How, whether, why: Causal judgments as counterfactual contrasts

Tobias Gerstenberg[1] (tger@mit.edu), Noah D. Goodman[2] (ngoodman@stanford.edu), David A. Lagnado[3] (d.lagnado@ucl.ac.uk) & Joshua B. Tenenbaum[1] (jbt@mit.edu)

[1]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139

[2]Department of Psychology, Stanford University, Stanford, CA 94305

[3]Experimental Psychology, University College London, London WC1H 0AP

Abstract

How do people make causal judgments? Here, we propose a *counterfactual simulation model* (CSM) of causal judgment that unifies different views on causation. The CSM predicts that people's causal judgments are influenced by whether a candidate cause made a difference to *whether* the outcome occurred as well as to *how* it occurred. We show how *whether-causation* and *how-causation* can be implemented in terms of different counterfactual contrasts defined over the same intuitive generative model of the domain. We test the model in an intuitive physics domain where people make judgments about colliding billiard balls. Experiment 1 shows that participants' counterfactual judgments about what would have happened if one of the balls had been removed, are well-explained by an approximately Newtonian model of physics. In Experiment 2, participants judged to what extent two balls were causally responsible for a third ball going through a gate or missing the gate. As predicted by the CSM, participants' judgments increased with their belief that a ball was a *whether-cause*, a *how-cause*, as well as *sufficient* for bringing about the outcome.

Keywords: causality; counterfactuals; mental simulation; intuitive physics.

Introduction

How do people make causal judgments? What role do counterfactual thoughts about what might have been play? Philosophers have proposed many different frameworks for thinking about causality. Some have argued that causation is fundamentally about dependence – what it means for C to have caused E is that E did somehow depend on C (Lewis, 2000; Woodward, 2003). Others maintain that causation is about processes: C caused E if there was a (physical) process that started with C and produced E (Dowe, 2000). Still others argue for a pluralistic view and point to two or more different concepts of causation (Beebee, Hitchcock, & Menzies, 2009; Hall, 2004). If one takes a look at the empirical evidence, one also gets the impression that people's causal judgments are very much a mixed bag. Some studies find that people's judgments are strongly influenced by information about the exact way in which C brought about E (Lombrozo, 2010; Walsh & Sloman, 2011; Wolff, 2007), whereas others find that people mostly care about counterfactual dependence – whether E would still have happened if C had been absent (Chang, 2009; Gerstenberg, Goodman, Lagnado, & Tenenbaum, 2012, 2014).

Our aim in this paper is to unify these different views. We argue that the core notion that underlies people's causal judgments is that of difference-making. However, there are several ways in which a cause can make a difference to the effect. It can make a difference to *whether* the effect occurred, and it can make a difference to *how* the effect occurred (cf. Lewis, 2000). While dependence-theories traditionally focus on the first type of difference-making, process-theories highlight the second type. Here, we propose a model that combines these different views by showing how each type of difference-making can be captured in terms of different counterfactual contrasts defined over the same generative model of the domain (cf. Schaffer, 2005; Woodward, 2011).

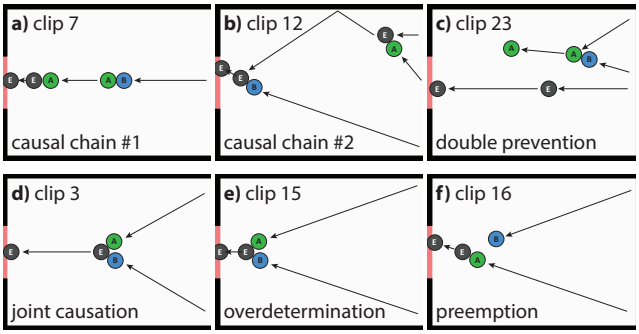


Figure 1: Diagrams of a selection of clips shown in the experiment.

Model

The *counterfactual simulation model* (CSM) explains people's causal judgments in terms of counterfactual contrasts operating over an intuitive domain theory. Here, we illustrate the workings of the model by focusing on people's intuitive understanding of physics. In previous work, we have shown that people's causal judgments in situations which feature a single collision event, are well-captured by assuming that they compare what actually happened, with what they think would have happened if the candidate cause had been removed from the scene (Gerstenberg et al., 2012, 2014). We now extend the CSM to handle more complex situations that involve the interaction of several candidate causes. We will see that people's causal judgments in these more complex cases can be explained if we assume that people consider different kinds of counterfactual contrasts.

