Adults hold two parallel causal frameworks for reasoning about people's minds, actions and bodies

Joseph Outa & Shari Liu
{jouta1, sliu199}@hopkins.edu
Department of Psychological and Brain Sciences,
Johns Hopkins University, USA

Abstract

Understanding other people involves making sense of their physical actions, mental states, and physiological experiences, yet little is known about the causal beliefs we hold across these domains. Across two exploratory studies, we measured these beliefs and their use in social cognition. In Study 1 (N = 50, M age = 39.44y), US adults (1) freely sorted and (2) reported causal beliefs about events of the mind, body, and actions. Representational similarity analysis (RSA) revealed two causal frameworks: one representing the 3 distinct latent categories, and another expressing causal relationships across them. Study 2 (N = 100, M age = 39.95y) demonstrated that adults flexibly apply either framework depending on the task, using the latent causes for trait inference, and causal beliefs to plan interventions on other agents. These findings suggest that intuitive theories of other people include both a sense of which capacities"go together" and their causal connections within and across domains.

Keywords: intuitive theories; intuitive psychology; intuitive biology; causality; similarity; common-sense understanding

Introduction

Every day, we are faced with the challenge of understanding other people: beings with psychological capacities to think and desire, physical capacities to walk and jump, and physiological capacities to feel hungry and tired. How are our mental representations of these capacities structured? One research tradition has approached this by identifying the latent factors that people use to organize various aspects of mental life (Weisman et al., 2017, 2021, 2025; Malle, 2019; Gray et al., 2007). By measuring people's beliefs about which capacities tend to co-occur and applying dimensionality reduction techniques to make sense of their underlying structure, these studies have revealed that people use dimensions like "agency" and "experience" to organize different sorts of agents (e.g. God, babies, robots) (Gray et al., 2007), and dimensions like "mind", "heart" and "body" to organize different sorts of experiences that agents can have (Weisman et al., 2017). Another research tradition has emphasized the existence of abstract and causally coherent systems of beliefs, termed intuitive theories of psychology and biology, which function to explain, predict, and intervene on other people's behaviors, and make sense of them as living systems (Carey, 1985, 2011; Gopnik & Wellman, 1994; Hatano & Inagaki, 1994; Wellman & Gelman, 1998). These two research traditions have proceeded in parallel without substantial cross-contact, raising questions about how they relate to each other. Here we provide an account that unifies them. We propose that we can think of these two representations (latent dimensions that organize "seeing" and "hearing" as events of the mind; and causal beliefs about whether seeing leads to hearing) as different frameworks through which our causal understanding structures our representations of other people.

The first goal of the current work is to characterize how human adults organize mental events, actions, and physiological events relative to each other. Prior work shows that people distinguish between thoughts and the objects they refer to (Wellman & Estes, 1986; Johnson & Wellman, 1982), and events of the body from events of the mind (Berent et al., 2022; Weisman et al., 2017, 2021). Both adults and children represent "the mind" as causing experiences like remembering and thinking; and "the biological body" as giving rise to events like feeling sick. We shall call this the *Latent Causes* framework: An intuitive theory of other people that structures their experiences in terms of their underlying latent causes. Yet, no experiment to date has studied mental events, mechanical actions, and bodily events in the same participants, nor tested different hypotheses about how people could in principle organize these events. This is the first contribution of the current work.

The second contribution is to contrast people's intuitions about the latent causes of our psychological, physical and biological capacities (e.g. whether the same underlying factor causes the ability to see and the ability to feel hungry) with people's causal beliefs about how specific events connect to each other (e.g. whether seeing can make someone feel hungry). One possibility is that the Latent Causes framework imposes domain-specific constraints on people's causal representations about individual events: People may expect events with a shared latent cause (such as thinking and remembering, both caused by "the mind") to be more causally connected with each other, than events with different latent causes (such as thinking and being hungry, caused by "the mind" and "the biological body" respectively). Suggestive evidence for this comes from findings that even though older children can learn new causal beliefs connecting events of the mind with events of the body from conditional dependencies, younger children struggle to attribute a psychological cause (such as thinking about show and tell) to a biological experience (such as having a tummy ache), especially when presented alongside an equally plausible biological causal explanation (such as

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

eating strawberries) (Schulz & Gopnik, 2004; Schulz et al., 2007). A second possibility is that people's causal expectations about individual events go beyond the domain boundaries implied by the Latent Causes framework. Theory of mind, for example, has been studied as a cognitive process that connects observable physical actions to their unobservable mental causes (Frith & Frith, 2005; Gopnik & Wellman, 1992; Leslie et al., 2004). However, less is known about the specific pattern of causal relationships within and across psychological, physical and biological experiences and capacities. We shall call this the *Direct Causes* framework: An intuitive theory of other people that is related to but distinct from the Latent Causes framework, and that structures their experiences in terms of causal relationships between mental events, actions, and physiological events, connecting each domain of events to themselves, and to each other. Whereas prior research has tested a few of these causal connections (such as whether feeling nervous can make someone feel sick; Schulz et al., 2007) in children, we are missing a specified larger-scale description even of adults' intuitions about the causal connections across these 3 domains.

