Zero-Sum Structure of Everyday Causal Attribution

Greta Mohr[1], Shasha Alvares[2], David A. Lagnado[3] & Samantha Kleinberg[4]
[1,3]Department of Experimental Psychology, University College London, 26 Bedford Way, London, WC1H 0AP, UK
[2,4]Computer Science Department, Stevens Institute of Technology, 1 Castle Point Terrace, Hoboken, NJ, 07030, USA

Abstract

Causal explanations for specific events (token causes) are widespread in daily life, from finding causes of a patient's symptoms to understanding why a policy was effective. Despite the importance of token causality in a wide-range of real-world scenarios, most of what is known about how people attribute causality is based on simple physics or highly constrained study set-ups. It is not yet known how people attribute responsibility when faced with both observed and unobserved factors. To address this, we conducted two experiments to examine how beliefs influence causal attribution and what information people infer indirectly. In both experiments, we found that people treated causality as a fixed resource to be split between observed and other factors, and that outcome valence influenced both causal and counterfactual judgments. On the other hand, prior beliefs about actions did not predict causal attributions.

Keywords: token causality; causal explanation; diabetes

Introduction

Causal explanations for events are a critical way we learn about and interact with the world. We ask why a patient's symptoms improved, why a car accident happened, or why a policy was effective. All of these questions target token causes, that is, causes of specific events occurring at a particular time and place rather than causes of illness, car accidents, or successful policies in general. Despite the prevalence of such reasoning, it is a difficult undertaking as we lack complete information and must rely on background knowledge to make sense of events and their alternatives. Understanding how people successfully perform this task is critical for understanding human reasoning and aiding it, yet there is no comprehensive theory of how people attribute token causality.

Prior work suggests that people largely think about token causality in terms of counterfactuals, comparing what happened to what could have happened had events unfolded differently (Kahneman & Miller, 1986; Kominsky et al., 2015). People reliably identify some events as more responsible for an outcome than others (e.g., a dropped match rather than presence of oxygen causing a forest fire), and cognitive science has aimed to understand how people make such attributions. The counterfactual simulation model (CSM) suggests people mentally simulate alternatives, and attribute causality to events that made a difference to whether and how an outcome occurred (Gerstenberg, 2024; Gerstenberg et al., 2015). However, it is an open question why people consider some counterfactuals and not others. It has been suggested that people consider how surprising or abnormal an event is (Halpern & Hitchcock, 2015; Icard et al., 2017), and consider counterfactuals that include normal events (e.g., oxygen present) rather than abnormal events (e.g., dropped match). Systematic variation of underlying event probabilities, though, shows that whether people prefer the normal or abnormal event varies with causal structure and the probabilities of alternatives (Morris et al., 2019). Norm violations, epistemic states, and intentionality also play an important role in how people assess both cause and blame judgments (Kirfel et al., 2022).

However, much of this research has focused on controlled set-ups where participants have full knowledge. The CSM is based on attributing causality for physical events (e.g., billiard balls colliding) and other studies use examples where participants receive complete information about causes (e.g., made-up diseases, drawing colored balls, etc.) with simple structures. Participants may use their knowledge of physics to predict where a ball may move, but this is fundamentally different from everyday events where the alternatives are less apparent, and outcomes may be more challenging to envision. For example, when considering alternatives to an individual exercising, one might consider a different form of activity, different duration or intensity, or a different schedule.

Counterfactual and causal judgments can dissociate, though, so counterfactual judgments alone might not be sufficient for understanding causal attributions (Alicke et al., 2015; Mandel, 2003). Recent work examining explanation of everyday events replicated this finding and further showed that people attribute causality to abrupt and gradual events differently, and this differed by domain, with gradual causes of health being considered more responsible than abrupt ones (Cheung et al., 2025). This suggests that it is critical to study token causality with realistic stimuli and understand how people use background knowledge to enrich the materials provided.

Taken together, existing research on token causality suggests people consider alternatives to how events occurred, and use these to guide causal judgments. However, there is not yet a theory that can explain which counterfactuals people consider, or how they allocate responsibility across events. We aimed to address this gap by examining 1) whether a priori beliefs about actions (e.g., surprise, difficulty) influence causal attribution, 2) how people attribute causality to both observed and unobserved factors, and 3) what assumptions and inferences people make beyond the information given. To do this we conducted two experiments explaining a real-world example, focusing in Experiment 1 on explanation with a single cause, and in Experiment 2 with multiple causes of an event.