Let us illustrate these different contrasts via the example of a simple causal chain (see Figure 1a). Ball E and ball A are initially at rest. Ball B then enters the scene from the right, hits ball A which subsequently hits ball E, and E goes through the gate. To what extent are balls A and B responsible for E's going through the gate?

Whether-dependence First, we may consider what would have happened if either ball had been removed from the scene. That is, we assess whether each ball's presence made a difference to whether or not ball E went through the gate. Formally, we define the probability that a candidate cause C was a *whether-cause* of a target event e as

$$P_W(C,e) = P(e' \neq e | S, remove(C)). \quad (1)$$

We first condition on what actually happened in the situation S – whether ball E went through the gate, the movements of the candidate cause balls, as well as the positioning of the walls and the gate. We then consider a counterfactual situation in which the candidate cause had been removed from the scene ($remove(C)$), and evaluate the probability that the outcome would have been different from what it actually was ($e' \neq e$). The more certain we are that the outcome event would have been different, the greater our subjective degree of belief that C was a *whether-cause*.

In the causal chain, ball B is a *whether-cause* of E's going through the gate. If B had been removed from the scene, ball E would have just remained at rest in front of the gate (see Figure 2a). Ball A, in contrast, is not a *whether-cause*. If A had been removed, ball E would still have gone through the gate – it would have been knocked in by B.

How-dependence If *whether-causation* was all that mattered then ball A shouldn't receive any responsibility at all for E's going through the gate. However, there clearly is a sense in which ball A made a difference to the outcome. For one, it was ball A that actually knocked E through the gate – there was direct transfer of force from A to E. So while ball A didn't make a difference to *whether* E went through the gate, it clearly made a difference to *how* E went through the gate. We define the probability of *how-causation* as

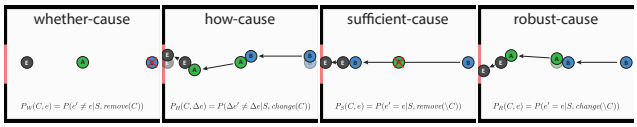
$$P_H(C, \Delta e) = P(\Delta e' \neq \Delta e | S, change(C)). \quad (2)$$

Again, we first condition on what actually happened (S). Now, we consider a different kind of counterfactual contrast. Rather than imagining what would have happened if the candidate cause had been removed from the scene, we simulate what would have happened if the cause had been somewhat different ($change(C)$). In our domain of colliding billiard balls, we may think of the *change* operation as a small perturbation applied to a ball's spatial position.[1] We then assess whether the outcome event would have been different from how it actually was ($\Delta e' \neq \Delta e$).

Note that there is an important difference in how the outcome event is construed depending on whether we assess P_W or P_H. For P_W, we construe the outcome event (e) *broadly*: did E go through the gate or did it not. For P_H, in contrast, we construe the outcome event (Δe) *finely*: exactly how, where, and when did E go through the gate. In the causal chain, ball B does qualify as a how-cause of E's going through the gate (see Figure 2b). If B's spatial location had been somewhat different, E would have gone through the gate differently from how it actually did. For the same reason, Ball A also qualifies as a *how-cause* of E's going through the gate.

So far, the CSM has two components: *whether-causation* and *how-causation*. In our running example, ball B is both a *whether-cause* and a *how-cause*, whereas ball A is only a *how-cause*. The CSM thus predicts that B will be judged more responsible for the outcome than A.

[1] We could also consider changes to C's velocity, direction of motion, or mass. What sorts of changes are relevant will be dictated by people's intuitive understanding of what factors might make a difference (a change in color, for example, is unlikely to be considered).



(a) (b) (c) (d)

Figure 2: Illustration of the different types of counterfactual contrasts applied to ball B.

