Supplementary Information for "A simple definition of 'Intentionally'"

1. How normality interacts with causal structure. An informal explanation in the context of Study 5

Here we give an informal explanation for why the models by Quillien (2020) and Icard et al. (2017) predict the effects that they do in the context of the scenario used in Study 5 (see the original publications for technical details).

The models assume that people generate counterfactuals as a function of their normality. Therefore, they assume that when computing actual causal strength, people will mostly generate counterfactuals where Mr A votes Yes, and Mr B votes No.

- In the conjunctive structure, given that Mr B votes No, manipulating Mr A's vote does not change the outcome: regardless of whether Mr A votes Yes or No, the request is denied. Therefore (across most counterfactuals that people generate) Mr A's vote does not make a difference to the outcome.

    By contrast, given that Mr A votes Yes, Mr B's vote makes a difference to the outcome: if he votes No, the request is denied, while if he votes Yes the request is accepted. Therefore the models predict that in the conjunctive structure, Mr B is more causal than Mr A.

    -In the disjunctive structure, given that Mr B votes No, Mr A's vote makes a difference for the outcome: the request is accepted if and only if Mr A votes Yes.

    By contrast, given that Mr A votes Yes, Mr B's vote does not make a difference to the outcome: the request is accepted regardless of whether Mr B votes Yes or No. Therefore for most counterfactuals that people will generate, manipulating Mr A's vote, but not Mr

B's vote, will change the outcome. Therefore the models predict that in the disjunctive structure, Mr A is more causal than Mr B.

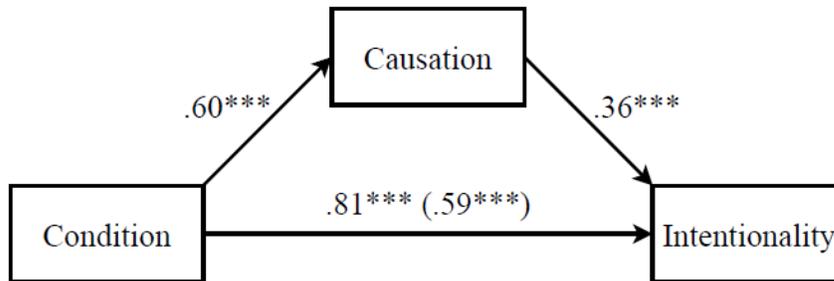## 2. Mediation analyses for studies 2 and 4

Figure S1: Mediation diagram for study 2 (Deviant vs Normal causal link). Standardized regression coefficients for the relationship between experimental condition and intentionality ratings, as mediated by causation ratings. In parentheses is the standardized coefficient for intentionality as predicted by experimental condition, controlling for causation. Experimental condition is coded as 0 = Deviant Link, 1 = Normal Link. The indirect effect for the mediation model was $\beta = .22$***.
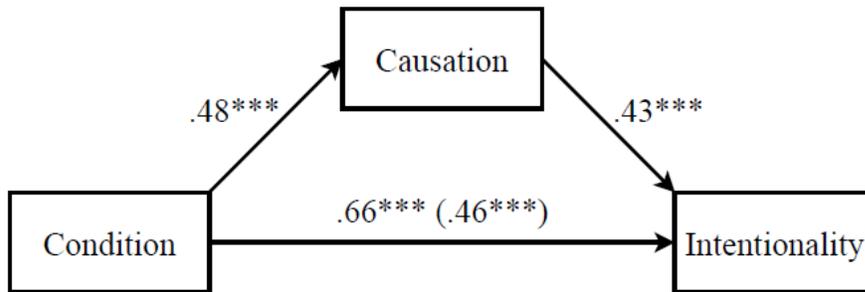
Figure S2: Mediation diagram for study 4 (Causal side-effect effect). Standardized regression coefficients for the relationship between experimental condition and intentionality ratings, as mediated by causation ratings. In parentheses is the standardized coefficient for intentionality as predicted by experimental condition, controlling for causation. Experimental condition is coded as 0 = Help, 1 = Harm. The indirect effect for the mediation model was $\beta = .21^{***}$.

3. Wording and Order effects in Study 5

We conducted the following Anovas.