Experiment Overview

We aimed to examine token causal explanation in a realistic scenario that is representative of how people attribute causality in daily life. To this end we used a vignette about reducing risk of diabetes with lifestyle changes, since the disease is prevalent, and people receive information about diet and exercise and can be expected to have beliefs about both. Experiment 1 focused on causal attribution with a single action, providing a case where alternative causes are implicit. Experiment 2 examined multiple actions within the same vignette.

Experiment 1

While causal attribution has been studied in constrained settings, it is still not known how people determine causes of everyday events, where they lack complete causal knowledge. To address this, we first focus on the case of a single action, to examine how expectations influence explanations and the counterfactual alternatives people imagine.

Method

Participants We recruited 144 U.S.-based adult participants via Prolific (M_{age} = 42.9, SD_{age} = 14.5). They were 52% female, 46% male, 2% nonbinary; and mainly identified as White (75.7% White, 13.9% Black or African American, 9.7% Asian, 13.9% Hispanic/Latino, 0.7% Native Hawaiian/ Pacific Islander) with varied diabetes experience (Type 1 diabetes = 0.7%; Type 2 diabetes = 13.9%; Not sure what type = 2.1%; caregiver for someone with diabetes = 26.4%, no experience 24.3%, and other experience 43.7%). An *a priori* power analysis using effect sizes from a previous study (Cohen's $d \approx 0.80$), indicated 22 participants/group was required for 95% power (α = .05). We oversampled to enable exploratory analyses (36 participants/group). Participants provided informed consent and were paid at a rate of £6/hour. All participants were included in analyses.

Materials We developed a base vignette that included context (green), an action (blue), and an outcome (pink) and was presented across two screens.

Screen 1:

> John's father had type 2 diabetes, so when John's own routine A1C blood test came back slightly elevated, he felt concerned that he was beginning to develop it as well. (An A1C test shows your average blood sugar level over the past 2–3 months). For several months following the test results, John began choosing water instead of sugary drinks during the day.

Screen 2:

> When he returned to the doctor several months later, his blood sugar levels were lower.

There were four versions of the vignette (2 action x 2 outcome). The actions were replacing sugary drinks with water (drink action) and joining a community fitness class once a week (exercise action). The outcomes were blood sugar being lower (good outcome) or unchanged (no change outcome).

Procedure We used a between-subjects design with participants each seeing one vignette. They made predictions before seeing the outcomes, then after made causal and counterfactual judgments.

Outcome predictions After seeing the context and action (screen 1) participants made predictions using sliders from 0-100:

- How likely is it that John's blood sugar will get better or get worse?
- How surprised are you that John chose this action?
- How difficult do you think it will be for John to make this lifestyle change?
- How difficult do you think it will be for John to sustain this lifestyle change?

Causal attributions Next, participants saw the outcome (pink) and made causal judgments on a scale from "did not cause it at all" (0) to "caused it entirely" (100) for the action (drinks or exercise) and for other factors (i.e. *To what extent did other factors cause John's blood sugar to be lower?*). We used a free text box to collect what other factors participants had in mind. Participants further rated how likely it is John has diabetes, and if this would be more or less likely if his father did not have diabetes.

Counterfactual judgments Finally, they made counterfactual judgments, rating 0-100 whether the outcome would have been less or more likely if John had done the following actions:

- replaced his usual sugary drinks with water and started weekly exercise?
- only started weekly exercise but not replaced his usual sugary drinks with water?
- replaced his usual sugary drinks with water but also started smoking regularly?
- made no changes to his lifestyle?

The highlighted items show text for the drinks condition, and were replaced with similar exercise text for that condition. Finally, participants were asked "Are there any other factors you think could have changed the outcome?" (free text response).

Participants then completed a demographic questionnaire with age, gender, ethnicity, diabetes experience, and self-rated knowledge about diabetes on a 0–10 scale.

Data Analysis Statistical analyses were conducted in MATLAB (2025b). Group differences between actions were assessed using independent-samples t-tests with unequal variances (Welch's t-tests). Linear regression analyses were performed using MATLAB's `fitlm` function. All continuous predictors were z-scored prior to modeling. The reported coefficients β reflect the standardized regression weights.

