Causal judgments in complex situations

Xinyi Lu[1*], Christos Bechlivanidis[2] & Tobias Gerstenberg[1]
[1]Department of Psychology, Stanford University
[2]Department of Experimental Psychology, University College London
[*]xinyilu@stanford.edu

Abstract

The Counterfactual Simulation Model (CSM) proposes that people make causal judgments by mentally simulating what would have happened in relevant counterfactual scenarios. The model assumes that causal judgments incorporate two aspects: whether-causation (would the outcome have occurred without the candidate cause?) and how-causation (did the candidate cause affect how the outcome occurred?). The CSM accurately predicts causal judgments for simple billiard ball scenarios (2 or 3 balls). Here, we explore whether it also captures people's judgments in more complex situations (up to 12 balls). We measured whether-causation (Experiment 1), how-causation (Experiment 2), and causal judgments (Experiment 3) for billiard ball scenarios varying in complexity. A model combining both whether- and how-causation best predicted causal judgments, though a how-causation-only model performed almost as well. The CSM's whether-causation predictions diverged from human judgments in complex scenarios, while how-causation predictions remained robust. This suggests that when participants make causal judgments, they might shift from counterfactual simulation toward simpler force-tracking heuristics as complexity increases.

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

Introduction

The white billiard ball caused the black ball to go into the pocket. Throwing a stone at a window caused it to shatter. The concept of causation is central to our understanding of the world – we naturally describe not only what happened but also explain why it happened.

Yet, understanding how people make causal judgments has proven challenging due to seemingly conflicting empirical findings. In philosophy, two major frameworks offer competing accounts. Process theories define causation through spatiotemporally continuous processes that transfer physical quantities like force from cause to effect (Dowe, 2000; Salmon, 1984). The stone caused the window to shatter because force was transferred from stone to glass. Dependence theories define causation through counterfactual dependence: the effect would not have occurred without the cause (Lewis, 1973; Pearl, 2000; Woodward, 2003). The stone caused the window to shatter because the window would not have shattered if the stone hadn't been thrown.

Psychological research has produced mixed results that appear to support both frameworks. Some studies find that causal judgments are strongly influenced by information about physical processes and mechanisms (Ahn et al., 1995; Shultz, 1982), while others demonstrate the importance of counterfactual dependence (Mandel, 2003; Walsh & Sloman, 2011). Some have concluded from these conflicting findings that people possess fundamentally different concepts of causation (Cartwright, 2004; Hall, 2004).

The Counterfactual Simulation Model

The Counterfactual Simulation Model (CSM; Gerstenberg et al., 2021) provides a unified account of these findings by postulating that people's causal judgments about physical interactions are sensitive to different aspects of causation. The model was originally developed and tested using billiard ball collision scenarios, where participants typically judged whether one ball (e.g., ball A) caused another ball (e.g., ball B) to go through a gate.

WHETHER-CAUSATION captures whether a candidate cause made a difference to *whether* an outcome occurred. To assess this, the model removes the candidate cause from the scene and simulates whether the outcome (broadly construed) would have been different. For example, ball A was a whether-cause of ball B's going through the gate, if ball B would not have gone through the gate without ball A. This aspect captures the notion of counterfactual dependence.

HOW-CAUSATION captures if a cause made a difference to *how* an outcome occurred. To assess this, the model applies a small perturbation to the candidate cause and simulates whether the outcome (finely construed) would have been any different. For example, ball A was a how-cause of ball B's going through the gate, if ball B would have gone through the gate (slightly) differently if ball A had been lightly perturbed. This aspect integrates the notion of a force transfer from cause to effect into a counterfactual framework.

How-causation, as implemented in the model, is binary (ball A either was a how-cause or it wasn't), while whether-causation is graded. To compute whether-causation, the model runs several noisy simulations about what would have happened without the candidate cause and then takes the proportion of cases where the outcome would have been different from what actually happened. So, the clearer it is that the outcome would have been different without ball A, the more of a whether-cause ball A was. Notice that this means that, in the CSM, computing how-causation is less computationally intensive than computing whether-causation.

