UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Are counterfactuals necessary for actual causation judgments?

Permalink
https://escholarship.org/uc/item/4885s5jd

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors
Hale, Randall
Cheng, Patricia

Publication Date
2026

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org

Powered by the California Digital Library
University of California

Are Counterfactuals Necessary for Actual Causation Judgments?

Randall Hale (randall.hale@ucla.edu)[1] & Patricia Cheng[1]
[1]Department of Psychology, UCLA

Abstract

Two descriptive accounts of actual causation judgment, the Counterfactual Simulation Model (CSM) and the Counterfactual Effect Size Model (CESM), propose that actual-cause judgments for an observed event are derived from operations over mentally simulated counterfactual alternatives of the event. We argue that counterfactual models face computational and conceptual obstacles and propose the Retrieval of Invariant Causal Knowledge (RICK) model, which utilizes forward mental simulation to map type-level causal knowledge directly to token causal events. We demonstrate two predicted shortcomings of the CSM and compare with our model's performance.

Keywords: actual causation; counterfactual simulation; causal reasoning

Introduction

Counterfactuals figure centrally in many normative theories of causal inference. In philosophy, the counterfactual tradition traces back to David Lewis (1973) and has since been developed into several influential frameworks, including Pearl's (2000) structural equation modeling and Woodward's (2003) interventionist account. More recently, these normative counterfactual accounts have been extended to descriptive accounts of human causal judgment, in models such as the Counterfactual Simulation Model (CSM; Gerstenberg et al., 2021) and the Counterfactual Effect Size Model (CESM; Quillien & Lucas, 2024), both of which propose that actual-cause judgments for an observed event are derived from operations over mentally simulated counterfactual alternatives of the event.

In its basic form, a counterfactual compares what happens in the observed world with what happens in a hypothetical world where one had intervened on a particular variable or set of variables. If the outcomes of the actual and counterfactual worlds differ, the reasoner has strong grounds for judging the variable as causal. However, in certain cases, intervening on a genuine cause fails to yield a different outcome. There are many well-known complications of this sort that counterfactual theories contend with, such as preemption, overdetermination, trumping, temporal asymmetry, and so on. These challenges have prompted a variety of specialized counterfactual tests, each tailored to handle a particular problematic case that previous theories could not accommodate. As a result, counterfactual theories have become increasingly complex (Menzies & Beebee, 2025). Excessive complexity is a problem for any theory, but it is especially problematic in cognitive modeling, where the relevant computations must remain psychologically plausible. Moreover, models relying on counterfactual simulation face another fundamental issue, which is that counterfactual simulation often does not supply information that is *not already given* in the knowledge necessary to generate those counterfactuals in the first place. While there are certainly cases where inferences cannot be drawn directly from the knowledge available to the reasoner, and must instead be extracted through other means, including simulation (Aronowitz & Lombrozo, 2020), both the CSM and CESM make a stronger claim, that counterfactual simulation is the central mechanism of actual causation judgment. We argue instead that, excepting rare edge cases, counterfactuals are unnecessary for yielding causal judgments that align with human intuition.

To illustrate that counterfactual simulation is often unproductive, consider a scenario where the passenger of a car observes the driver press the gas pedal, shortly after which the car accelerates forward. If counterfactual simulation were necessary to judge whether pressing the gas pedal caused the car to move, the passenger would need to simulate whether the car would have moved had the driver *not* pressed the pedal. But, without prior knowledge that there exists a causal relationship between the events *pressing a gas pedal* and *car accelerating*, the passenger would be unable to simulate what the counterfactual outcome would be. If specifying the counterfactual already presupposes the relevant causal dependency, why not directly apply that knowledge to the token instance?

In this paper, we demonstrate two predicted shortcomings of the CSM: 1) the computational demands of counterfactual simulation necessarily scale with the number of candidate causes in the judged scenario, and 2) causal attribution judgments derived from counterfactuals are highly unstable in noisy environments, and are only viable within simple, controlled causal scenarios.

In Experiment 1, we adopt stimuli similar to those used by Gerstenberg et al. (2021): participants view animations of colored balls colliding and judge which ball causes a gray 'effect' ball to pass through a goal. According to the CSM, each candidate cause is first subject to a 'difference-maker' counterfactual, which screens out causally irrelevant candidates. Difference-makers are then evaluated on four causal dimensions: whether, how, robustness, and sufficiency. Robustness and sufficiency each require two simulations, meaning $n + 6dm$ counterfactual simulations are required for each

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society* ©2026 the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

scenario (n: number of candidate causes, dm: number of difference-makers).

