Adults' use of latent vs direct causes for counterfactual reasoning

Background

This study builds on a prior pre-registered experiment conducted in our lab (Study 1, June 2025: [link redacted for future double-blind review]) which investigated the hypothesis that adults use two representational spaces to organize mental, physical and bodily events. One space groups events based on underlying latent factors, and another space describes how specific events are directly causally connected to each other. Study 1 had two main findings regarding the structure of our representations of other agents' capacities: (1) adults represent 3 latent factors namely "the mind", "the mechanical body" and "the biological body" that organize mental events, actions, and physiological events, and (2) a distinct representation of how these events causally relate to each other.

An open question regarding the two representational spaces is what kind of content they articulate. One possibility is that these two spaces are intuitive causal-explanatory framework theories about how other people work. Prior work has shown that intuitive theories are common-sense, abstract, causal-explanatory frameworks that help us navigate the world (Gerstenberg & Tenenbaum, 2017). They help people reason about events from different domains: Intuitive psychology, for example, conceives of actions as causally connected to mental states, which allows people to explain, predict and intervene on other agents' minds and behaviors. Intuitive theories also help people reason about events within the same domain: children's intuitive understanding of the biological domain helps them explain, predict and plan interventions on bodily states to prevent states of illness and disease, and to infer that members of living species that share internal structure, not superficial perceptual features, will have similar properties (Degn et al., 2025; Gelman & Markman, 1987).

Here, we propose to test the hypothesis that the judgments that we measured in Study 1 constitute two distinct intuitive theories that help us make sense of other agents as mental beings, physical actors, and living systems in terms of (1) Latent Causes and (2) Direct Causes. If the representational spaces measured in Study 1 constitute intuitive theories, then even when not asked to, adults should recruit these beliefs for prediction, explanation, counterfactuals, and intervention. If these two spaces represent two distinct ways of reasoning causally about other agents, then we should be able to predict when people will use beliefs about Latent Causes vs Direct Causes.

In summary, in the current work we propose to test for the functional roles of the Latent Causes and Direct Causes frameworks as distinct intuitive causal theories in common-sense social cognition.

General Alternative Hypotheses

The two frameworks (Latent Causes and Direct Causes) support common-sense reasoning: We hypothesize that the two causal frameworks measured in Study 1 are the basis for our intuitive reasoning within and across the domains of mind, action and body. We predict that adults will spontaneously make use of these representations for inference, explanation, intervention, and counterfactual reasoning.

The two frameworks do not support common-sense reasoning: An alternative hypothesis is that while adults can report their beliefs about how mental events, actions, and physiological events are organized when explicitly asked to (like in Study 1), they do not make use of these beliefs during everyday social cognition.

Current Study

Prior work has shown that people's intuitive causal reasoning enables them to think counterfactually, about whether an outcome still would have occurred if some prior event hadn't happened (Gerstenberg, 2024; Lewis, 1973; Lombrozo, 2010; Woodward, 2005). This study examines whether participants use representations of direct causes (i.e. whether A can make B happen), more so than latent causes (i.e. what is the common cause of A and B) to make counterfactual judgments. Specifically, the study presents a sequence of events (i.e. a preceding event followed by a target event). And then it asks whether a **target** event would occur if the preceding event was absent, and manipulates whether the preceding event was a **cause** of the target event or was **similar** to the target event.

We constructed a stimulus set based on the items and results of Study 1. The set consists of 15 triads of items, each triad consisting of a **target** item (such as "get tired") and a choice set of two items: an event that was most **similar** (i.e. close by in the Sorting Task RDM, but far away in the Causal Task RDM) to the target, and an event that was most **causally relevant** (i.e. close by in the Causal Task RDM but far away in the Sorting Task RDM). Thus, we chose these items based on their joint causal and similarity distance from the target: similar items had lowest similarity distance and largest causal distance to the target, whereas causal items had lowest causal distance and highest similarity distance to target. We constrained the similar option to be from the same domain as the target (this was already true for 13/15 target items), and the causal option to be from a different domain (this was already true for 14/15 target items). The full set of 15 item triads and their respective distances are listed in Figure 1.