This computational approach has proven successful in predicting human judgments. Gerstenberg et al. (2021) show that a model that combines how-causation and whether-causation accurately captures participants' causal judgments across a range of situations that featured two or three billiard balls.

The Challenge of Complexity

Despite its success in these simple scenarios, a critical question remains unanswered: Does counterfactual simulation predict causal judgments in more complex scenarios, such as those involving more interacting balls?

We suspect that the two components of the CSM – whether-causation and how-causation – are differentially affected by increasing complexity. While fully reconstructing *how* an outcome occurred in a fine-grained fashion would be computationally difficult, the model's implementation of how-causation requires merely detecting *if* the outcome would have changed at all. Thus, how-causation is computationally equivalent to tracking (a chain of) physical contacts: did ball A contact ball B directly or indirectly at any point in time? This only requires tracking ball A's trajectory and detecting if it directly collided with ball B, or if there was a chain of collisions that connects A with B. In contrast, whether-causation requires full counterfactual simulation – imagining what would have happened if the candidate cause were removed – which faces several challenges as complexity increases.

First, complex scenarios are harder to track perceptually; numerous studies have shown that multiple object tracking is limited to a handful of objects (e.g., Meyerhoff et al., 2017; Pylyshyn & Storm, 1988).

Second, in multi-step causal chains, small uncertainties in simulating early events can propagate and amplify, making later events increasingly unpredictable (Smith & Vul, 2013). Thus, even if people can simulate simple interactions fairly accurately, their simulations of complex cascades can become highly uncertain.

Third, in simple scenarios with 2–3 balls, there is often a single clear causal path from cause to effect. In complex scenarios with many more balls, however, multiple pathways of causal influence may exist: even if ball A had not hit ball B, another ball might have hit ball B instead and produced the same outcome. Evaluating whether A was truly necessary requires accurately simulating this complex web of counterfactual trajectories.

Fourth, accurate counterfactual simulation of a complex cascade involves imagining the trajectory of multiple objects. However, recent work suggests that our capacity for mentally simulating multiple entities is likely severely restricted (Balaban & Ullman, 2025).

Therefore, we might expect how-causation to be much easier to compute than whether-causation as complexity increases. Here, we test the limits of counterfactual simulation by examining how scene complexity affects judgments of whether-causation and how-causation, and if whether- and how-causation continue to predict causal judgments in complex scenarios.

Experiment Overview

We operationalized scene complexity by increasing the number of balls in billiard ball scenarios. We created 20 billiard ball videos that featured ball A (labeled and in green) and ball B (labeled and in blue), as well as optional unlabeled gray balls, in which ball B always went through a gate on the left. There were 4 videos in each complexity level (2, 3, 6, 9, or 12 balls). Figure 1B shows examples of videos in each complexity level. Videos were generated using the physics engine Pymunk (Blomqvist, 2025) and ranged from 7–13 seconds.

We ran three experiments using the same 20 videos, programmed in jsPsych (de Leeuw, 2015). Figure 1A shows an overview. Experiment 1 collected whether-causation judgments (would ball B have gone through the gate without ball A?). Experiment 2 collected how-causation judgments from a different group (did ball A influence how ball B went through the gate?). Experiment 3 collected causal judgments from a third group (did ball A cause ball B to go through the gate?).

This approach allows us to test three research questions:

1. Do whether- and how- judgments become more uncertain as scenario complexity increases?
2. To what extent do whether- and how-causation explain causal judgments as scenario complexity increases?
3. How well does the CSM capture people's whether- and how- judgments as scenario complexity increases?