Sufficiency Our test for *whether-causation* captures whether the candidate cause's presence was necessary for the outcome to occur. It has also been argued that sufficiency is an important aspect of causation: we prefer causes that bring about the outcome without requiring any other causes (Woodward, 2006). Again, we define sufficiency in terms of a counterfactual contrast. The probability that a candidate cause was sufficient for bringing about the outcome is

$$P_S(C, e) = P(e' = e | S, remove(\backslash C)). \quad (3)$$

After having conditioned on what happened (S), we assess the probability that the same outcome (broadly construed) would still have happened ($e' = e$) in a counterfactual situation in which we removed all other candidate causes apart from the cause under consideration ($remove(\backslash C)$).[2]

In the causal chain, ball B was a sufficient cause of E's going through the gate (see Figure 2c). Even if we had removed the other candidate cause (ball A) from the scene, ball E would still have gone through the gate. Ball A, in contrast, was not sufficient. If B had been removed from the scene, E would not have gone through the gate.

Robustness Finally, it has also been argued that people's causal judgments are influenced by robustness (Lewis, 1986; Lombrozo, 2010; Woodward, 2006). Causal relationships are robust to the extent that they would have continued to hold even if the conditions had been somewhat different. We define the probability that a cause was *robust* as

$$P_R(C, e) = P(e' = e | S, change(\backslash C)). \quad (4)$$

After having observed what actually happened (S), we consider whether the same outcome (broadly construed) would still have happened ($e' = e$) even if the other candidate causes had been somewhat different ($change(\backslash C)$). In the causal chain, the robustness of each of the candidate causes is somewhat compromised. Considering the robustness of ball B, there is a good chance that E would not have gone through the gate in a counterfactual situation in which ball A's position was changed. More generally, the longer a causal chain, the less robust each of the candidate causes becomes. A small perturbation to any one of the balls would be sufficient for the chain to fail.

[2] What objects are included in the set of candidate causes is an empirical question. In our experiments reported below, we stipulate the set of candidate causes. Generally, people might have different intuitions about which causes are worth considering (e.g. whether or not the wall should be included in the set).

Putting it all together

Now we have all the puzzle pieces we need. The CSM predicts that 'good' causes are *whether-causes* and *how-causes* that are *sufficient* for bringing about the outcome in a *robust* way. The prototypical case which meets all of these requirements is the Michottean launching event (Michotte, 1946/1963). In a launching clip, ball E is initially at rest, ball C enters the scene and collides with ball E, causing it to move. C is a *whether-cause* of E's motion (E wouldn't have moved if C had been removed from the scene); C is a *how-cause* of E's motion (E would have moved differently if C's initial position had been changed); C is a *sufficient* cause of E's motion (in this case, sufficiency is trivial as there are no alternative candidate causes that could be removed); and C is a *robust* cause of E's motion (again, there are no alternative causes whose position could have been somewhat different).

With all the puzzle pieces in our hands, we still need to say something about how to put them together. Before applying the different counterfactual contrasts, we need to determine the set of candidate causes in a particular situation. We only consider causes that actually made a difference. We define the probability that a cause made a difference as

$$P_{DM}(C, \Delta e) = P(\Delta e' \neq \Delta e | S, remove(C)). \quad (5)$$

A cause made a difference if the outcome event (finely construed) would have been different had the candidate cause been removed from the scene. The CSM predicts that only if our subjective degree of belief is high that the cause made a difference, do we continue to consider the other aspects of causation (P_W, P_H, P_S, and P_R).

Before discussing the experimental tests of the model, let us briefly illustrate the CSM's predictions by applying it to some example cases. Figure 1c depicts a 'double prevention' scenario. E is headed toward the gate. Ball A threatens to knock E off the path. Ball B, however, knocks ball A out of the way. B thus prevents A from preventing E (hence 'double prevention'). Let us focus on ball B's causal status: B made a difference in the actual situation. B is a *whether-cause* (E would not have gone through the gate if B had been removed) but not a *how-cause* (E would have gone through the gate in exactly the same way even if we had perturbed B's initial position somewhat). B was *sufficient* for E's going through the gate (E would have gone through the gate even if ball A had been removed), and B was a *robust* cause (E would most likely have gone through the gate even if A's position had been changed).[3]

Finally, let's consider the 'preemption case' shown in Figure 1f. Here E is at rest in front of the gate and ball A knocks E through the gate shortly before ball B would have done the same (hence, A preempts B from unleashing its causal power). Ball A made a difference according to Eq. 5 whereas ball B did not. Ball A was not a *whether-cause* since E would have gone through the gate even if ball A had been removed. Ball A does qualify as a *how-cause* though, and it was also *sufficient* and *robust*.