Since the number of required simulations grows with dm by a factor of 6, the CSM predicts longer judgment times as dm (or n) increases. We tested this prediction by comparing judgments across scenarios with 1, 2 and 3 DMs, measuring the time to make a causal judgment.

Additionally, to account for participant uncertainty, the CSM injects noise wherever the counterfactual diverges from the actual scenario (i.e. at any collision not actually observed)[1]. In scenarios where multiple candidate causes interact, simulated outcomes become increasingly sensitive to initial conditions, as they involve simulating more unobserved collisions (a manifestation of the 'butterfly effect') . In consequence, causal attribution tends toward uniformity across difference-makers, dramatically diverging from human causal judgments. This divergence is especially pronounced for the proximal cause; even in cases with dense interactions among candidates, participants reliably identify which ball ultimately causes the effect ball to enter the goal, showing no comparable loss of discrimination.

We seek a more computationally efficient and robust method for applying prior knowledge to explain token events, and propose the Retrieval of Invariant Causal Knowledge (RICK) model as an alternative to counterfactual simulation. As the name suggests, our model differs fundamentally in how we consider causal knowledge to be represented in memory. For the CSM, causal knowledge consists of abstract physical laws that are used to simulate interactions between objects, analogous to a video game engine. This knowledge does not directly yield causal generalizations, rather, causal judgments must be derived from simulated outcomes generated from that knowledge. Notably, this characterization of causal knowledge does not align with the typical assumptions about causal knowledge in the causal induction literature, where type-level representations of causal relations are thought to be the output of the induction process (Cheng, 1997; Gong et al., 2025).

We first define the relevant constructs to formalize the application of type-level knowledge to token events. Then, we present a minimal characterization for a type-level causal relation that may be applied to the physical collision events used by Gerstenberg et al. (2021) and demonstrate how this application may resolve the aforementioned problems faced by counterfactual theories.

Event Structure and Causal Attribution

We treat events, rather than objects or entities, as the primitives over which causal judgment operates. In our framework, an *event* is a change in the state of a variable, where the variable represents a particular dimension of an entity or system. For an entity X that, at a given moment, takes value v_i along dimension V, we define an event E as a change in the value of V from v_i to v_j. An event category, in turn, is a schematized representation of a particular kind of state change. For any entity possessing dimension V, the event category $\mathcal{E}$ denotes the change from v_i to v_j. To illustrate, consider the claim: *the balloon popped as a result of rising to high altitude*. The causal event corresponds to the state of a variable representing the pressure of the atmosphere at the balloon's location, which decreases as altitude increases, while the outcome event corresponds to a state transition of a variable representing the balloon's structural integrity, from intact to not intact. This type of transition is schematized by the event category *popped*.

Instantiations of event categories are referred to as *event tokens*, and serve as the input for actual cause judgment. A variable may be represented at any level of abstraction, which determines the set of values it may assume. Most often, the set of meaningful values associated with a variable does not include the entire set of perceptually distinct states. Instead, perceived states must be mapped onto a coarser set of values specified by the variable's level of abstraction. For instance, a variable representing the angle of a door relative to its frame may take values in the condensed set open, closed, even though the underlying physical state varies continuously over a range of angles between 0° and 120° (presumably the set of perceptually distinct states lies somewhere between these abstractions). Differences within the physical range of values may be perceptually discriminable yet causally irrelevant, and are therefore collapsed into the same representation (*open* or *closed*). We assume that the level of abstraction specified for a variable is determined by the observer's present goals, attentional resources, or other situational constraints, and may therefore fluctuate.

The use of events, rather than objects or entities, as the units of causal judgment may appear at odds with the way causal claims are often expressed. For example, one might say that *the forest fire was caused by an arsonist*. In this case, the cause (the arsonist) is framed as an entity category. Indeed, attributing causal status to entities is often necessary for assigning responsibility, blame, or credit for an outcome (Lagnado & Channon, 2008). Entity attribution can be derived from an event-based representation by tracing backward from the causal event to the variable whose state change constitutes that event, and then identifying the entity or system to whom the variable refers. When an entity is implicated as the bearer or participant in causally relevant state changes along this chain, it may inherit causal responsibility in an entity-based description.