All pre registrations, materials, data, and analyses are available here: https://github.com/cicl-stanford/complex_causation_cogsci2026

Experiments 1 & 2: Whether- and How-Judgments

Methods

Participants For each experiment, we recruited 50 participants from Prolific (Experiment 1: $M_{\text{age}} = 43$, $SD = 15$, gender: 25 male, 25 female; race: 37 White, 3 Black/African American, 6 Asian, 3 Multiracial/Mixed, and 1 other; Experiment 2: $M_{\text{age}} = 41$, $SD = 14$, gender: 22 male, 27 female, 1 non-binary; race: 39 White, 6 Black/African American, 3 Asian, 1 American Indian/Alaska Native, and 1 Multiracial/Mixed). Participants were paid $2 for approximately 15 minutes, plus a performance-based bonus of up to $2 (Experiment 1 mean = $1.35; Experiment 2 mean = $1.62).

Materials and Procedure The procedure was identical across both experiments except for the prompt and instructions. Figure 1A shows an overview of the prompts used in each experiment. All participants completed 20 trials in randomized order. On each trial, a video clip played twice in succession, after which participants responded to a prompt using a slider scale anchored at "definitely not" (0), "unsure" (50), and "definitely yes" (100).

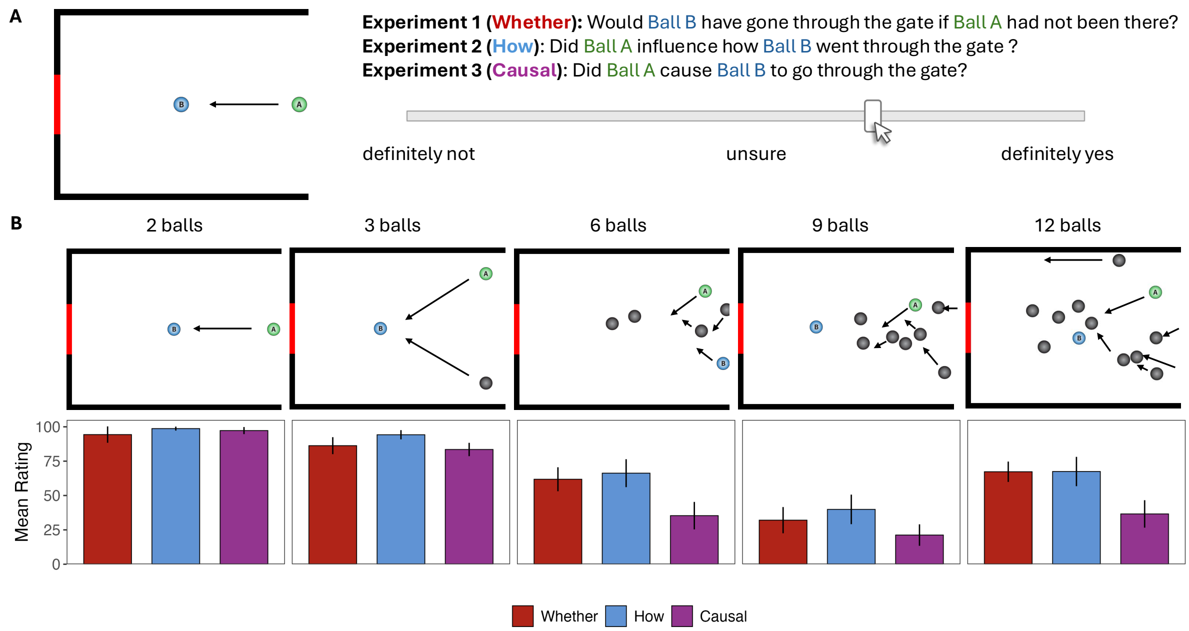

Figure 1: **Experiment overview and examples**. (A) Overview of the three experiment prompts. (B) Example videos from each complexity level (2, 3, 6, 9, and 12 balls). Bars represent mean judgments for whether-causation (Experiment 1), how-causation (Experiment 2), and causal judgments (Experiment 3). Error bars show 95% confidence intervals.

Before beginning the experimental trials, all participants were given instructions that explained the respective prompt along with an example video. Experiment 1 instructions asked participants to consider what would have happened to ball B if ball A had not been present. Experiment 2 instructions explained that "influence" meant ball A either directly contacted ball B (A→B) or indirectly influenced ball B through a chain of collisions (A→gray ball→B).

