_{\text{agent}} \rightsquigarrow \lambda f_{\langle v, t \rangle} \lambda x \lambda e.\text{agential}^{+/st}(e, x) = d \wedge f(e)$$

In a positive context, we expect the function *agential*st to be preferred to that of *agential*⁺ in (12), for the former yields a stronger meaning (as *agential*st asymmetrically entails *agential*⁺). This default preference for the stronger meaning accounts for why in Studies 1 and 2, sentences with an animate subject are rated less well in the low agency context. But the availability of the weaker meaning also explains why some participants accept these sentences in the same context (see Figures 1 and 2).

References

- Childers, J. B., & Echols, C. H. (2004). 21/2-year-old children use animacy and syntax to learn a new noun. *Infancy*, 5(1), 109–125.
- Cruse, D. (1973). Some thoughts on agentivity. *Journal of Linguistics*, 47.
- Fauconnier, S. (2012). *Constructional effects of involuntary and inanimate agents: A cross-linguistic study*. Unpublished doctoral dissertation, Katholieke Universiteit Leuven, Leuven.
- Johnson, S. C. (2000). The recognition of mentalistic agents in infancy. *Trends in Cognitive Sciences*, 4(1), 22–28.
- Kamp, H. (1975). Two theories about adjectives. In E. Keenan (Ed.), *Formal semantics of natural language* (pp. 123–155). Cambridge, MA: Cambridge University Press.
- 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(2), 117–139.
- Kennedy, C. (2013). *Projecting the adjective: The syntax and semantics of gradability and comparison*. Routledge.
- Kirfel, L., & Lagnado, D. (2021a). Causal judgments about atypical actions are influenced by agents' epistemic states. *Cognition*, 212, 104721.
- Kirfel, L., & Lagnado, D. (2021b). *Changing minds—epistemic interventions in causal reasoning*. (PsyArXiv)
- Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis*, 63(3), 190–194.
- Kratzer, A. (1996). Severing the external argument from its verb. In *Phrase structure and the lexicon* (pp. 109–137). Springer.
- Lagnado, D. A., & Channon, S. (2008). Judgments of cause and blame: The effects of intentionality and foreseeability. *Cognition*, 108(3), 754–770.
- Leslie, A. M. (1984). Infant perception of a manual pick-up event. *British Journal of Developmental Psychology*, 2(1), 19–32.
- Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. University of Chicago press.
- Lombrozo, T. (2010). Causal-explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions. *Cognitive psychology*, 61(4), 303–332.
- Murray, D., & Lombrozo, T. (2017). Effects of manipulation on attributions of causation, free will, and moral responsibility. *Cognitive science*, 41(2), 447–481.
- Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of child language*, 17(2), 357–374.
- Quillien, T., & German, T. C. (2021). A simple definition of 'intentionally'. *Cognition*, 214, 104806.
- Rissman, L., & Majid, A. (2019). Thematic roles: Core knowledge or linguistic construct? *Psychonomic bulletin & review*, 26(6), 1850–1869.
- Rose, D. (2017). Folk intuitions of actual causation: a two-pronged debunking explanation. *Philosophical Studies*, 174(5), 1323–1361.
- Sassoon, G. (2013). A typology of multidimensional adjectives. *Journal of semantics*, 30(3), 335–380.
- Sassoon, G., & Fadlon, J. (2017). The role of dimensions in classification under predicates predicts their status in degree constructions. *Glossa: a journal of general linguistics*, 2(1).
- Schwenkler, J., & Sytsma, J. (2020). *Reversing the norm effect on causal attributions*.
- Strickland, B., Fisher, M., Keil, F., & Knobe, J. (2014). Syntax and intentionality: An automatic link between language and theory-of-mind. *Cognition*, 133(1), 249–261.
- Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, 69(1), 1–34.