The free-text questions were coded independently by two coders using a binary procedure (code present/absent for a response), with disagreements discussed. Mean inter-code reliability was 93%. The categories for what other factors may have caused the outcome were: 1) diet changes, 2) exercise changes, and 3) other influences. The categories for the counterfactual question on other factors that may have changed the outcome were 1) multiple actions, 2) specific alternative action/cause, and 3) uncertainty or lack of information.

Results

Outcome Predictions As shown in Figure 1, participants rated that John's health was more likely to improve with the drinks action, compared to the exercise action ($t(142) = -2.63$, $p = .03$). Participants rated the exercise action as more difficult to sustain compared to the drinks action ($t(142) = 2.63$, $p = .009$) but there was no significant difference in difficulty of making each change ($t(142) = 1.89$, $p = .06$) or surprise ($t(142) = 1.66$, $p = 0.10$).

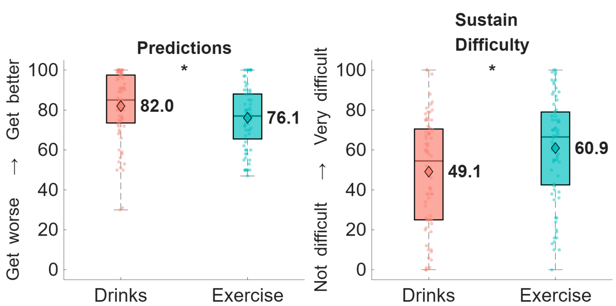

Figure 1: Experiment 1 (from left to right panel) outcome predictions and sustain difficulty ratings for each action. Boxes show medians and inter-quartile ranges; diamonds indicate means. *p< .05

Causal Attributions As shown in Figure 2, outcome type influenced how participants attributed causality for actions and other factors. A linear model showed that causal ratings for actions were higher for good outcomes ($\beta = 1.44$, $p < .001$) and were negatively predicted by causal ratings for "other factors" ($\beta = -0.17$, $p = .003$). Causal ratings were not predicted by the specific action ($\beta = 0.09$, $p = .37$), change difficulty ($\beta = -0.05$, $p = .39$), or sustain difficulty ($\beta = 0.06$, $p = .22$).

Counterfactual Judgments Participants rated counterfactuals of how much worse or better the outcome would have been than what was observed had John made no lifestyle changes, taken both actions, taken only the other action, or taken the original action and smoked. Ratings of 50 indicate no change in outcome. Figure 3 shows participants rated the counterfactual with no lifestyle changes differently depending on the original action and observed outcome (higher for exercise: $\beta = 0.45$, $p = .04$, lower for observed good outcome:$\beta = -0.87$, $p < .001$), with no action × outcome

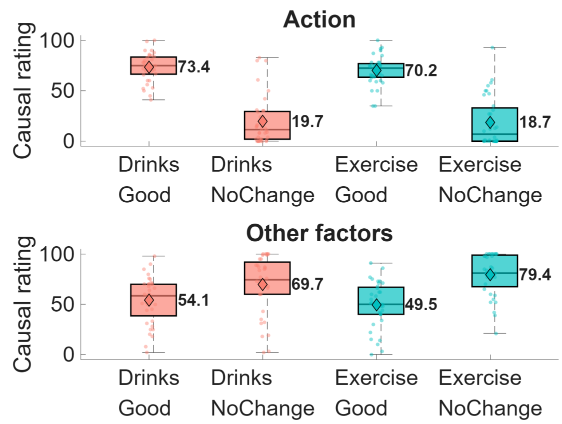

Figure 2: Causal attribution ratings for focal action (top) and 'other factors' (bottom). Diamonds indicate means.

interaction ($\beta = -0.49$, $p = .12$). When they saw a good outcome, participants judged that having taken both actions would have led to an even better outcome (M =61.8, $\beta = 0.65$, $p < .001$), while in the no change condition ratings indicated a similar outcome was expected ($M = 47.2$) and there was no effect of original action ($\beta = 0.09$, $p = .58$). Participants did not imagine that taking the other action would have changed the outcome (outcome: $\beta = 0.11$, $p = .50$, action: $\beta = -0.00$, $p = .98$). When participants imagined adding smoking to the original action, there was no difference by outcome ($\beta = -0.08$, $p = .65$) or action ($\beta = -0.17$, $p = .32$) and responses clustered near the midpoint.