Results and Discussion

For whether-causation judgments, participants rated whether ball B would have gone through the gate if ball A had not been there. We reverse-coded these judgments by subtracting participants' ratings from 100 to yield the probability that ball B would have missed the gate in the absence of ball A. How-causation judgments were used directly, as they already captured whether ball A influenced how ball B went through the gate.

Effect of Complexity on Certainty Do whether- and how-judgments become more uncertain as scenario complexity increases? To test this question, we transformed judgments into certainty scores using the formula: certainty = $|\text{judgment} - 50| \times 2$, yielding values from 0 (maximally uncertain, at the scale midpoint) to 100 (maximally certain, at either extreme). Figure 2 shows participants' judgment certainty for different numbers of balls.

For each experiment, we fit hierarchical Bayesian models predicting certainty from number of balls with random intercepts and slopes for participants using the `brms` package in `R` (Bürkner, 2017). The random slopes were removed for Experiment 1 due to convergence issues.

Experiment 1 (Whether): Certainty about whether ball B would have gone in without ball A decreased as the number of balls increased ($\beta = -2.15$, 95% CrI $[-2.58, -1.73]$), from $M = 89.8$ ($SD = 7.79$) for the 2-ball videos to $M = 62.1$ ($SD = 6.77$) for the 12-ball videos.

Experiment 2 (How): Certainty about whether ball A influenced ball B also decreased when more balls were present in the scene ($\beta = -1.32$, 95% CrI $[-1.78, -0.86]$), from $M = 86.8$ ($SD = 3.26$) for the 2-ball videos to $M = 78.21$ ($SD = 1.23$) for the 12-ball videos.

Both experiments demonstrated that participants were less certain about their judgments in more complex scenarios. The effect was numerically stronger for whether-causation than how-causation, though both showed reliable decreases in certainty with increasing complexity. We performed an exploratory analysis (not preregistered) comparing certainty between whether- and how-judgments. A hierarchical Bayesian model revealed that how-judgments were more certain than whether-judgments ($\beta = -4.22$, 95% CrI $[-5.92, -2.52]$); the decrease in certainty as the number of balls increased was not steeper for whether-judgments than how-judgments (interaction $\beta = -0.42$, 95% CrI $[-0.88, 0.04]$).

Experiment 3: Causal Judgments

Experiment 3 collected causal judgments for the same set of stimuli used in Experiments 1 and 2.

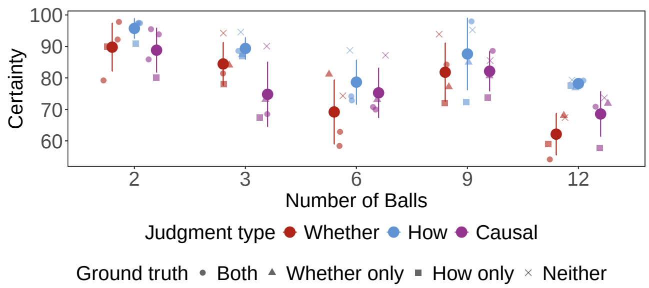

Figure 2: **Mean certainty ratings** across different numbers of balls for whether (Experiment 1), how (Experiment 2), and causal (Experiment 3) judgments. Large points show the overall mean for each complexity level. Small points show means for individual videos, with shapes indicating the ground truth based on whether ball A was only a whether-cause (with no noise added in the counterfactual simulation), only a how-cause, both, or neither. Error bars show 95% confidence intervals.

Methods

Participants and Materials We recruited 50 participants from Prolific ($M_{\text{age}} = 41$, $SD = 11$, gender: 24 male, 25 female, 1 non-binary; race: 36 White, 5 Black/African American, 5 Asian, 1 American Indian/Alaska Native, 2 Multiracial/Mixed, and 1 undisclosed) who were paid $3 for approximately 15 minutes.