[3]Note that whether B qualifies as being sufficient and robust depends on whether we include latent causes, such as the cause of E's motion, in the set of candidate causes. If we included the cause of E's motion in the set, then B would not be sufficient and its robustness would be lower.

Experiment 1: Counterfactual judgments

The CSM assumes that people consider different counterfactual contrasts when making causal judgments. To test the plausibility of this assumption, we first need to make sure that people are capable of simulating the relevant counterfactuals. In this experiment, we directly asked participants to make counterfactual judgments about whether ball E would have gone through the gate if either ball A or ball B had been removed from the scene.

Modeling counterfactual judgments

We model people's counterfactual judgments by assuming that their intuitive understanding of the domain approximately follows the laws of Newtonian physics. This 'Noisy Newtons' approach has been applied successfully in a range of situations (e.g. Battaglia, Hamrick, & Tenenbaum, 2013; Sanborn, Mansinghka, & Griffiths, 2013) and we have shown in previous work that people's counterfactual judgments for simple collision cases are well-explained within this framework (Gerstenberg et al., 2012, 2014).

In order to predict participants' counterfactual judgments, we use the same physics engine that was used to generate the actual clips and generate counterfactual situations by simply removing the candidate ball from the scene. Whether or not ball E would have gone through the gate in this situation follows deterministically from each ball's initial position and velocity. However, people don't have direct access to the outcome in the counterfactual world. They need to make use of their intuitive physical theory to mentally simulate what would have happened.

We capture people's uncertainty by introducing noise in the counterfactual simulation from the point at which the candidate ball would have collided with one of the other balls (cf. Smith & Vul, 2013). At each time step in the simulation, we introduce a small random perturbation drawn from a Gaussian distribution to the other ball's velocity. For example, consider the *causal chain #2* shown in Figure 1b. Would ball E have gone through the gate if ball A had not been present in the scene? To simulate people's judgments, we remove ball A from the scene, and add noise to ball E's velocity from the point at which balls A and E would have collided. In this case, the chances that E ends up going through the gate in the noisy simulation are low (it would only go through if the noise happened to sufficiently perturb ball E's velocity down towards the gate). To simulate the counterfactual of whether E would have gone through the gate if B had not been present in the scene, we remove B from the scene and introduce noise to E's velocity at the time at which B and E would have collided. Here the chances that E would have gone through the gate is high. Since B only collides with E shortly before it entered the gate, a very large perturbation to E's velocity would

be required to prevent E from going through the gate.

We fit the Noisy Newton model to people's judgments by finding the value for the noise parameter which leads to the highest correlation with people's judgments. The noise parameter refers to the value of the standard deviation (SD) of the Gaussian distribution from which the random perturbations to a ball's velocity are drawn. The greater the SD the more noise is introduced into the simulations. For each level of SD, we generate two samples of 1000 noisy worlds: one in which ball A is removed, and one in which ball B is removed. We then determine the probabilities $p(e' \neq e|S, remove(A))$ and $p(e' \neq e|S, remove(B))$ by counting the number of worlds in which the outcome would have been different from what it actually was.

Methods

Participants & Design 80 participants ($M_{age} = 33.4$, $SD_{age} = 10.1$, 34 female) were recruited via Amazon Mechanical Turk. Half of the participants answered counterfactuals involving ball A, the other half involving ball B. Each participant saw 32 clips.[4]

Procedure Participants viewed each clip twice before answering the question: "Would ball E have gone through the gate if ball A/B had not been present?". Participants indicated their response on a slider whose endpoints were labeled "definitely no" and "definitely yes". The midpoint was labeled "unsure". After having answered the question, participants received feedback by viewing the same clip again whereby either ball A or ball B was turned into a 'ghost ball' that didn't collide with the other balls and stopped moving at the point at which it would have first collided. This was done to remind participants of what the actual clip had looked like. On average, it took participants 18.1 (SD = 4.63) minutes to complete the experiment.