We assume that the observer possesses a vast repertoire of learned causal relations. But in building a model of actual cause judgment, the modeler must still decide what those relations are taken to specify. This may seem uncomfortably flexible, since it appears to permit the positing of whatever

[1]This procedure is incorrectly described by Gerstenberg et al. (2021); they specify adding a perturbation for every timestep after the point of counterfactual divergence (as in a 'random walk'). In their actual implementation of the CSM, however, they inject noise only at the frame where the counterfactual simulation diverges (this only occurs at collisions), applying the perturbation to whichever balls are involved in the counterfactual collision; we adopt the latter procedure in our implementation.

type-level causal relation best fits the case at hand. Our aim, however, is not to specify the precise representational structure of learned causal relations, but to make a more limited point: for judgments of this kind, even a minimally specified type-level causal rule can yield a more faithful account than counterfactual simulation. Human causal knowledge is likely far richer than the sparse rule we employ here, but that richer structure need not be specified for the present argument. In applying our framework to the physical collision scenarios used, we define the type-level causal relationship as: collisions between physical objects can cause changes in the velocity of one or both colliders, formalized below:

$$\text{collision}(a, b) \rightarrow \Delta v_a \lor \Delta v_b$$

This rule requires minimal assumptions about any other properties of a or b and applies invariantly across a wide range of physical interactions. While the rule may fail in some cases (ex. a collision destroys the structural integrity of an object such that it can no longer be treated as an encapsulated entity), such possibilities do not arise for the scenarios we tested[2].

In the present stimuli, the outcome of interest is whether the grey ball passes through the goal. Although an episode may contain multiple collision events prior to that outcome, only one of them (the proximal-cause collision) directly changes the grey ball's velocity. To connect the continuous dynamics of motion to the judgment participants are asked to make, the perceivable state of the effect ball, including its position and velocity, must be mapped onto the constrained set $\{goal, miss\}$. At this level of abstraction, the outcome event is defined as the grey ball's transition from $miss$ to $goal$. Because the grey ball begins motionless, it remains unambiguously in the $miss$ state until it is struck by the proximal cause. Applying the $collision \rightarrow \Delta v$ rule to the observed event sequence therefore reveals the proximal cause's causal role directly: it is the collision event that produces the change in the grey ball's velocity that initiates the $miss \rightarrow goal$ transition. From this event-level attribution, an entity-level attribution follows straightforwardly: the non-grey ball participating in that collision is identified as the causal entity because it is the participant in the causally relevant event.

This resolves the attribution problem for the proximal cause, but the problem becomes more difficult for more distal candidate causes. Consider the two-difference-maker case in which ball A collides with ball B, ball B collides with the effect ball, and the effect ball enters the goal. Here, the $A \times B$ collision does not itself produce the outcome event; rather, its influence on the grey ball is mediated by the later $B \times Effect$ collision.

[2] We acknowledge that visual contact between objects may not always be judged as the cause of the objects' subsequent motion. For example, in Michotte's 'launching' stimuli, one scenario featured a ball that oscillated horizontally between two points. Given that the change in this ball's motion was predetermined by this behavior, contact with another ball was not judged as causal for the ball's subsequent motion. We argue that this is a case of overdetermination; the collision rule we specify would produce a change in the balls motion in the absence of the other cause that is responsible for the oscillation.

The attribution problem therefore depends on whether the observer can map the earlier collision onto the later $miss \rightarrow goal$ transition. If forward simulation of the observed scene makes that mapping apparent, attribution to the distal collision is possible. If the outcome-relevant consequences of the earlier collision are not readily recoverable at that point, the mapping becomes difficult, and attribution should weaken accordingly. This account also predicts cases in which distal, or 'root', causes receive the strongest attribution. Suppose three squares are shown on the screen, two moving only vertically and a third moving rightward from the left. The rightward-moving square collides with the first, transferring horizontal velocity; that square transfers horizontal velocity to the second; and the second then strikes the effect ball into the goal. Because neither mediating square initially possessed horizontal velocity, the horizontal motion that eventually reaches the effect ball is easily traceable to the first square. In such a case, the root cause should receive the bulk of causal attribution. More generally, we suggest that people prefer root causes when the mapping from an earlier event to the outcome remains clear, but we do not include an additional parameter to account for root-cause bias in the present implementation. Regardless, the preference for proximal causes observed in the present collision stimuli arises when that mapping becomes more difficult, causing attribution to more distal causes to deteriorate.