Procedure The experimental procedure was identical to Experiments 1 and 2 except the prompt and instructions (see Figure 1A for overview).

Results and Discussion

Whether- and how-judgments predict causal judgments Our primary question was how well empirical whether- and how-judgments (from Experiments 1 and 2) predicted causal judgments (Experiment 3). Figure 3 shows the relationship between three model variants and participants' causal judgments. We fit a hierarchical Bayesian model predicting causal judgments from mean whether-causation judgments (reverse-coded) and mean how-causation judgments for each video, with random intercepts and slopes for participants. To quantify overall model fit, we compared predicted values (using population-level effects only) to observed mean causal judgments for each of the 20 videos. The model achieved excellent fit ($r = .941$, RMSE $= 10.67$), indicating that whether- and how-judgments captured well how people make causal judgments in our videos.

We compared the full model to a whether-only and a how-only model using approximate leave-one-out cross-validation. The combined model provided the best predictive performance. However, the how-only model performed only slightly worse (ELPD difference $= -57.4$, SE $= 13.3$), while the whether-only model showed substantially worse fit (ELPD difference $= -195.8$, SE $= 22.9$). Thus, both whether- and how-judgments were predictive of causal judgments. Notably, how-causation was a substantially stronger predictor ($\beta = 0.72$, 95% CrI $[0.57, 0.86]$) than whether-causation ($\beta = 0.19$, 95% CrI $[0.02, 0.35]$).

To examine how model performance changes with complexity, we used a rolling window approach. Since each complexity level had only 4 videos, each window combined 12 videos from 3 consecutive complexity levels (so the first window included 2-, 3-, and 6-ball videos; the second included 3-, 6-, and 9-ball videos; the third and last included 6-, 9-, and 12-ball videos). For each window, we fit the 'how + whether model'. Figure 4A shows how the contributions of whether-causation and how-causation to causal judgments changed with increased complexity. As the window moves rightward (increasing ball count), the predictive power of both whether-causation and how-causation decline. Nevertheless, how-causation consistently receives stronger weight than whether-causation in predicting participants' judgments across all complexity levels.

Individual Differences We observed a strong negative correlation between participant whether- and how-slopes ($r = -0.96$), suggesting these may represent competing strategies or a trade-off in how people judge causation. Thus, for each participant, we examined how well their judgments were explained by the three different model variants. To test the models, we z-scored individual participants' judgments, and restricted the priors over the regression weights to be positive. Model comparison via leave-one-out cross-validation revealed substantial individual variation: while the full model best explained the aggregate pattern, most individual participants were best fit by simpler models: 58% ($n = 29$) by the "how-only model" and 24% ($n = 12$) by the "whether-only model", with only 18% ($n = 9$) best fit by the "how + whether model".

Comparing CSM predictions to human judgments How well does the CSM capture people's whether- and how-judgments as scenario complexity increases? To answer this question, we compared human whether- and how-judgments obtained in Experiments 1 and 2 to the whether- and how-predictions generated by the CSM.

For whether-causation, the model removes the candidate cause ball A and then introduces noise to capture people's uncertainty about what would have happened. We apply noise in the form of random perturbations to the direction of the first ball that ball A contacts at each time step in the simulation after the time point at which ball A and that ball made contact in the actual scenario. We generated 1000 noisy samples for each of the 20 video clips using different degrees of noise ranging from $\theta = 0$ to $\theta = 5$ in steps of 0.1, where θ refers to the standard deviation of the Gaussian distribution from which the perturbations to the velocity vector were drawn. For each video, we computed the CSM's predicted probability that ball B failed to reach the gate if ball A had been removed from the scene. We fit the noise parameter to participants' judgments by minimizing the sum of squared errors between