Results

Figure 3 shows a scatter plot of participants' mean counterfactual judgments and the predictions by the best-fitting Noisy Newton model. For example, participants thought that there was a high chance that ball E would have gone through the gate if ball A had been removed in clip 7 (see Figure 1a) and the model correctly captures this. It also correctly predicts that people consider it unlikely that E would have gone in, if ball A (or B) had been removed in clip 3 (see Figure 1d).

The model that explains participants' counterfactual judgments best, uses a noise parameter of $SD = 1.6°$ and results in a correlation of $r = 0.88$ with $RMSE = 19.05$. A deterministic physics model (i.e. $SD = 0°$) does worse with $r = 0.82, RMSE = 30.28$. The correlation of the Noisy Newton model with participants' judgments decreases, for noise values greater than 1.6°.

Discussion

The results of Experiment 1 demonstrate that people are capable of simulating what would have happened in a counterfactual situation in which one of the balls had been removed from the scene. In previous work, we had shown that participants' counterfactual simulations were accurate for simple cases with two balls (Gerstenberg et al., 2012), and situations that involved additional objects such as bricks or teleports (Gerstenberg et al., 2014). Here, we show that people's mental simulations of counterfactuals are well-captured by a Noisy Newton model even in more complex situations that involve the collisions of several balls.

[4] You can take a look at the clips here: http://web.mit.edu/tger/www/demos/contrasts.html

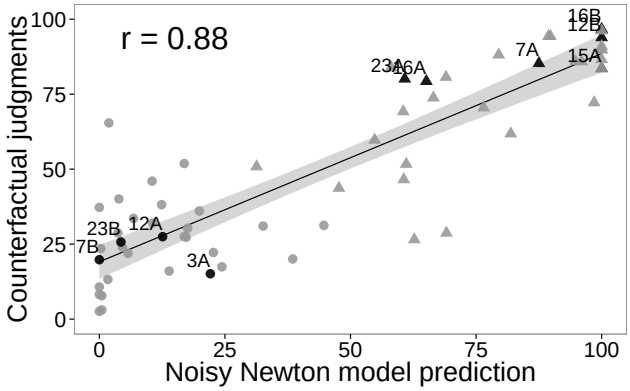


Figure 3: Scatterplot of the Noisy Newton model ($SD = 1.6°$) and participants' mean counterfactual judgments. ○ = cases in which ball E would have missed, △ = cases in which E would have gone in.

Experiment 2: Causal responsibility judgments

Experiment 1 established that people are able to mentally simulate what would have happened in different counterfactual situations. In Experiment 2 we now want to see how the different counterfactual contrasts that the *counterfactual simulation model* (CSM) postulates, influence people's causal judgments.

Methods

Participants & Design 41 participants ($M_{age} = 33.7$, $SD_{age} = 10.5$, 21 female) were recruited via Amazon Mechanical Turk. This experiment used the same set of 32 clips as in Experiment 1. In half of the clips ball E went through the gate, whereas in the other half it missed.

Procedure Participants viewed each clip three times before answering the question: "To what extent were A and B responsible for E (not) going through the gate?". The question was adapted based on the outcome of the clip. Participants indicated their responses on two separate sliders, one for each ball. The endpoints of the sliders were labeled "not at all" and "very much". On average, it took participants 21.2 (SD = 4.96) minutes to complete the experiment.

Results

In order to evaluate participants' causal responsibility judgments, we will consider three different versions of the CSM which differ in terms of the number of counterfactual contrasts they consider. The simplest model, CSM_{W}, tries to explain participants' judgments merely in terms of *whether-dependence*. Another version of the model, CSM_{WH}, also considers *how-dependence*. Finally, the CSM_{WHS} model in-

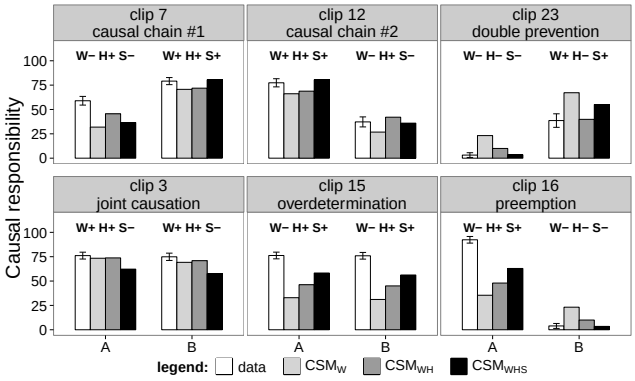


Figure 4: Mean causal responsibility ratings (white bars) and model predictions of different versions of the CSM (shaded bars) for a selection of cases. Error bars denote ± 1*SEM*. The labels above the bars indicate each ball's causal status. For example, Ball *A* in clip 7 is neither a *whether-cause* nor *sufficient* but a *how-cause*.