First, we conducted a Wording * Agent * Structure mixed Anova on causation ratings. There was a 2-way interaction between Wording and Agent, $F(1,127) = 6.33$, $p = .01$, such that, averaging across causal structures, the norm-violating agent was rated as more causal than the norm-conforming agent, but only for the "Caused" wording of the causation question (see Figure S3). No other interaction or main effect involving Wording was significant, all $Fs < .22$, all $ps > .64$.
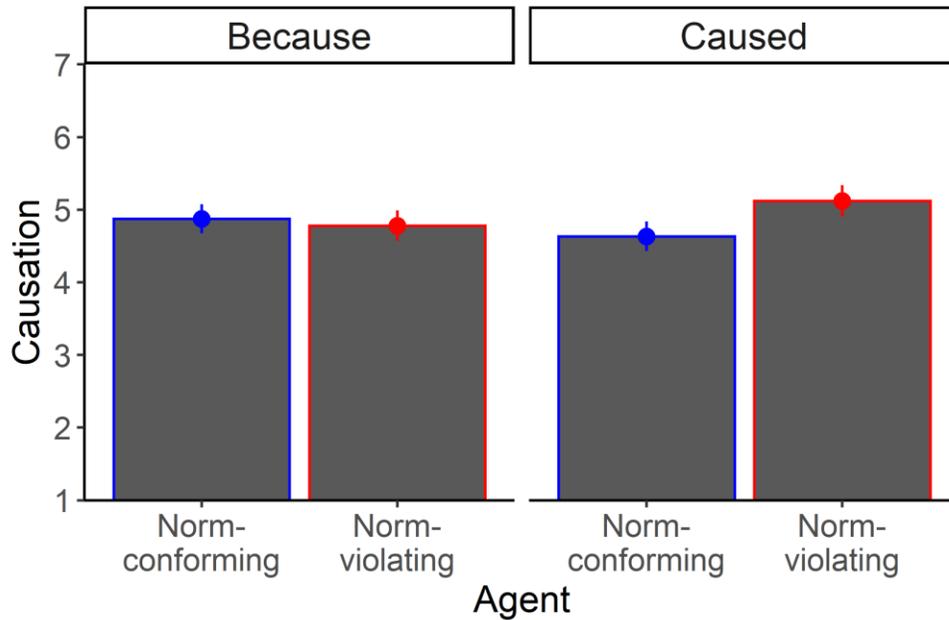
Figure S3. Interaction between Agent and Wording, for the Causation question

Second, we conducted three 2*2*2 Anovas, with Order, Agent and Structure as predictors, and intentionality, causation, and inference as outcome variables, respectively.

We only found significant effects involving Order for the Anova involving the causation question. Specifically, there was an Agent*Structure*Order 3-way interaction, $F(1,127) = 4.60$, p = .03, such that in the disjunctive causal structure, the norm-violating agent was seen as less causal than the norm-conforming agent only when the norm-violating agent was presented last (see Figure S4). There was also an Agent * Order 2-way interaction, $F(1,127) = 9.63$. p = .002, such that, averaged across causal structures, the norm-violating agent was rated as more causal than the norm-conforming agent, but only when the norm-conforming agent was presented first (see Figure S5). There was no main effect of Order, $F(1.127) = .22$, p = .64.