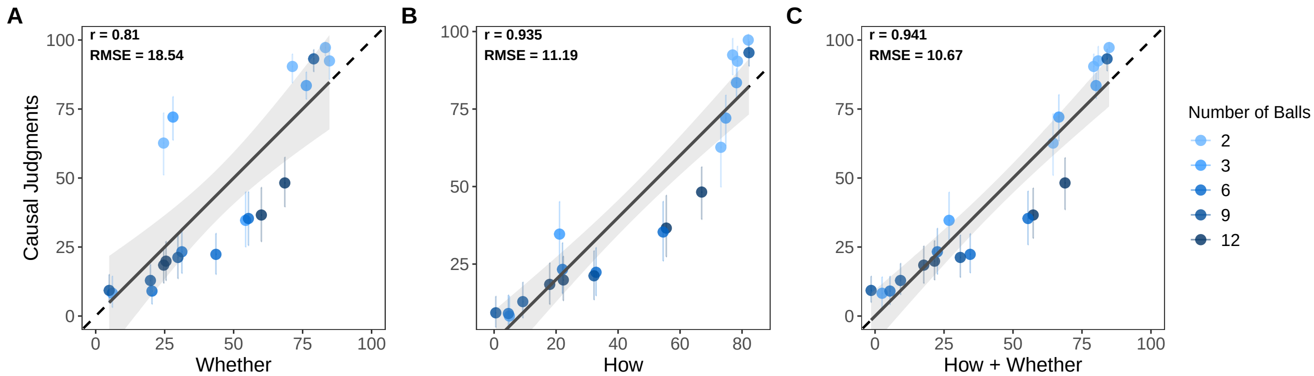

Figure 3: **Predictions versus empirical causal judgments for each of the 20 videos**. Each panel shows a different predictive model: using whether-judgments only (A), how-judgments only (B), and how + whether judgments (C). Points represent individual videos, colored by the number of balls in the video. Solid lines indicate actual fit. Error bars show bootstrapped 95% confidence intervals.

model predictions and participants' mean whether-judgments. The best fitting model had a noise value of $\theta = 1.9$ with $r = .694$ and RMSE = 28.11.

For how-causation, the CSM applies a very small perturbation to ball A's initial position and then checks whether ball B's outcome trajectory would have been different from what it was in the actual video. If so, ball A was a how-cause of ball B's going through the gate. This yields a binary measure: A is either a how-cause or not a how-cause.

Figure 5 shows the overall correlations between the CSM whether and how predictions and the corresponding human judgments. The correlation for whether was higher than for how. We suspected that people have greater difficulty simulating whether-causation for more complex cases involving more balls. However, since we had only 4 videos per condition, we used the rolling window approach to examine this relationship. Figure 6A shows the correlation between whether- and how-predictions across different complexity windows. As the window moves rightward (increasing ball count), we observe a growing divergence of human whether-judgments from the CSM's ground truth. Notably, this does not seem to occur for how-judgments, which participants continued to predict well, even in videos with many balls. Figure 6B shows that at higher ball counts, the correlation between human whether- and how-judgments (shown in black) approaches ceiling, even when there is no such correlation in the ground truth of the CSM (shown in blue). This suggests that participants increasingly rely on the same kinds of information to judge whether- and how-causation as complexity increases.

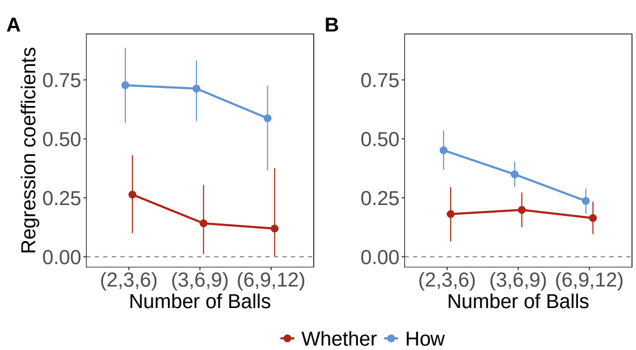

Figure 4: **Predicting causal judgments from whether-causation and how-causation across rolling complexity windows**. Each point represents the regression coefficients from the 'how + whether model' fit to a window of 12 consecutive videos (sorted by number of balls). (A) Using human whether and how judgments as predictors. (B) Using CSM whether- and how-predictions as predictors. Points represent posterior means, error bars show 95% credible intervals.