cludes *sufficiency* as an additional component. Table 1 shows the weights that the different models put on the different predictors and Table 2 shows the probabilities of the different counterfactual contrasts for the subset of cases shown in Figure 1. Figure 4 shows participants' causal responsibility judgments together with the predictions of the three different versions of the CSM for the same selection of clips.

To get P_W and P_S, we simply used participants' counterfactual judgments from Experiment 1. To determine P_H, we ran the physics model and generated a sample of situations in which we applied a small perturbation to the candidate ball. We then checked the proportion of cases in which E went through the gate differently from how it actually did. In cases in which the ball made no difference to the actual outcome (i.e. $P_{DM} = 0$), the other predictors were capped at 0.

The CSM_W, which only considers *whether-causation* as a predictor, struggles with several situations. In the causal chain #1 (Clip 7), participants gave a high rating to ball *B* but also a relatively high rating to ball *A*, even though $P_W(\text{A}, e)$ is very low (see Table 1). Further, the model overpredicts ratings to ball *B* in the double prevention case (Clip 23). While $P_W(B, e)$ is high in this case, participants' judgment was relatively low. Finally, it struggles with the cases in which the outcome is overdetermined and where each ball individually made no difference to whether E went through the gate (Clips 15 and 16). Over the set of 32 cases, the CSM_W accounts for merely 50% of the variance in participants' judgments. This clearly shows that participants' causal responsibility judgments in these clips cannot be explained merely in terms of

Table 1: Regression results for different versions of the CSM.

	CSM_W	CSM_WH	CSM_WHS
P_W	0.59^{***}	0.40^{***}	0.36^{***}
P_H		0.30^{***}	0.22^{***}
P_S			0.32^{***}
Constant	23.18^{***}	9.99	3.46^{*}
R^2	0.50	0.69	0.82
F Statistic	61.56^{***} (df = 1; 62)	66.54^{***} (df = 2; 61)	90.68^{***} (df = 3; 60)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 2: Probabilities of the different counterfactual contrasts for the subset of cases shown in Figure 1. P_W and P_S are based on participants' counterfactual judgments in Experiment 1.

Clip	7		12		23		3		15		16	
Ball	A	B	A	B	A	B	A	B	A	B	A	B
P_{DM}	100	100	100	100	0	100	100	100	100	100	100	0
P_W	15	80	73	6	0	74	85	78	16	13	21	0
P_H	100	100	100	100	0	0	100	100	100	100	100	0
P_S	20	85	94	27	0	80	22	15	87	84	97	0

whether-dependence.

A model which also considers how-dependence significantly improves the fit to participants' judgments with 69% variance accounted for overall. As Figure 4 shows, it correctly predicts a higher rating for ball *A* in the causal chain #1 and a lower rating for ball *B* in the double prevention case. It also gets closer to participants' judgments in situations in which the outcome was overdetermined. However, the CSM_WH predicts a relatively large difference between the joint causation case (Clip 3) and the overdetermination case (Clip 15). In the joint causation case, P_W is high for both balls, and they both also made a difference to how E went through the gate. In the overdetermination case, both balls made a difference to how E went through the gate while P_W is very low. In contrast to this prediction, the results show that participants' judgments are almost identical in the two cases.

The CSM_WHS explains this pattern of results by assuming that participants' also care about sufficiency. While in the case of joint causation, both causes are necessary but neither is sufficient, in the case of overdetermination, neither cause is necessary but they each are individually sufficient. By taking both aspects into account, the CSM_WHS correctly predicts that participants' judgments are similarly high in both situations. Overall, the CSM_WHS accounts for 82% of the variance in participants' judgments (see Figure 5 for participants' judgments for the full set of 32 different clips together with the predictions of the CSM_WHS).