The two other Anovas (for intentionality and for inference) failed to show any significant main or interaction effects involving Order, all Fs < 1.45, all ps > .23.
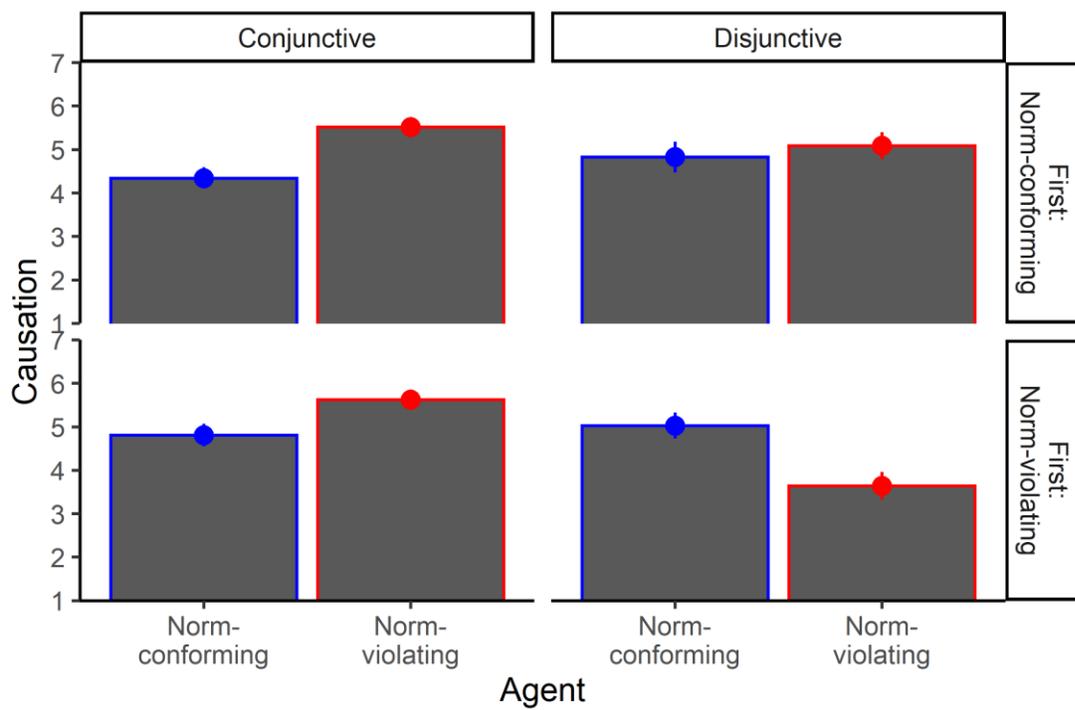
Figure S4. Interaction between Agent, Order and Structure, for the Causation question.
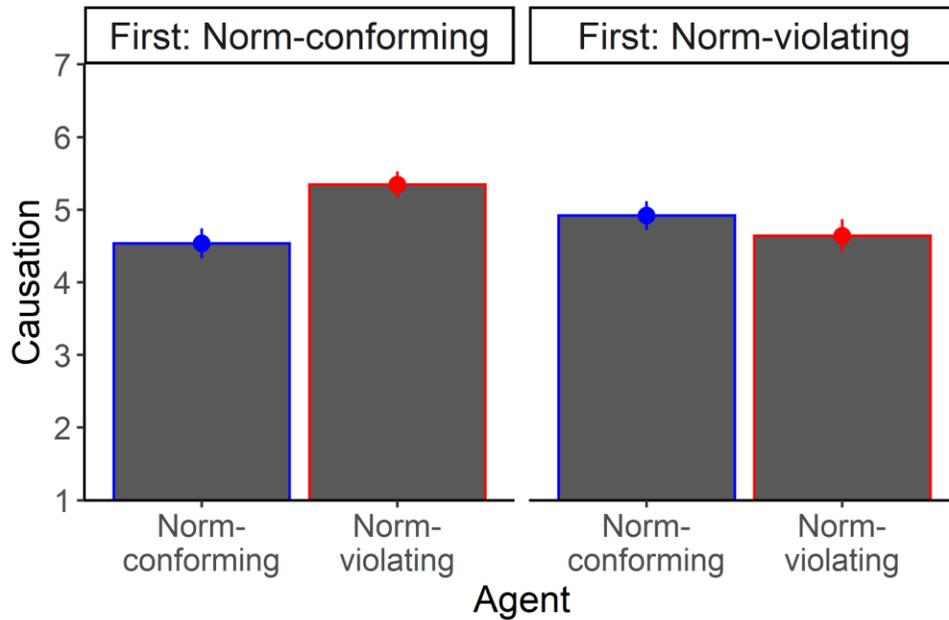
Figure S5. Interaction between Agent and Order, for the Causation question.

4. Study 5B

In study 5, we asked participants to indicate how much they agree that the agent intentionally gave new computers to Professor Smith. A reviewer suggested that the effects we found may have been driven by variation in how much people agree that the agent *gave* new computers to Professor Smith. If we assume that participants interpreted "give" as meaning something like "cause to have", this raises the possibility that effects of normality on intentionality ratings had nothing to do with intentionality. Instead, they would be explained by variation in intuitions about whether the agent's decision caused Professor Smith to get new computers.

This alternative explanation would require that participants interpret "give" in a purely mechanistic sense, devoid of considerations of intentionality. However, research suggests that causative verbs, such as "kill", "bend", or "break", are more complex than, for example, "cause

to die" (Fodor, 1970). In particular, many causative verbs imply intentionality (DeLancey, 1983; Schlesinger, 1989; Wolff, 2003; Pinker, 2007), and "give" seems to be one of them (Newman, 1996). Consider, for instance, that if Alice throws a piece of cake in the trash, where it later gets retrieved by a hungry Bob, Alice did not thereby *give* a piece of cake to Bob.