In sum, in the current work, we conducted a larger-scale study to measure people's beliefs about the latent causes of a variety of mental events, actions, and physiological events, as well as their beliefs about how these events are causally connected. We asked 3 main research questions. First, when people are explicitly asked to categorize mental events, actions, and physiological events any way they wish, which categories do they use to organize these events? Second, when people are asked to instead report causal connections between the same events, do they use a similar or distinct representation to organize their causal beliefs? And lastly, do people intuitively and flexibly use these two frameworks in commonsense social cognition, even when they are not explicitly asked to?

Study 1

In Study 1, we measured and compared people's beliefs about the organization of mental events, actions, and bodily events - both how similar and distinct they are from each other, and how they can cause each other. We demonstrate that adults hold two distinct representational spaces: one for organize mental events, actions, and physiological events along those category boundaries, and a distinct web of causal beliefs connecting pairs of these events within and across categories.

Methods

Participants We recruited fifty US adults (Mean age = 39.44, range = [23,65]) from Cloud Research Connect to participate in an online experiment. All participants provided informed consent to participate, and were paid $7.50 for 30 minutes of their time. Study procedures were approved by the Institutional Review Board at Johns Hopkins University. The demographics of our sample were as follows: 16 female, 34 male; 6 Hispanic 43 Non-hispanic; 1 American Indian, 11 Asian, 6 Black/African American, 30 White, 1 White/Hispanic. None of the participants met our exclusion criteria, which was failing two or more of our four attention checks (see below).

Design and Procedure In Study 1, adult participants saw 15 items spanning various mechanical, psychological and biological capacities (see Figure 1, matrix axes for the full list). These items were drawn directly from prior work (Berent et al., 2022; Weisman et al., 2017, 2021). They were chosen to vary roughly similarly in phrasing and thematic spread. For events of the body (e.g. hunger and illness), we included both drawn out and phasic bodily changes. For actions, we included both object-directed and non-object directed actions. For mental events, we included both perceptual and cognitive events. Using these 15 items, We aimed to measure and interpret people's intuitions about (1) the similarities and differences and (2) the causal relations between all events. Therefore, after providing informed consent, all participants engaged in a Sorting Task and a Causal Task, presented in a random order. Participants' responses in each of the two tasks was highly similar regardless of the order that they completed them (for participants who completed the Sorting Task, first vs second, $r = 0.86$; for the Causal task, $r = 0.95$). Therefore, we did not investigate the role of task order in the subsequent analyses.

In the Sorting Task, participants were told that they would see 15 cards describing experiences a person might have, and that their job was to organize the items into groups based on how similar they were. This was followed by a short video demonstration, showing how to drag the items into an example configuration using placeholder items. During the main task, participants saw a circular sorting canvas with items placed in random positions outside the canvas. Participants were asked to take as long as they needed to drag and drop items in order to group them inside the canvas. Participants submitted their responses by clicking a 'continue' button, which only appeared once all items were placed. Participants were not told about about the three *a priori* categories.

In the Causal Task, participants were told they would see pairs of events a person might experience, and their job was to report whether experiencing one event could make someone experience the other. Participants gave responses for 210 pairs of the 15 items: this included every item paired with every other item in both orders (item A causing item B and vice versa), but excluded items paired with themselves. For each pair, participants reported on a continuous slider scale the extent to which experiencing one event (e.g. jumping up and down) could make someone experience the other (e.g. feel tired), with responses ranging from "definitely not" to "definitely yes". Trials were presented in groups of 15, either organized by the potential cause (e.g. Can jumping up and down make someone [...]?), or by the potential effect (Can [...] make someone jump up and down?). For all participants, the particular order of items within a group, and the order of groups, were presented in a random order.