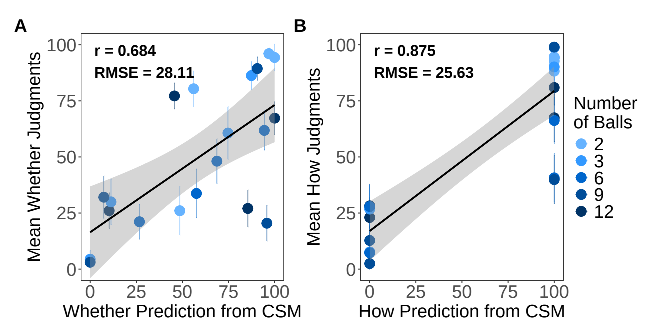

Figure 5: **CSM predictions versus human judgments**. (A) Whether-causation: CSM whether prediction versus mean human whether judgments. (B) How-causation: CSM how prediction versus mean human how judgments. Points represent individual videos, colored by the number of balls present. Error bars show 95% confidence intervals.

General Discussion

The Counterfactual Simulation Model (CSM) proposes that people make causal judgments by mentally simulating what would have happened in relevant counterfactual scenarios (Gerstenberg et al., 2021). We tested whether the model's core assumptions hold in complex scenarios where mental simulation may become difficult, uncertain, or unreliable. We systematically varied the number of billiard balls present in the videos (2–12 balls) and measured whether-causation judgments (Experiment 1), how-causation judgments (Experiment 2), and finally causal judgments (Experiment 3). Participants' uncertainty about both whether- and how-causation increased as the number of balls increased, with certainty about whether-causation decreasing more than certainty about how-causation.

Whether- and how- still predict causal judgments

A combined model using both whether and how judgments successfully predicted causal judgments across our range of simple to complex billiard ball scenarios, suggesting that participants relied on both whether- and how-causation when judging causality. However, the how-only model performed almost as well as the combined model, and was the best model for the majority of individual participants, suggesting that how-causation was the primary factor driving causal judgments across our scenarios.

We observed that as the number of balls in a video increased, people's how judgments remained well aligned with the CSM's how-causation predictions but increasingly diverged for whether-causation (Figure 6A). This suggests that as scenarios become more complex and difficult to simulate, participants' judgments may shift away from the counterfactual simulation process as predicted by the CSM.

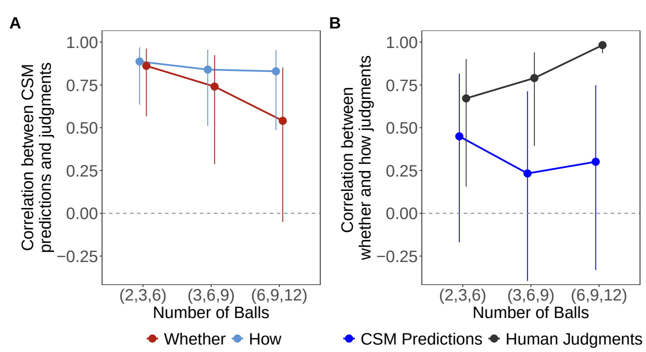

Figure 6: **Correlations computed across rolling complexity windows**. Each point represents the correlation computed over a window of 12 consecutive videos (sorted by number of balls). (A) CSM predictions versus human judgments for whether-causation and how-causation across complexity windows. (B) Correlation between whether- and how-judgments for human participants and CSM. Error bars show bootstrapped 95% confidence intervals.

Potential strategy shift in complex scenarios

If participants are engaging in counterfactual simulation less, what are they doing instead? One clue is the high correlation between human whether- and how-judgments, which approaches ceiling for complex scenarios – even when judgments are much less correlated with the ground truth (Figure 6B). This suggests that participants increasingly rely on the same kinds of information for both whether- and how-judgments as complexity increases.