Nonetheless, it is possible that in the specific context of the vignette we used, people interpreted the verb "to give" in the purely mechanistic sense of "cause to have", with no implications of intentionality. In order to rule out this possibility, we designed a vignette that was similar to the one used in study 5, but where Mr A and Mr B did not intentionally cause Professor Smith to get new computers. We asked whether each agent gave new computers to Professor Smith, and whether their decision caused Professor Smith to get new computers. If participants tend to interpret "give" in a purely mechanistic causal sense, then their agreement with the "give" statement should closely match their agreement with the "cause" statement. If intentionality is an important component of the semantics of giving in the current scenario, then agreement with the "give" statement should be lower than agreement with the "cause" statement.

Participants.

Sample size was determined with the goal of having 99% power to detect an effect of the same size as the abnormal inflation for intentionality ratings in study 5. Therefore, we planned to recruit 150 participants in each between-subject condition. Anticipating an exclusion rate similar to that of study 5 (about 33%), we decided on a target sample size of 450 participants.

In total, we recruited 452 US residents on Prolific. We excluded from analysis 99 participants who failed a catch item (typing the answer to "what is 12-4?") or either one of three

comprehension questions (see below), yielding a final sample of 353 participants (186 female, 2 other).

Stimuli and Procedure.

We used a 2 (Question) * 2 (Normality) mixed design, with Question manipulated Between-subjects, and Normality manipulated within-subjects.

Participants read the following vignette:

At an accounting firm, a committee is in charge of auditing the financial situation of other institutions. The committee must issue an opinion about whether the institution they are evaluating is in good financial situation. The committee has two members, Mr A and Mr B. In order for the committee to give a positive evaluation, it must be that both committee members vote Yes. While Mr A almost always votes Yes, Mr B is notorious for almost always voting No.

Today, the committee is examining the financial situation of an institution. The committee members do not know the name of the institution they are evaluating, or even which kind of institution it is. All they see is a long list of numbers. Both committee members examine the numbers carefully. Then, they both cast their vote at the same time. As usual, Mr A voted Yes; surprisingly, Mr B also voted Yes. Since both committee members voted Yes, the accounting firm declares that the institution is in good financial situation.

Unbeknownst to the committee members, the institution they evaluated is the local university. The dean asked for an audit because he was not sure whether he could afford to give research funding for professors this year. After he receives the result of the audit, he decides to give

research funding to the professors who requested it. One of them is Professor Smith, who gets the funding she requested to get new computers.

Participants in the Cause condition were asked two causation questions (one for each committee member), followed by two inference questions. Participants in the Give condition were asked two give questions, followed by two inference questions. The order in which committee members appeared in the questions was randomized across participants but fixed within-participant. Depending on condition, participants were asked how much they agreed with the following statements, on a 1-7 likert scale (1: strongly disagree, 7: strongly agree):

- Professor Smith got new computers because [Mr A / Mr B] voted Yes.

- [Mr A /Mr B] gave new computers to Professor Smith.

- The fact that [Mr A / Mr B] voted Yes tells us that it was important for him that Professor Smith get new computers.

Additionally, participants were asked the following three comprehension questions, on the same page just below the Cause or the Give question:

- In order for a request to be approved, how many committee members need to vote Yes? (One / Two / Three / Impossible to tell)

- Both committee members usually reach the same decision most of the time. (True / False / Impossible to tell)

- The committee members know they are evaluating a university (True / False / Impossible to tell)

Participants who failed either question were excluded from analysis.