Four attention check questions were interspersed randomly through the experiment, and participants were excluded if

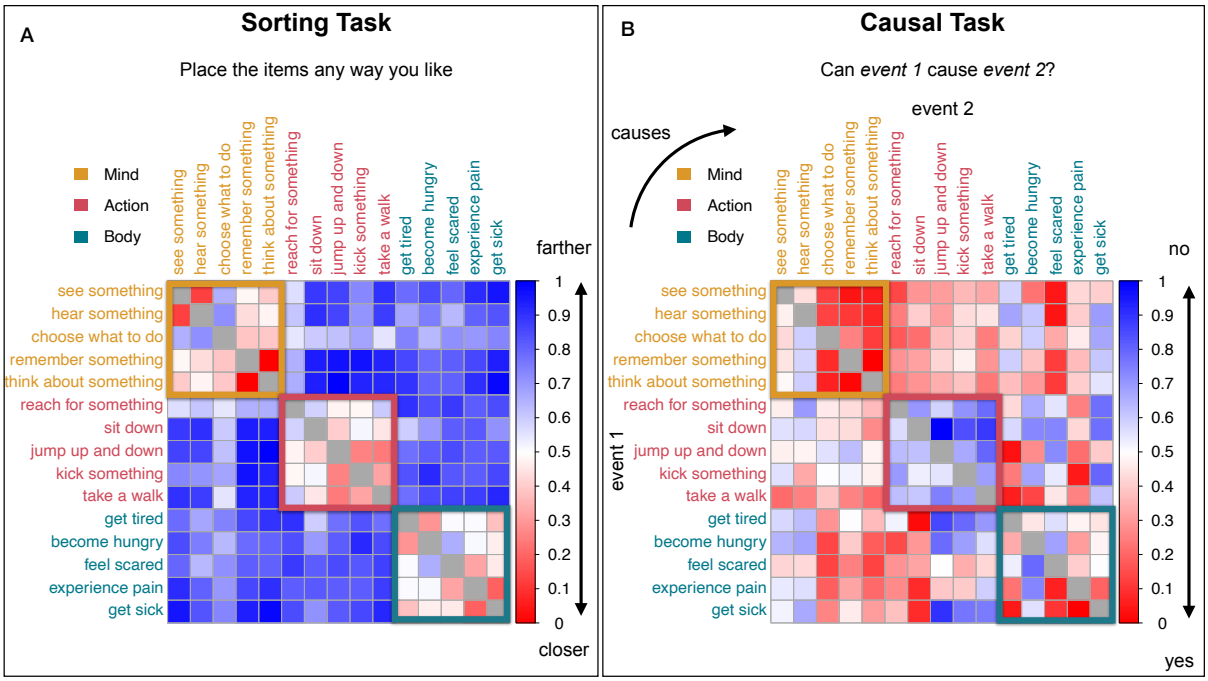

Figure 1: (A-B) Results for Study 1. (A) Group-averaged representational dissimilarity matrix (RDM) for responses in the Sorting Task, achieved by computing the normalized Euclidean distance between the sorted items for each participant, and averaging across participants. (B) Group-averaged RDM for the Causal Task, achieved by computing the normalised causal ratings between all pairs of items for each participant, and averaging across participants. In (A) and (B), closer items (Sorting Task) and more causally connected items (Causal Task) are plotted in hotter colors, and farther items (Sorting Task) and less causally connected items (Causal Task) are plotted in cooler colors. Note that the RDM from the Sorting Task is by design symmetrical, while the Causal Task RDM is asymmetrical. Items are color-coded by a priori domain categories.

they failed at least two of them (none met this criteria). Upon completing both tasks, participants saw a debriefing statement explaining the purpose of the study, and were asked to optionally report their demographic information.

Tasks for both Studies 1 and 2 were implemented using jsPsych version 7 (de Leeuw, 2015) and hosted as a webpage on Github Pages (live demos: Study 1 and Study 2). Participant responses were collected on a private OSF repository using Datapipe (de Leeuw, 2024). All anonymized data, and analysis scripts required to reproduce the results are available at `https://github.com/liulaboratory-experiments/mind-body-action-causation`.

Representational similarity analysis

We used representational similarity analysis (RSA; (Kriegeskorte et al., 2008) to visualize, compare, and interpret participants' responses for both tasks. In brief (see SI §1.1-1.2 for details), for each participant we built a representational dissimilarity matrix (RDM) that expresses the spatial distance between items in the Sorting Task, and the "causal distance" between items in the Causal task; for both of these tasks, values closer to 0 indicate smaller distance, and values closer to 1 indicate a larger distance. We built an RDM per participant per task by computing the normalized distance between pairs of items: for the Sorting Task, the Euclidean distance between item positions; for the Causal Task, the numerical slider response on a given trial, which ranged from 0 (definitely not) to 100 (definitely yes).