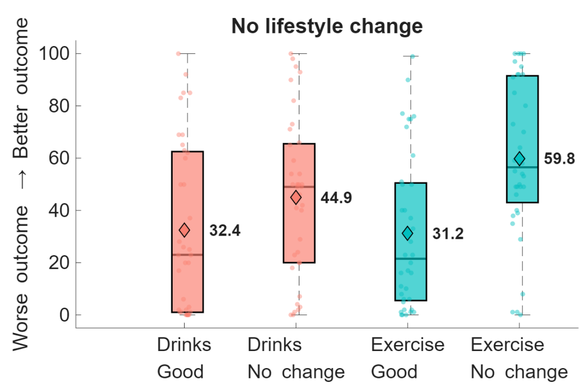

Figure 3: Counterfactual outcome ratings, shown by action and outcome. Diamonds indicate means.

Diabetes Likelihood Estimations Participants rated that John was likelier to have diabetes in the no change condition ($\beta = -0.87$, $p < .001$) and when they rated his action as more difficult to sustain ($\beta = 0.31$, $p < .001$). There was no effect of initial outcome predictions ($\beta = 0.04$, $p = .56$) or action ($\beta = 0.06$, $p = .70$). When participants were asked if John was more or less likely to have diabetes if his father did not, initial diabetes likelihood was a positive predictor ($\beta = 0.20$, $p = .043$) and outcome expectation was a negative predictor ($\beta = -0.24$, $p = .005$). There was no effect of sustain difficulty ($\beta = -0.08$, $p = .38$), action ($\beta = -0.15$, $p = .38$), or outcome ($\beta = 0.07$, $p = .69$).

Free text responses

Participants frequently reasoned beyond the described actions, drawing inferences about other actions and additional causes. Of the 122/144 responses to what other factors may have caused the outcome 67.1% mentioned diet changes, 26.7% mentioned exercise changes, and 49.2% mentioned other factors. For the good outcome, participants made assumptions about additional changes beyond those described ("*he probably ate a little better too and that helped*" and "*perhaps he started exercising more and eating healthier*"). For the no change outcome, participants mentioned aspects of the actions such as frequency ("*[exercising] once a week is not enough*") and factors such as genetics.

For the counterfactual question on what other factors could have changed the outcome, 88/144 participants responded. Of these, 69.55% mentioned specific alternative actions/causes, 37.4% mentioned multiple actions, and 10.9% mentioned uncertainty/lack of information. Some of the specific alternative actions/causes participants mentioned were dietary factors, weight and activity ("*sedentary lifestyle*"), medication ("*maybe taking a GLP1*"), genetics, stress, and sleep.

Discussion

We found that causal judgments reflected a zero-sum structure, with increased attribution to other factors corresponding to reduced attribution to the focal action. This effect was stronger for no change outcomes. While participants made different predictions for each action, they attributed the same degree of causality to both. The difference in counterfactual judgments when the actual outcome was good versus no change shows that people infer additional information about what happened in the actual world (beyond that given in the scenario) and carry this over to their counterfactual inference. For example, when John exercised and the actual outcome was good, people judged the counterfactual outcome (if John had made no lifestyle change) to be worse, whereas they rated it better when the actual outcome was no change. The difference must lie in additional inferences made as a result of the good versus no change outcome in the actual world, perhaps inferences about why John's condition persisted despite exercise. This finding shows that people's counterfactual inferences are not the same as hypothetical predictions because they incorporate extra inferences about what occurred in the actual world.

Experiment 2

Participants in Experiment 1 seemed to allocate a fixed amount of causality between the action and other factors, and judged counterfactuals differently based on observed outcomes. We designed Experiment 2 to test this further by presenting both actions within the same vignette and adding questions to probe the difference between predictions and counterfactuals.