Model Specification

Let an episode be represented as an ordered sequence of observed collision events,

$$\mathcal{T} = (c_1, c_2, \ldots, c_n)$$

where each event token c_t is a collision between two objects at time t. For the present stimuli, we restrict attention to the realized support set terminating in the outcome event, that is, the subset of observed collisions that feed into the actual proximal collision with the effect ball. A collision is included in this support set if it is the proximal collision itself, or if its collided object later serves as the collider in a support collision. This allows multiple earlier collisions to support the same downstream event. Collisions outside this realized support set are assigned zero attribution.

For each collision event in the realized support set, the model first assigns directional roles to the two participants. In the present stimuli, because the balls travel from right to left, we define the *collider* as the ball that is further to the right at the moment of contact, and the other ball as the *collided* object. This convention is used only to identify which participant is treated as the upstream source of motion in the event token. Entity-level attribution is then assigned to the collider associated with that event.

The first extracted feature is *collision magnitude*, an unsigned measure of how much a collision turns the collided object. Let c denote a collision in which collider A strikes collided object B, and let $\hat{v}_B^-$ and $\hat{v}_B^+$ denote the normalized pre- and post-collision directions of B's motion. Collision magnitude for collision c is defined as

$$M(c) = \frac{1}{\pi} \arccos\big(\mathrm{clamp}\left(\hat{v}_B^- \cdot \hat{v}_B^+, -1, 1\right)\big)$$

This quantity maps no directional change to 0 and a full reversal of direction (180°) to 1. Because this term depends only on the size of the directional change, not on speed or target direction, the contribution of the collision to the eventual outcome is left to the second feature.

The second extracted feature is *mapping ease*, which captures how readily the consequences of an observed collision project forward to the outcome event. For each observed collision c, let s_c^+ denote the post-collision state of the scene immediately after that event. This state includes the positions and velocities of all objects after the collision has been resolved. From s_c^+, the model performs repeated forward simulations under uncertainty and computes the proportion of simulated trajectories in which the effect ball enters the goal. Formally, if O denotes the outcome event $miss \rightarrow goal$, mapping ease is defined as

$$E(c) = P(O \mid s_c^+)$$

and is estimated by Monte Carlo forward projection:

$$\hat{E}(c) = \frac{1}{N} \sum_{r=1}^{N} \mathbf{1}\{O \text{ occurs in rollout } r \text{ from } s_c^+\}$$

where uncertainty in the rollouts serves as a proxy for the observer's uncertainty at that timepoint. Importantly, these forward simulations are not counterfactual interventions on candidate causes. The observed collision is held fixed, and the model asks only how apparent the eventual outcome becomes from that actual post-collision state onward. Each support collision is therefore evaluated from its own post-collision state.

Because a single candidate ball may contribute more than one support collision, the implementation aggregates collision-level features to the ball level. For candidate ball i, the exported collision-magnitude feature is the maximum collision-magnitude value across support collisions in which i serves as the collider:

$$M(i) = \max_{c:\,\mathrm{collider}(c)=i} M(c)$$

Likewise, the exported mapping-ease feature for ball i is the maximum mapping-ease value across its support collisions:

$$E(i) = \max_{c:\,\mathrm{collider}(c)=i} \hat{E}(c)$$

In addition, the implementation defines a deterministic support gate,

$$G(i) = \begin{cases} 1 & \text{if candidate } i \text{ is involved in a supporting collision} \\ 0 & \text{otherwise} \end{cases}$$

The final descriptive form of the RICK model is a linear response model over these two predictors. Let $J(i)$ denote the predicted causal responsibility judgment for candidate ball i. Then

$$J(i) = G(i)\,[\beta_0 + \beta_M M(i) + \beta_E E(i)]$$

where β_0 is an intercept and β_M and β_E are the weights on collision magnitude and mapping ease, respectively. The gate ensures that candidates with no outcome-relevant support collision receive predicted causal responsibility of zero. In the present implementation, the free weights are estimated by regression against participant attribution judgments. The noise parameter used in the forward projections for mapping ease is not estimated from these causal-responsibility judgments, but from a separate prediction task described below.