Given these RDMs, one per participant per task, we investigated several research questions. First, we measured and compared the average agreement across individuals for each task by computing a group-level noise ceiling. We did this by computing Kendall's τ between each participant's RDM and the average of all the other participants' RDMs, iterating through participants, and averaging across folds. Second, we compared the similarity between the RDMs from the two tasks by measuring Pearson's r between each pair of RDMs within participants. Third, to make sense of and compare the representational spaces elicited by the two tasks, we compared people's empirical RDMs to 4 theoretical RDMs (see SI§1.3):

(1) **Physical and Psychological**: Inspired by the hypothesis that people draw a clear boundary between events of the physical world and events of the mental world (Berent et al., 2022), this model organizes events according to two latent categories: the mind (5 items) and the body (10 items).

(2) **Mind, Action and Physiology**: Organizes the items according to the 3 *a priori* latent causes: the mind, physical

actions of the body, and physiological/biological bodily processes.

(3) **Fine-grained Mind, Action, and Physiology**: This model makes one further distinction within each latent category proposed in (2) (e.g. perceptual vs cognitive events; object directed vs non-object directed actions; drawn out vs phasic bodily changes).

(4) **Phrase embeddings**: This model uses a Universal Sentence Encoder to represent the 15 events as high dimensional vectors, derived from large scale corpora using machine learning techniques (Cer et al., 2018). The model reflects the hypothesis that phrases are encoded as distributed semantic feature spaces (McClelland & Rogers, 2003; Laurence & Margolis, 1999), with their relationships computed as cosine distances between vectors.

For each of these four models, we computed Kendall's τ between each participant's RDM and the theoretical RDM expressed by the model. For RDMs from the Sorting Task, we took the values from one half of the off-diagonal values, since the responses in this task were necessarily symmetrical; for RDMs from the Causal Task, we took the values from both halves of the off-diagonal matrix, since people gave ratings about both whether one item could cause a second, and vice versa. Then, we compared the distribution of Kendall's τ across the four different models for the Sorting and Causal Tasks.

Results and Discussion

The results of Study 1 are summarized in Figure 1. First, the noise ceiling was higher in the Causal Task (noise ceiling τ = 0.42) than in the Sorting Task (τ = 0.26), and this finding was not driven by having twice the number of observations in the Causal task (upper diagonal noise ceiling, τ: 0.44; lower diagonal τ: 0.41). This indicates that people agreed with each other more when they reported their beliefs about which events cause other events, than when they sorted the events into categories.

Second, a linear mixed-effects model ($\tau \sim$ `RDM_type` + (1 | `subject_id`)) showed that the Mind, Action and Physiology model performed best, outperforming the Fine-grained Mind, Action and Physiology model ($\Delta\tau = -0.060$), the Physical and Psychological model ($\Delta\tau = -0.110$), and the Phrase Embeddings model ($\Delta\tau = -0.125$), all $ps < .001$; see Supplementary Figure 1). This model's performance (τ = 0.252) approached the noise ceiling for this task, suggesting that it explained nearly all the explainable variance in people's responses. Thus, participants made use of the 3 *a priori* latent cause categories when asked explicitly to sort the 15 events in any way they wished.

By contrast, none of the models explained people's responses in the Causal Task (Kendall's τ = -0.036 to -0.08; see Supplementary Figure 1). Instead, we observed asymmetrical causal judgments that freely crossed the domain boundaries revealed in the Sorting Task. Adults reported that mental events are more likely to cause actions (mean rating = 0.32, 95% confidence interval [0.31, 0.34]) than the other way around (0.41 [0.40, 0.43]). Adults also were more likely to reject the claim that actions could directly cause each other (0.61 [0.59, 0.63]), compared to whether mental events could cause each other (0.27 [0.26, 0.29]), and bodily events could cause other bodily events (0.38 [0.36, 0.40]). People also held distinctive causal beliefs for specific events within and across domains. For example, people agreed that perception (seeing and hearing) could cause cognition (deciding, thinking, remembering) (0.15 [0.13, 0.17]) more so than vice versa (0.48 [0.44, 0.51]). Thus, participants made use of a distinct set of representations when they were explicitly asked to report whether the 15 events can or cannot cause each other, vs when they were asked to sort the events any way they wished.

Study 2

Study 1 suggests that when people are asked explicitly to sort versus consider the direct causal connections between mental events, actions, and physiological events, they make use of two distinct representational spaces. In Study 2, we studied whether adults make use of these same distinct spaces, when they are not explicitly asked to do so, during two classic social cognition tasks: inferring which traits co-occur in agents and intervening on agents to bring about intended outcomes.