One interpretation is that in highly complex scenarios with many balls, detailed counterfactual simulations become computationally untenable. Instead, participants may fall back on simpler heuristics of tracking force transfer, analogous to how-causation. This would explain why (1) how-causation was a stronger predictor of causal judgments compared to whether-causation in this set of videos; (2) human whether judgments diverge from the CSM's whether-predictions as the number of balls in a video increases, but how-predictions remained accurate; and (3) participants' whether- and how-judgments appear increasingly correlated as the number of balls in a video increases. This could represent a rational strategy shift in which participants abandon costly counterfactual reasoning in favor of more efficient causal cues when uncertainty is high. We also note that participants falling back on "how" to judge "whether" represents an ecologically valid strategy. We selected videos such that the correlation between how-causation and whether-causation was relatively low (see Figure 6B). However, across a large number of randomly generated videos like the ones that participants saw, the correlation between how- and whether-causation in the ground truth is in fact much higher (r = 0.691, 95% CI [0.679, 0.704]).

Limitations and Future Directions

A few limitations of the present work should be noted. First, we used only 20 videos across 5 complexity levels, providing limited power for analyses within each level. This precluded traditional interaction analyses, and while our rolling window approach provides suggestive evidence, it should be interpreted cautiously as exploratory. Second, using the number of balls as a crude complexity measure conflates several factors, such as the number of collisions, the length of the causal chain between A and B, and the number of alternative causal pathways. Future research with larger stimulus sets could address both issues by directly manipulating these factors to provide more definitive tests of the effect of complexity.

One line of work we are interested in is examining if how-causation and whether-causation are weighted by participants' certainty about each aspect when making causal judgments. If people integrate whether and how information weighted by their confidence in each aspect, we might expect whether-causation to contribute less to overall causal judgments in complex scenarios. A model that incorporates this aspect of confidence or certainty could provide a more nuanced account of how people adaptively combine multiple kinds of information when judging causality.

Acknowledgments

XL was supported by a postdoctoral fellowship from the Natural Sciences and Engineering Research Council of Canada (NSERC). TG was supported by grants from the Stanford Institute for Human-Centered Artificial Intelligence (HAI) and from the Cooperative AI Foundation.

References

Ahn, W., Kalish, C. W., Medin, D. L., & Gelman, S. A. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition*, *54*(3), 299–352.

Balaban, H., & Ullman, T. D. (2025). The capacity limits of moving objects in the imagination. *Nature Communications*, *16*, 5899.

Blomqvist, V. (2025, November). Pymunk [A easy-to-use pythonic rigid body 2d physics library].

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28.

Cartwright, N. (2004). Causation: One word, many things. *Philosophy of Science*, *71*(5), 805–819.

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1–12.

Dowe, P. (2000). *Physical causation*. Cambridge University Press.

Gerstenberg, T., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2021). A counterfactual simulation model of causal judgments for physical events. *Psychological Review*, *128*(5), 936–975.

Hall, N. (2004). Two concepts of causation. In J. Collins, N. Hall, & L. A. Paul (Eds.), *Causation and counterfactuals* (pp. 225–276). MIT Press.

Lewis, D. (1973). Causation. *Journal of Philosophy*, *70*(17), 556–567.

Mandel, D. R. (2003). Judgment dissociation theory: An analysis of differences in causal, counterfactual and covariational reasoning. *Journal of Experimental Psychology: General*, *132*(3), 419–434.

Meyerhoff, H. S., Papenmeier, F., & Huff, M. (2017). Studying visual attention using the multiple object tracking paradigm: A tutorial review. *Attention, Perception, & Psychophysics*, *79*, 1255–1274.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge University Press.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, *3*(3), 179–197.

Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world*. Princeton University Press.

Shultz, T. R. (1982). Rules of causal attribution. *Monographs of the Society for Research in Child Development*, *47*(1), 1–51.

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.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford University Press.