This formulation is more parsimonious than the CSM. Whereas the CSM requires separate weights for each of its causal dimensions, RICK requires only two feature weights.

This specification yields the qualitative patterns described above. Proximal causes receive strong support when the collision that directly changes the effect ball's motion produces a substantial directional update and the outcome is highly projectable from that point. More distal causes are supported insofar as their observed collisions substantially alter downstream motion and those consequences remain recoverable under forward projection. Because multiple upstream collisions can contribute to the same downstream ball, a late nudge need not erase an earlier steering event; rather, each support collision is evaluated separately before the ball-level predictors are aggregated. When an earlier collision produces a distinctive change in motion that is preserved through later events, the root cause may receive the bulk of attribution. When that mapping becomes difficult, support for distal causes weakens and the proximal cause tends to dominate.

Experiment 1

In each scenario, participants provided causal responsibility judgments for one of the three candidate causes, randomly selected at each trial. We categorized the candidate causes relative to its reverse order in the causal sequence, such that the ball that actually makes contact with the effect ball (the proximal-cause) is the first difference-maker, the ball in the middle of the chain is the second difference-maker, and the ball at the start of the chain is the third difference-maker. Since only one candidate ball makes contact with the effect ball in the one difference-maker condition, the second and third ordering is arbitrarily selected.

Methods

Sample and Procedure We recruited 40 participants located in the United States from the Prolific academic platform. Participants completed a response-key training block in which four keys (F, SPACE, J, K) were mapped to three colored balls (red, blue, green) and the MISS outcome. In the experimental task, participants viewed animations of three colored balls moving leftward toward a stationary grey "effect" ball positioned in front of a goal. In each trial, one colored ball ultimately caused the grey ball to enter the goal. Participants selected which colored ball 'caused' the outcome. After making a selection, they provided causal responsibility judgments for each candidate ball on a 0–100 scale indicating agreement with the statement "The color ball caused the grey ball to go through the goal."

In a final section, participants viewed simplified trials involving a single candidate cause and the effect ball. Each animation terminated when the distance between the balls reached 50% of its initial value, after which participants predicted whether the outcome would be a goal or a miss. Participants completed 20 such trials, which were used to estimate the noise parameter used by the CSM and in RICK's forward projections.

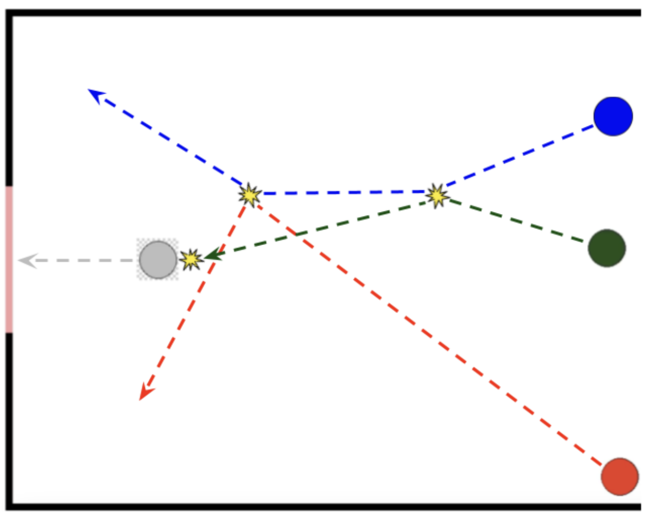

Figure 1: A 3 difference-maker dynamic collision event.

