Adults' use of latent vs direct causes for Explanation

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; Gopnik & Wellman, 1994). 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 demonstrated that both adults and children can spontaneously generate, seek and endorse explanations for everyday phenomena (Lombrozo, 2006). Explanation is a cognitive function enabled by our causal-explanatory common-sense theories of the world (Gerstenberg & Tenenbaum, 2017; Lombrozo, 2006; Wellman & Liu, 2007); and it in turn helps guide inductive reasoning, hypothesis generation and causal learning (Lombrozo, 2016; Walker et al., 2014, 2017). In the prior study from our lab (Study 1), adults were asked whether one event can make a second event happen, and reported their responses on a sliding scale. One open question is whether adults call upon these same causal judgments to explain why something happened, even when not explicitly asked to. Thus we designed a task where participants rated how satisfying different explanations are for a target event (e.g. “getting tired”): half the time, they considered an event that other participants rated as strongly causally relevant (“jumping up and down”), and half the time, they considered an event that other participants rated as very similar (“feeling scared”).

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, “feel scared”) 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, “jump up and down”). 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 evaluate explanations, they are more likely to rely on representations of direct causes (i.e. whether A can cause B), than representations of latent causes (i.e. whether A and B are both caused by a common third variable). Specifically, when people are asked to rate how satisfactory an explanation is for a target event (e.g. target = get tired; causal choice = jump up and down; similar choice = feel scared), they will rate the causal choice as more satisfactory than the similar choice.

Dependent Variable

How satisfactory is the explanation (0-100%).

Independent Variable

Whether the explanation of the target outcome involves referring to **a similar event** or **a causal event.**

Procedure

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

> Welcome! In this game you will be asked to consider different explanations for an event. You will be presented with a target event (e.g. fall down) and read an explanation for that event (e.g. stepping over something). You will rate **how satisfying** this explanation is on a sliding scale.

The next pages will present 30 trials, randomly interspersed. Trials will consist of an explanation of a target event that references either a causal event, or a similar event. The test question asks subjects to evaluate the suitability of the explanation on a slider response scale. The trials will read as follows:

> Consider the following statement: Someone {target event} because they {similar or causal event}. To what extent is this a satisfactory explanation?

Each of 15 targets will be shown twice, once with a causal event as the explanation, and once with a similar event as the explanation. Below the test question will be a continuous slider scale with 5 labels: Not at all satisfying, Slightly satisfying, Somewhat satisfying, Quite satisfying, and extremely satisfying. The 30 trials will be presented in shuffled order.

There will be two randomly interspersed attention checks, and participants will be excluded if they fail one or more attention checks.

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

Analysis Plan

To test for our prediction, namely, that people will rate the explanation that references a causal event as more satisfactory than the explanation that references a similar event, we will use a linear mixed effects model using the lmer() function from the lmerTest R package. The model predicts people's satisfaction ratings for the causal versus similar event explanations, while accounting for repeated observations within subjects. The model specification will be: **satisfaction_rating ~ explanation_type + (1 | subject_id)**. 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.

Since there are two explanation types (i.e. similar and causal), we will set the reference group to be the causal explanation type, thus the intercept will represent the satisfaction rating assigned to the causal explanation. The slope coefficient will capture the difference between the satisfactory rating of the causal event explanation and the similar event explanation. We will confirm our first prediction if the slope coefficient is significantly greater than 0.

Sample Size and Stopping Rule

Power analysis for Confirmatory Analysis

We used the R package *simr* to estimate the power curve for replicating the predicted result. Based on a pilot study (N = 100), the sample size needed to achieve 80% power to replicate the slope coefficient for the model is N<10.

Stopping rule and exclusions

Considering the sample required to replicate the effects (N < 10), 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 <1% exclusion rate from the pilot study, we will conservatively collect data from 105 participants, prior to exclusions.

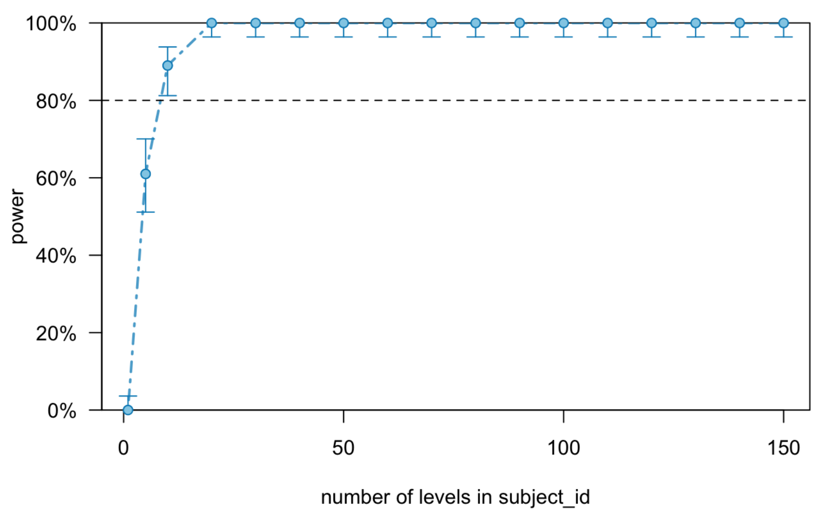

Figure 2: Power curve showing the power that would be achieved (y-axis) using different sample sizes (x-axis) to replicate the fixed for the satisfaction_rating ~ explanation_type + (1 | subject_id) model.

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

Gopnik, A., & Wellman, H. M. (1994). The theory theory. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the Mind* (pp. 257–293). Cambridge University Press. https://doi.org/10.1017/cbo9780511752902.011

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences*, *10*(10), 464–470. https://doi.org/10.1016/j.tics.2006.08.004

Lombrozo, T. (2016). Explanatory preferences shape learning and inference. *Trends in Cognitive Sciences*, *20*(10), 748–759. https://doi.org/10.1016/j.tics.2016.08.001

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

Walker, C. M., Lombrozo, T., Legare, C. H., & Gopnik, A. (2014). Explaining prompts children to privilege inductively rich properties. *Cognition*, *133*(2), 343–357. https://doi.org/10.1016/j.cognition.2014.07.008

Walker, C. M., Lombrozo, T., Williams, J. J., Rafferty, A. N., & Gopnik, A. (2017). Explaining constrains causal learning in childhood. *Child Development*, *88*(1), 229–246. https://doi.org/10.1111/cdev.12590

Wellman, H. M., & Liu, D. (2007). Causal reasoning as informed by the early development of explanations. In *Causal Learning* (pp. 261–279). Oxford University PressNew York. https://doi.org/10.1093/acprof:oso/9780195176803.003.0017