Method

Participants We used the effect sizes of Experiment 1 (difference in prediction between drinks and exercise $g = 0.36$) to inform a power analysis. We determined ≈97 participants per action condition were needed, and targeted 100 participants per condition. We recruited 200 U.S.-based adult participants via Prolific ($M_{\text{age}} = 45.1$, $SD_{\text{age}} = 15.1$). Participants were majority male (54.0% male, 45.0% female, 0.5% nonbinary and 0.5% prefer not to say) and mainly identified as White (68.0% White, 16.5% Asian, 10% Black or African American, and 6% Hispanic, Latino/a, or Spanish origin, 0.5% Native Hawaiian/ Pacific Islander) with varied diabetes experience (Type 1 diabetes = 3.5%; Type 2 diabetes = 8.0%; Not sure what type = 1.0%; caregiver for someone with diabetes = 26.0%, no experience 33.5%, and other experience 36.0%). All participants were included in analyses.

Materials We used the same overall design as in the first Experiment, and updated the vignette so it now described both changes. We revised the exercise action to be "regularly" rather than once a week as some participants in Experiment 1 noted this may be insufficient. The updated section is:

> For several months following the test results, John began regularly choosing water instead of sugary drinks during the day. John also joined a community fitness class recommended by a friend and attended regularly.

Procedure We again used a between-subjects design with 4 groups (2 outcomes x 2 action orders; N DrinksExercise good outcome: 49, no change: 51; ExerciseDrinks good outcome: 50, no change: 50). The outcomes were as in Experiment 1, and participants were randomized to see the drinks action first or exercise first (swapping the order of the blue sections above).

Participants made the same outcome prediction ratings as in Experiment 1, with the difference being they now rated surprise, change difficulty, and sustain difficulty for each action (6 ratings total). After the outcome was revealed, participants rated to what extent did each action and other factors cause the outcome? (0-Did not cause at all to 100-Caused entirely). We again provided a free text box and elicited what other factors participants had in mind.

Participants then made counterfactual judgments about whether the outcome would be more or less likely if John had made only one of the changes (drinks, exercise) or no lifestyle changes. Finally, we added free text questions to understand what information participants inferred:

- What else do you think changed as a result of John cutting out sugary drinks/starting to exercise regularly?
- What else had to change for John to make the lifestyle changes?
- What else would have led to a different outcome?

The demographics questions were the same as Experiment 1.

Data Analysis

We used the same data analysis method as in Experiment 1. Because Experiment 2 involved a mixed design, we additionally conducted within-participant comparisons using paired-samples t-tests.

Free-text responses were again independently coded by two coders. Inter-code reliability ranged from 98% to 99.5% (M=98.9%). The codes for the four free-text questions were:

Other causes: diet changes, other influences, genetics

Side effects: physical health/weight loss, energy level changes, diet changes, other factors

Supporting factors: motivation and mental attitude, diet changes, routine changes, other influences

Causes of different outcome: diet changes, medication, other influences

Results

Outcome Predictions We combined both action orders for analysis, as an independent-samples t-test showed no difference in outcome ratings between the two action order conditions ($t(198) = 0.15$, $p = .88$, ExerciseDrinks M =84.2, DrinksExercise M = 84.5). As shown in Figure 4, though, participants rated the exercise action as being a more difficult change to make ($t(199) = -3.78$, $p < .001$), more difficult to sustain ($t(199) = -6.14$, $p < .001$), and more surprising ($t(199) = -2.20$, $p = .029$).

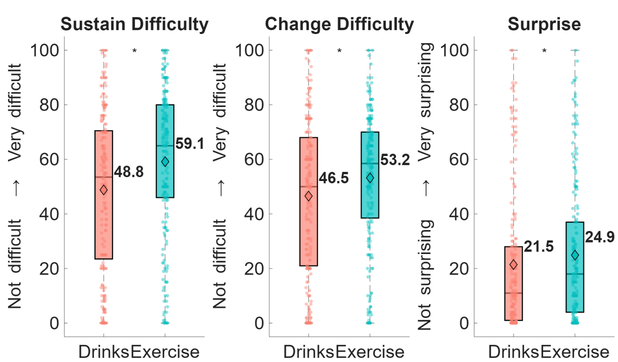

Figure 4: Experiment 2 predictions for each action. Diamonds indicate condition means. $*p < .05$.