Results

Causation ratings were much higher than Give ratings. In addition, we only found evidence of abnormal inflation for Causation ratings (see Figure 10).
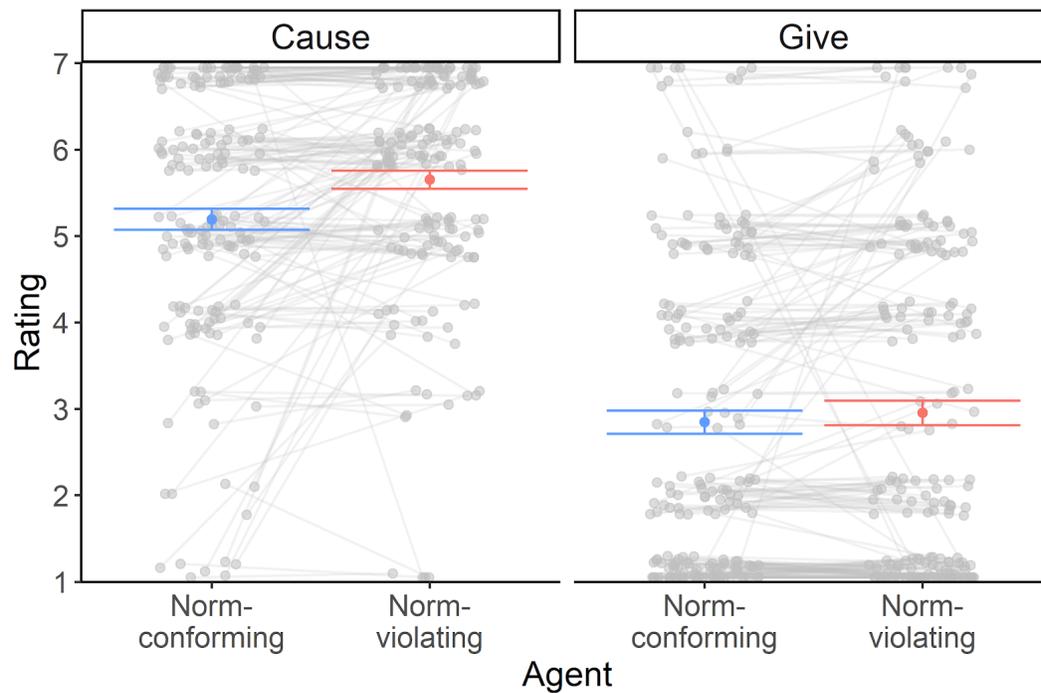

Figure S6. Causation and Give ratings for norm-conforming (Mr A) and norm-violating (Mr B) agents. Error bars represent standard errors of the mean. Individual data points are jittered for better visibility.

A 2*2 mixed Anova revealed a main effect of Question type, $F(1, 351) = 219.1$, $p < .001$, such that participants gave higher ratings to the Causation question ($M = 5.44$, $SD = 1.49$) than the Give question ($M = 2.88$, $SD = 1.93$).

Participants gave higher Causation ratings for the norm-violating agent (M = 5.67, SD = 1.36) than for the norm-conforming agent (M = 5.20, SD = 1.58), t(161) = 4.14, p < .001, d_z = .33. By contrast, participants gave similar Give ratings for the norm-violating agent (M = 2.94, SD = 1.99) and the norm-conforming agent (M = 2.83, SD = 1.88), t(190) = 1.26, p = .21. There was a significant interaction between Question type and Normality, such that the abnormal inflation effect was stronger for Causation compared to Give ratings, F(1, 351) = 6.88, p = .009, $\eta^2_{partial}$ = .02).

Seventy-eight percent of inference ratings were below the midpoint of the scale, confirming that most participants understood that the agent's decisions cannot tell us anything about their attitude toward Professor Smith. Analyzing only data from the participants (N=204) who gave inference ratings of 1 (the lowest point on the scale) does not qualitatively change any of the above results.

Discussion

In a scenario very similar to the one used in study 5, but where the agents cause Professor Smith to have computers in a non-intentional manner, participants are reluctant to say of an agent that he *gave* computers to Professor Smith. This suggests that participants in study 5 considered intentionality to be an important component of the meaning of "give". Therefore, even assuming that their answers to the intentionality question were entirely driven by their judgment of whether the agent *gave* the computers, these ratings would still arguably reflect their concept of intentionality. In addition, we found no evidence of an abnormal inflation effect for Give ratings, which suggests that in the absence of intentionality, people may not engage in computations of causal strength when they judge whether someone gave something (alternatively, the absence of abnormal inflation could be due to a floor effect).

References

DeLancey, S. (1983). Agentivity and causation: data from Newari. *Proceedings of the 9th annual meeting of the Berkeley Linguistics Society* (pp. 54–63). Berkeley, CA: Berkeley Linguistics Society.

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

Icard, T.F., Kominsky, J.F., & Knobe, J. (2017). Normality and actual causal strength. *Cognition*, *161*, 80-93

Newman, J. (1996). *Give: A cognitive linguistic study*. Walter de Gruyter.

Pinker, S. (2007). *The stuff of thought: Language as a window into human nature*. Penguin.

Quillien, T. (2020). When do we think that X caused Y?. *Cognition. 205.*

Schlesinger, I. M. (1989). Instruments as agents: on the nature of semantic relations. *Journal of Linguistics, 25*, 189–210.

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