Results and Discussion

To test whether judgment time differed as a function of the number of difference-makers, we fit a linear mixed-effects model predicting the time to causal judgment (JT) from the number of difference-makers (DM) and preemption ($PREEMPT$), including baseline reaction time (BRT), stimuli duration (DUR), and time since last collision (TLC) as covariates. Each participant contributed 18 trials (6 per condition), and repeated measures were handled by a participant random effect. From this model, we obtained estimated marginal means (EMMs) of JT for each level of DM. Our primary hypothesis test consisted of 3 contrasts derived from these EMMs: $\{DM_2 - DM_1, DM_3 - DM_2, DM_3 - DM_1\}$. Covariate-adjusted EMMs for the three DM levels were: DM_1 = 2.70 [2.42, 2.98], DM_2 = 2.78 [2.56, 2.99], DM_3 = 2.64 [2.41, 2.88] (see **Figure 2**). Comparing judgment time for each pair of DM levels, we found no significant differences: $DM_2 - DM_1$: $t(12.1) = 0.44$, $p = .67$, $DM_3 - DM_2$: $t(12.7) = -1.20$, $p = .25$, $DM_3 - DM_1$: $t(12.7) = -0.33$, $p = 0.75$. Taking the CSM as a process model, judgment time should increase with the number of difference-makers, since each additional difference-maker requires an additional counterfactual comparison. We did not observe this pattern. One possibility is that the relevant simulations are too fast to detect in the present design. However, this explanation weakens the psychological force of the CSM: in other mental simulation tasks, increased simulation demands produce measurable reaction-time costs, as in mental rotation (Provost & Heathcote, 2015), and imagined object tracking appears to be strongly capacity-limited, with even two moving objects producing a serial bottleneck (Balban & Ullman, 2025). Thus, although the null reaction-time result is not decisive, it provides little positive support for the claim that participants explicitly simulated each candidate counterfactual before making their causal judgment.

To quantify comparative model fit, we computed the absolute prediction error for each participant judgment relative to each model's predicted attribution and fit a linear mixed-effects model with fixed effects of model (RICK vs. CSM), number of difference-makers, and their interaction, together with random intercepts for participant and stimulus. This analysis revealed a strong main effect of model, $\chi^2(1) = 524.64$, $p < .001$, qualified by a robust model × difference-maker interaction, $\chi^2(2) = 295.78$, $p < .001$. Descriptively, mean absolute error was substantially lower for RICK than for the CSM overall (RICK: 11.9; CSM: 26.6). The two models were statistically indistinguishable in the one-difference-maker condition ($z = -0.89$, $p = .37$), but RICK showed dramatically lower error in the two-difference-maker condition (RICK: 11.8, CSM: 32.3; $z = -19.98$, $p < .001$) and in the three-difference-maker condition (RICK: 18.6, CSM: 42.8; $z = -23.11$, $p < .001$). These results are consistent with **Figure 3**: the CSM diverges sharply from human judgment once multiple candidate causes interact, whereas RICK preserves the concentration of attribution on the proximal cause much more effectively. Relative to the CSM, this yields a more parsimonious response model with fewer fitted weights while also providing substantially better fit in the multi-difference-maker conditions.

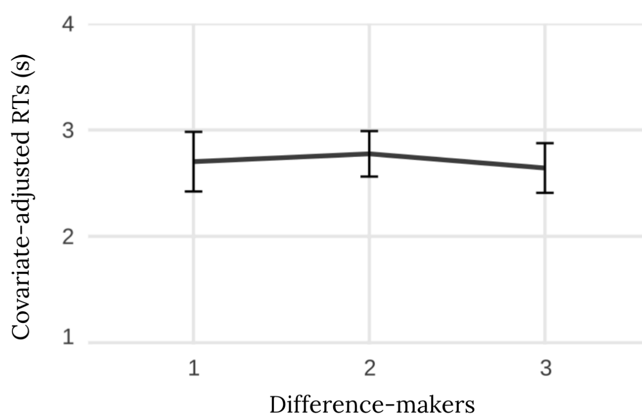

Figure 2: Covariate-adjusted causal judgment times (EMMs) for each difference-maker condition.