The results also provide some evidence for the role of robustness in people's causal judgments (see Figure 2d). The CSM_WHS predicts incorrectly that participants' judgments should be lower for ball *A* in the preemption case (Clip 16) than for ball *B* in the causal chain #1 (Clip 7). In the causal chain, ball *B* is not a robust cause of *E*'s going through the gate. If *A*'s position had been somewhat different then *E* might not have gone through the gate. Conversely, in the preemption case, ball *A* is a very robust cause of *E*'s going through the gate. Randomly perturbing the position of the alternative cause B, doesn't affect the robustness of the relationship between *A* and *E*.

Discussion

The results of Experiment 2 show that we can explain participants' causal responsibility judgments to a high degree of quantitative accuracy by assuming that people are sensitive to a number of different factors when making their judgments. A model that considers *whether-dependence*, *how-*

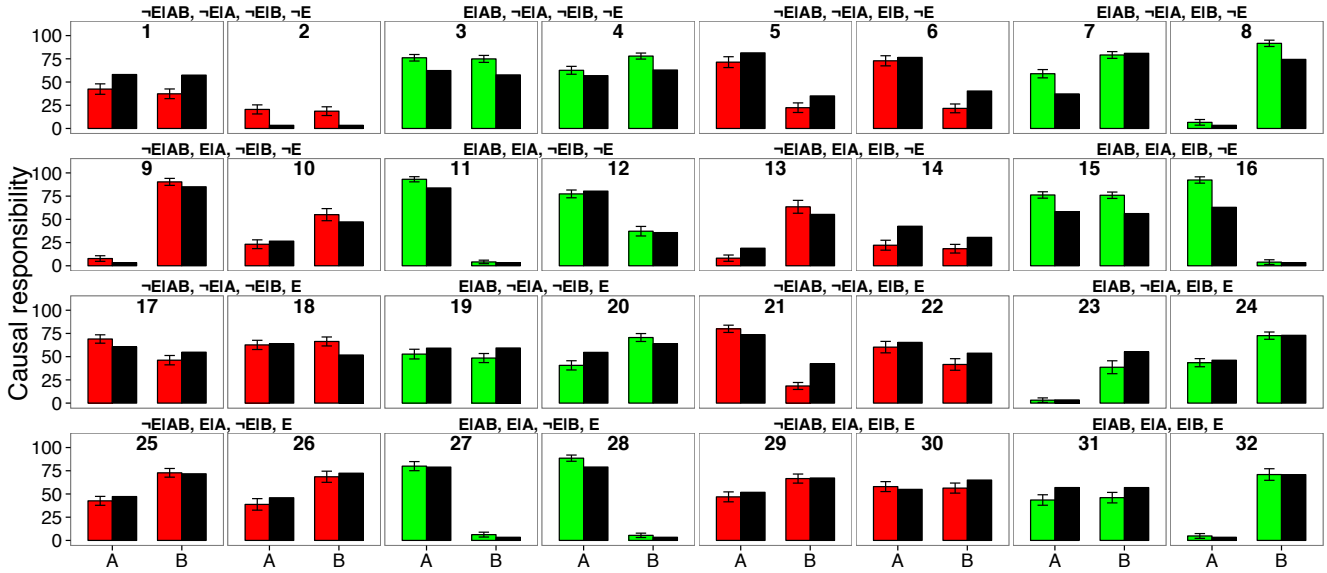


Figure 5: Mean causal responsibility (red = negative outcome, green = positive outcome) and model predictions (black bars). Error bars denote $\pm 1 SEM$. *Note*: The labels on top of each pair of clips indicate the actual outcome and the outcome in different counterfactual situations. For example, in Clips 11 and 12, ball E actually went through the gate ($E|AB$), it would have also gone through if only ball A had been present ($E|A$). However, it would not have gone through if only ball B ($\neg E|B$), or neither ball A nor ball B had been present ($\neg E$).

dependence, and *sufficiency* best explains participants' judgments. Even though there was some evidence for the importance of *robustness* in people's judgments, including it as a separate predictor did not significantly increase the model's fit.