Causal Attributions Figure 5 shows causal attributions for each action and "other." We again found that for the good outcome participants attributed more causality to observed actions, while for the no change outcome, participants attributed more causality to other factors. There was no difference between causal ratings of the two actions for either outcome (good: $t(98) = 0.99$, $p = .32$; no change: $t(100) = 0.22$, $p = .83$), and causal ratings for exercise and drinks were significantly positively correlated for both good ($r = .68$, $p < .001$) and no change outcomes ($r = .93$, $p < .001$). However, causal attributions to exercise and drinks were both negatively correlated with attributions to other factors (exercise: $r = -0.48$, $p = .001$; drinks: $r = -0.44$, $p = .001$).

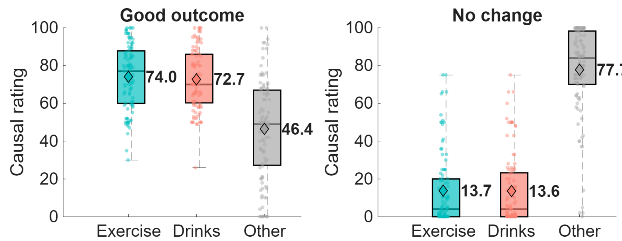

Figure 5: Experiment 2 causal attribution ratings for good (left panel) and no change outcomes (right panel). Diamonds indicate condition means.

Counterfactual Judgments Figure 6 shows counterfactual ratings. For the good outcome, when participants imagined John taking only one of the two actions, the drinks action was rated higher than exercise ($t(98) = -3.01$, $p = .003$). Both actions were still rated better than making no lifestyle change (only exercise: $t(98) = 9.12$, $p < .001$; only drinks: $t(98) = 11.08$, $p < .001$). For the no change outcome, there was no difference in rating between only drinks and only exercise ($t(100) = 1.82$, $p = .072$). Making no lifestyle change was rated better than only drinks ($t(100) = -2.60$, $p = .011$) and was not rated as significantly different than only exercise ($t(100) = -1.92$, $p = .057$). All ratings were lower for good outcomes (no lifestyle change: $t(198) = -10.31$, $p < .001$; only drinks: $t(198) = -2.89$, $p = .004$; and only exercise $t(198) = -5.42$, $p < .001$).

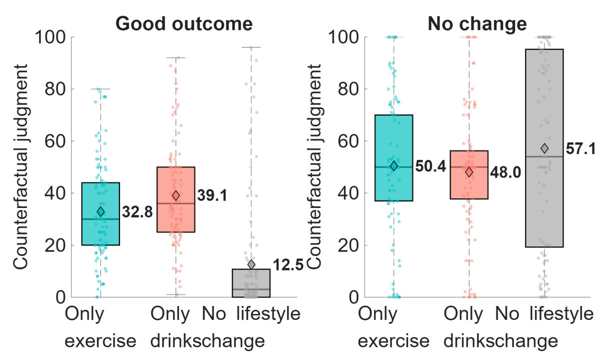

Figure 6: Experiment 2 counterfactual ratings shown separately for good (left panel) and no change (right panel) outcome conditions. Diamonds indicate condition means.

Free text responses All participants (N = 200) provided free-text responses. Examining other factors participants thought contributed to the outcome, similar to Experiment 1, 46.8% mentioned additional dietary changes ("*eating less desserts too*") and 49.7% mentioned other factors such as sleep, medicine, or untreated conditions. Notably, there was a difference in mentions of genetics between the good outcome (4%) and no change outcome (44.5%).

Our additional questions aimed to understand what changes participants thought may have been implicit. Considering what other effects may have followed from John's actions, 48.3% mentioned weight loss or improved fitness, and 14.8% suggested he had "*increased energy*". For good outcomes, 23.7% mentioned additional positive dietary changes such as lower calorie intake, while for no change outcomes 13.4% mentioned possible negative changes, including "*snacking more*" or "*more sugar from other foods*". For changes needed alongside the focal action 43% mentioned motivation and mental attitude ("*his mindset is key*"), 21.4% highlighted the need for lifestyle or routine changes. For causes that could have led to a different outcome, 50.2% mentioned diet changes ("*cutting out processed foods and unnatural sugars*"), and 17% mentioned "*taking medication*" and other influences such as "*keeping doctor appointments*".