Target and Choice Items			
Domain	Target Item	Causal Choice (Causal Distance, Freesort Distance)	Similar Choice (Causal Distance, Freesort Distance)
mind	see something	take a walk (0.2, 0.58)	hear something (0.42, 0.18)
	hear something	take a walk (0.25, 0.58)	see something (0.44, 0.18)
	choose what to do	experience pain (0.2, 0.57)	remember something (0.15, 0.36)
	remember something	take a walk (0.24, 0.61)	think about something (0.07, 0.15)
	think about something	get sick (0.25, 0.65)	remember something (0.08, 0.15)
action	reach for something	become hungry (0.21, 0.56)	kick something (0.54, 0.4)
	sit down	experience pain (0.13, 0.58)	jump up and down (0.54, 0.3)
	jump up and down	see something (0.29, 0.6)	sit down (0.83, 0.3)
	kick something	think about something (0.37, 0.6)	sit down (0.7, 0.42)
	take a walk	think about something (0.26, 0.58)	sit down (0.77, 0.33)
body	get tired	jump up and down (0.1, 0.54)	feel scared (0.42, 0.4)
	become hungry	take a walk (0.18, 0.55)	feel scared (0.67, 0.44)
	feel scared	see something (0.12, 0.54)	get tired (0.51, 0.4)
	experience pain	kick something (0.13, 0.51)	get tired (0.43, 0.41)
	get sick	see something (0.38, 0.62)	get tired (0.39, 0.34)

Figure 1: Stimuli for the current study based on group-averaged responses from two separate samples (total N = 151). The first two columns list the target events and their domains. The remaining two columns list the two options (the “causal choice” vs “similar choice”) associated with that target, and their similarity and causal distance from the target.

Hypothesis

We hypothesize that when people reason counterfactually, they rely on representations of direct causes (i.e. whether A can make B happen), more so than representations of latent causes (i.e. whether A and B are both caused by a third variable). If this is true, then when people see two events occur in sequence (A and B), they will be more likely to judge that B would not have happened if A hadn’t happened if A is a direct cause of B, than if A and B share a common cause.

Dependent Variable

Probability of target event occurring in the counterfactual (0-100%).

Independent Variable

Whether the omitted preceding event is **causal** vs **similar.**

Procedure

The experiment will have two within-subjects conditions. Subjects will be presented with the following instruction page:

> In this game, you will be asked to imagine what would happen in an alternative scenario. On each trial, there will be an observed event (for example, “a person took a walk, and then saw something”), and then you will be asked to consider whether the outcome (i.e. seeing something) would still happen if the first event (i.e. taking a walk) did not happen.

The next pages will present the two conditions, with 15 trials in each condition, but all trials will be randomly interspersed. Across both conditions, trials will consist of a test question involving a similar or casual event presented with a target, and a slider response scale.

Each trial presents either a similar event or a causal event, as follows:

> A person {similar event OR causal event}, then they {target event}. Suppose that they did not {similar event OR causal event}. Would they still have {target event}?

Each of the 15 targets will be shown twice, once with a preceding causal event, and once with a preceding similar event. Below the test question is a continuous slider scale with 5 labels: Definitely not, Maybe not, Unsure, Maybe yes, and Definitely yes. The 30 trials will be presented in shuffled order.

There will be two randomly interspersed attention checks, and participants will be excluded if they failed at least one attention check.

After the 30 trials, subjects will be presented with a debrief followed by a demographics questionnaire.

Analysis Plan

To analyze the effect of the preceding causal event on counterfactual judgments, we will run a linear mixed effects model usin,g the lmer() function from the lmerTest R package. The model predicts people’s judgments about the probability of target occurring in the counterfactual case, when the preceding event didn’t happen, based on whether the preceding event is similar or causal, while accounting for repeated observations within subjects. The model specification will be : **target_likelihood ~ preceding_event_type + (1 | subject_id)**.

We predict that people will produce lower probability judgments about the target still occuring, if the preceding event hadn’t happened, when the preceding event was causal than when it was similar to the target. Our threshold for statistical significance will be p = .05, two-tailed, and we will use the check_model() function from the performance package (Lüdecke et al., 2021) to conduct quality assurance.

Sample Size and Stopping Rule

Power analysis for Confirmatory Analysis

Based on a pilot study (N = 100), we used the R package *simr* to estimate the sample size needed to achieve 80% power to replicate the two model coefficients for the target_likelihood ~ preceding_event_type + (1 | subject_id) model: the intercept (i.e. the mean likelihood for the reference group, the similar preceding event, Estimate = 54.89), and the fixed effect (i.e. the difference between the likelihood of the reference group and the causal preceding event, Estimate = -19.32). Due to the large observed effect size, the sample size needed for 80% power to detect the effect of the intercept and fixed effect was N < 5 (Figure 2).