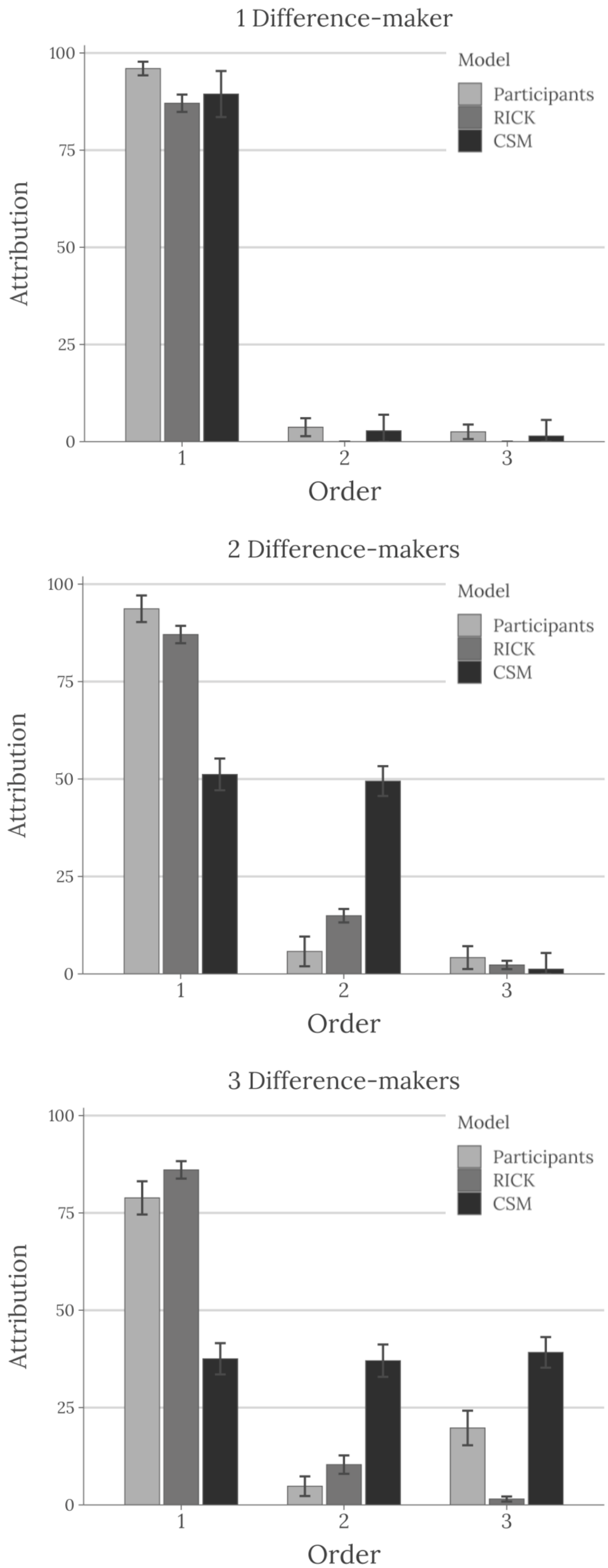

Figure 3: Model and participant attribution patterns across conditions.

Conclusion

We argue that counterfactual simulation is not necessary for explaining actual-cause judgments in dynamic collision events. Instead, observers can directly apply type-level causal knowledge directly to observed token events, using forward projection to determine how clearly each collision maps onto the outcome. We offer only a limited exposition for the RICK account, and future work will extend the model to a wider breadth of stimuli, and eventually to other domains beyond collision scenarios.

Materials Availability

All materials, code, and data associated with this project are available at: `https://github.com/rahalejr/rick`

References

Aronowitz, S., & Lombrozo, T. (2020). Learning through simulation. *Philosophers' Imprint*, *20*.

Balaban, H., & Ullman, T. D. (2025). The capacity limits of moving objects in the imagination. *Nature Communications*, *16*(1), 5899.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, *104*(2), 367–405.

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.

Halpern, J. Y., & Hitchcock, C. (2013). Compact representations of extended causal models. *Cognitive Science*, *37*(6), 986–1010.

Halpern, J. Y., & Pearl, J. (2005). Causes and explanations: A structural-model approach—part i: Causes. *British Journal for the Philosophy of Science*, *56*(4), 843–887.

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

Lewis, D. (2000). Causation as influence. *Journal of Philosophy*, *97*(4), 182–197.

Menzies, P., & Beebee, H. (2025). Counterfactual theories of causation [Fall 2025 edition]. In E. N. Zalta & U. Nodelman (Eds.), *The stanford encyclopedia of philosophy*.

Pearl, J. (2000). *Causality*. Cambridge University Press.

Provost, A., & Heathcote, A. (2015). Titrating decision processes in the mental rotation task. *Psychological Review*, *122*(4), 735–754.

Quillien, T., & Lucas, C. G. (2024). Counterfactuals and the logic of causal selection. *Psychological Review*, *131*(5), 1208–1234.

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