Discussion Participants' causal attributions depended largely on the observed outcome. That is, while there were different pre-outcome beliefs about exercise and replacing sugary drinks, causal attributions afterward were highly correlated. Further, attribution shifted toward "other causal factors" in a zero-sum fashion depending on the outcome. We again observed that participants made extra inferences about what happened based on the outcome: believing other positive changes were made to help bring about the good outcome, while for the no change outcome they inferred other factors that may have interfered with John's lifestyle changes.

General Discussion

We found that 1) people reliably attribute different degrees of causality to events, 2) causal attribution is treated as distributing a fixed amount of credit, and 3) outcome valence influences both causal and counterfactual judgments.

Participants split causality between observed and unobserved factors: when attributions were higher for observed causes, they were lower for unobserved ones, and vice versa. Notably, in Experiment 2 causal attribution for drinks and exercise were strongly correlated, suggesting that people may be viewing them as one conjunctive cause. Balancing causes such that they add up to a whole provides further support for the "causal pie" view in epidemiology (Rothman, 1976). Further, both causes in Experiment 2 receiving high ratings suggests participants recognized that the full 0 to 100 scale could be used independently for each action (e.g., average causal attribution for exercise and drinks added up to 147).

We found significant differences in how people judged causality and counterfactuals depending on outcome valence. For example, when John took up exercising but had no change in his glucose in Experiment 1, participants judged that he would have been better off had he not exercised, and this differed from counterfactuals when the same action resulted in a good outcome. Participants treated successful actions as causally efficacious, predicting that adding a second action would further improve outcomes, whereas no change outcomes were seen as inevitable. Further work is needed to understand this valence effect. We hypothesize that the no change outcome may violate expectations, leading to extra assumptions about other factors (e.g., genetic factors making exercise ineffective for John). This difference in counterfactual judgments when the actual outcome is good versus no change also shows that people distinguish between a genuine counterfactual (John exercised and improved; what if he had not exercised?) and a hypothetical (what if John did not exercise?). This important difference shows the added complexity of counterfactual inference (Pearl & Mackenzie, 2018).

A key open question for future work is to systematically examine what inferences people make about support factors, alternative causes, and side effects and how these influence causal attribution and distinguish it from counterfactual judgments. In particular, respondents used the outcome to make inferences about other actions John took. Disentangling alternative causes versus support factors may help further elucidate the causal pie hypothesis. Further, as we focused on the health domain, given the role of background knowledge, it is important to examine how results vary across domains.

Conclusion

Causal attribution for everyday events is critical but not well-understood. Through two experiments we found people treat causality as a fixed resource to be allocated between observed and unobserved factors, and that outcome valence reliably shaped causal and counterfactual judgments. These results suggest a need to study token causality with naturalistic stimuli and better understand the potential interaction between expectations, outcome valence, and causal attribution.

Acknowledgments

This work was supported by a grant from the John Templeton Foundation.

References

Alicke, M. D., Mandel, D. R., Hilton, D. J., Gerstenberg, T., & Lagnado, D. A. (2015). Causal conceptions in social explanation and moral evaluation: A historical tour. *Perspectives on Psychological Science*, *10*(6), 790–812.

Cheung, V., Leone, C., Lagnado, D., & Kleinberg, S. (2025). Causal and counterfactual reasoning about gradual and abrupt events. *CogSci*.

Gerstenberg, T. (2024). Counterfactual simulation in causal cognition. *Trends in Cognitive Sciences*, *28*(10), 924–936.

Gerstenberg, T., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2015). How, whether, why: Causal judgments as counterfactual contrasts. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *37*.

Halpern, J. Y., & Hitchcock, C. (2015). Graded causation and defaults. *The British Journal for the Philosophy of Science*.

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

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psych. Rev.*, *93*(2), 136.

Kirfel, L., Icard, T., & Gerstenberg, T. (2022). Inference from explanation. *JEP: General*, *151*(7), 1481.

Kominsky, J. F., Phillips, J., Gerstenberg, T., Lagnado, D., & Knobe, J. (2015). Causal superseding. *Cognition*, *137*, 196–209.

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

Morris, A., Phillips, J., Gerstenberg, T., & Cushman, F. (2019). Quantitative causal selection patterns in token causation. *PloS one*, *14*(8), e0219704.

Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. Basic Books.

Rothman, K. J. (1976). Causes. *American journal of epidemiology*, *104*(6), 587–592.