General discussion

Causality and counterfactuals are close kin. In this paper, we have shown how to explain people's causal responsibility judgments in terms of different counterfactual contrasts defined over an intuitive domain theory. We applied the *counterfactual simulation model* (CSM) to modeling judgments about collisions between several billiard balls within the domain of intuitive physics. Experiment 1 showed that people's counterfactual judgments closely follow the predictions of a noisy Newtonian model. In Experiment 2, we demonstrated that people's causal judgments are tightly linked to their counterfactual judgments. In previous work, in which we looked at less complex stimuli that featured a single collision event only, we found that participants' causal judgments were closely related to their subjective degree of belief that the candidate cause made a difference to whether the outcome occurred (Gerstenberg et al., 2012, 2014). By considering a more challenging set of situations, we found that participants' causal judgments go beyond simple *whether-dependence*. When making causal judgments, people also care about whether the cause influenced *how* the outcome happened and, whether they believe that the cause was sufficient (and robust) for bringing about the outcome.

The CSM defines a space of four counterfactual contrasts by applying two basic operations, *remove* and *change*, to different targets – either the candidate cause, or the alternative causes. While some have argued for the existence of two fundamentally different types of causation (e.g Hall, 2004), the CSM provides a framework of unification by showing how these different causal conceptions can be understood as different counterfactual contrasts operating over the same intuitive domain theory (Schaffer, 2005). Here we have focused on applying the CSM to people's causal judgments in the domain of intuitive physics. However, the counterfactual contrasts that the CSM postulates are defined on a sufficiently general level such that the model can be applied to any domain for which we are able to write down a generative model. In future work, we will apply the CSM to modeling people's causal judgments in other domains such as interactions between social agents.

Acknowledgments This work was supported by the Center for Brains, Minds & Machines (CBMM), funded by NSF STC award CCF-1231216 and by an ONR grant N00014-13-1-0333.

References

Battaglia, P. W., Hamrick, J. B., & Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, *110*(45), 18327–18332.
Beebee, H., Hitchcock, C., & Menzies, P. (2009). *The oxford handbook of causation*. Oxford University Press, USA.
Chang, W. (2009). Connecting counterfactual and physical causation. In *Proceedings of the 31th annual conference of the cognitive science society* (pp. 1983–1987). Cognitive Science Society, Austin, TX.
Dowe, P. (2000). *Physical causation*. Cambridge University Press.
Gerstenberg, T., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2012). Noisy Newtons: Unifying process and dependency accounts of causal attribution. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 378–383). Austin, TX: Cognitive Science Society.
Gerstenberg, T., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2014). From counterfactual simulation to causal judgment. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 523–528). Austin, TX: Cognitive Science Society.
Hall, N. (2004). Two concepts of causation. In J. Collins, N. Hall, & L. A. Paul (Eds.), *Causation and counterfactuals*. MIT Press.
Lewis, D. (1986). Postscript C to 'Causation': (Insensitive causation). In *Philosophical papers* (Vol. 2). Oxford: Oxford University Press.
Lewis, D. (2000). Causation as influence. *The Journal of Philosophy*, *97*(4), 182–197.
Lombrozo, T. (2010). Causal-explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions. *Cognitive Psychology*, *61*(4), 303–332.
Michotte, A. (1946/1963). *The perception of causality*. Basic Books.
Sanborn, A. N., Mansinghka, V. K., & Griffiths, T. L. (2013). Reconciling intuitive physics and newtonian mechanics for colliding objects. *Psychological Review*, *120*(2), 411–437.
Schaffer, J. (2005). Contrastive causation. *The Philosophical Review*, *114*(3), 327–358.
Smith, K. A., & Vul, E. (2013). Sources of uncertainty in intuitive physics. *Topics in Cognitive Science*, *5*(1), 185–199.
Walsh, C. R., & Sloman, S. A. (2011). The meaning of cause and prevent: The role of causal mechanism. *Mind & Language*, *26*(1), 21–52.
Wolff, P. (2007). Representing causation. *Journal of Experimental Psychology: General*, *136*(1), 82–111.
Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford, England: Oxford University Press.
Woodward, J. (2006). Sensitive and insensitive causation. *The Philosophical Review*, *115*(1), 1–50.
Woodward, J. (2011). Mechanisms revisited. *Synthese*, *183*(3), 409–427.