Stopping rule and exclusions

Considering the largest sample required to replicate all effects (N < 5), and that the effect size in the pilot could be inflated due to a small sample size, we conservatively estimate a target sample of 100 participants. To account for a 0% exclusion rate from the pilot study, we will conservatively collect data from 105 participants, prior to exclusions.

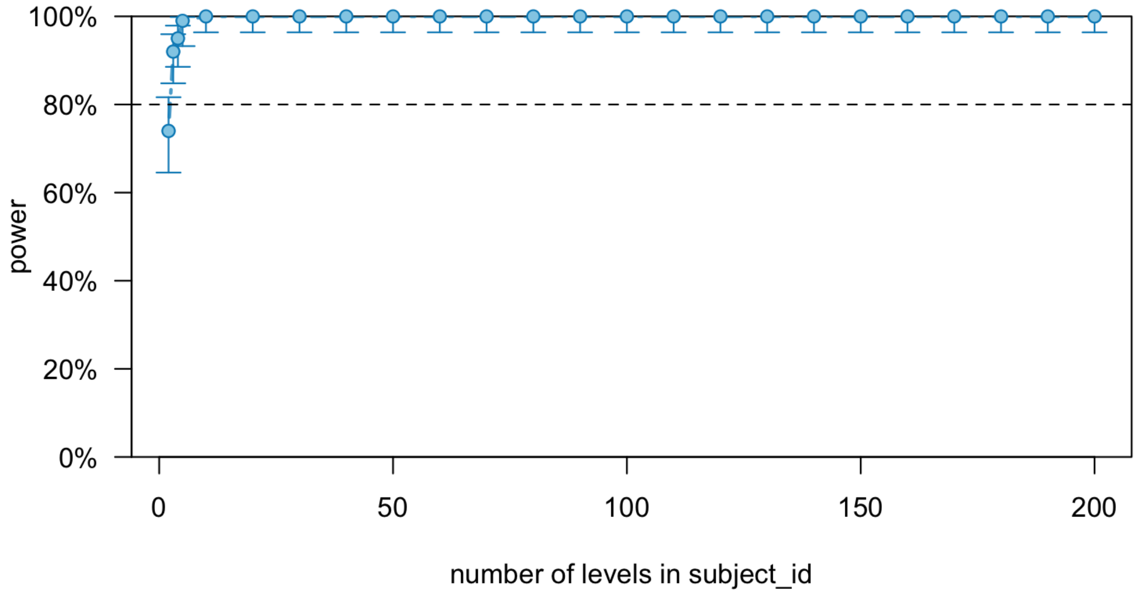

Figure 2: Power curve showing the power that would be achieved (y-axis) using different sample sizes (x-axis) to replicate the fixed effect, which is the difference between the likelihood of the target for the similar vs the causal preceding event for the `target_likelihood ~ preceding_event_type + (1 | subject_id)` model.

References

Degn, P., Fiber, Z., & Sullivan, J. (2025). Children's understanding of the causal mechanisms underlying disease prevention. *Cognitive Development*, *74*(101563), 101563. https://doi.org/10.1016/j.cogdev.2025.101563

Gelman, S. A., & Markman, E. M. (1987). Young children's inductions from natural kinds: The role of categories and appearances. *Child Development*, *58*(6), 1532. https://doi.org/10.2307/1130693

Gerstenberg, T. (2024). Counterfactual simulation in causal cognition. *Trends in Cognitive Sciences*, *28*(10), 924–936. https://doi.org/10.1016/j.tics.2024.04.012

Gerstenberg, T., & Tenenbaum, J. B. (2017). *Intuitive Theories* (M. R. Waldmann (ed.)). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199399550.013.28

Lewis, D. (1973). Causation. *The Journal of Philosophy*, *70*(17), 556. https://doi.org/10.2307/2025310

Lombrozo, T. (2010). Causal-explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions. *Cognitive Psychology*, *61*(4), 303–332. https://doi.org/10.1016/j.cogpsych.2010.05.002

Lüdecke, D., Ben-Shachar, M., Patil, I., Waggoner, P., & Makowski, D. (2021). Performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, *6*(60), 3139. https://doi.org/10.21105/joss.03139

Woodward, J. (2005). *Making things happen*. Oxford University Press.