Methods

Participants A new group of 100 US adults (Mean age = 39.95, range = [18, 73]) were recruited from Cloud Research Connect to participate in an online experiment, and compensated with $1 for 4 minutes of their time. The demographics of our sample were as follows: 45 female, 49 male 5 nonbinary; 4 Hispanic 94 Non-hispanic; 11 Asian, 12 Black/African American, 1 Multiracial, 1 Native Hawaiian/Pacific Islander, 75 White. After excluding one participant who failed an attention check, our final sample was 99.

Design and Procedure Our goal in Study 2 was to test whether adults would flexibly use the two sorts of representations we measured in Study 1, even when the task does not explicitly require them to. Therefore, we constructed a stimuli set consisting 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. closeby according to the Sorting Task RDM) to the target ("feel scared", causal distance = 0.48, similarity distance = 0.41), and an event that was most causally relevant (i.e. closeby according to the Causal Task RDM) to the target ("jump up and down", causal distance = 0.11, similarity distance = 0.56). For the rest of the paper, we will refer to these choice items as the "similar" and the "causal" options. 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 can be found in the panel labels of Figure 2B, and their respective distances are listed in Supplementary Table 1. We embedded the same 15 item triads into two tasks, described below.

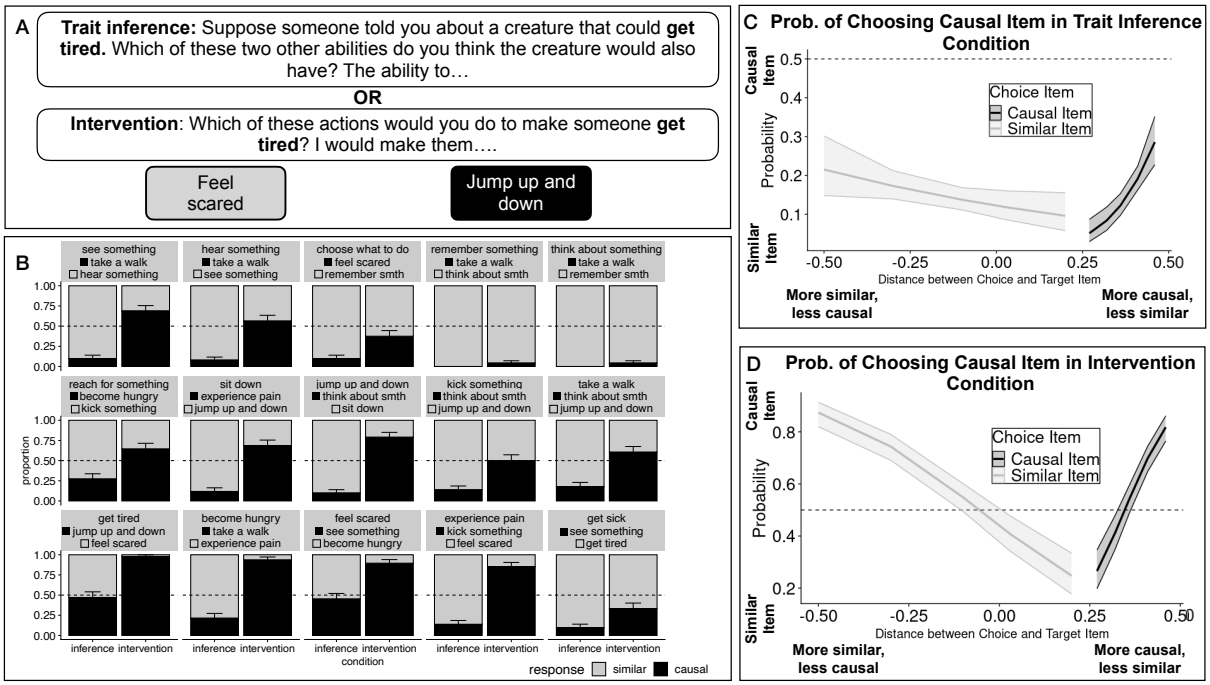

Figure 2: Methods and results for Study 2. (A) In the Trait Inference Task, participants selected which of two abilities they expect a hypothetical creature to have, given that they have a target ability. In the Intervention Task, participants selected which of two actions they would do to another agent to produce a target outcome. The choice sets were identical across both tasks. (B) Across all 15 items, participants were more likely to select the causal item in the Intervention Task than in the Inference Task. Error bars indicate confidence intervals. (C) In the Trait Inference task, people were more likely to choose the similar item if it was more similar and less causally relevant to the target, and if the causal item was less similar and more causally relevant to the target. (D) In the Intervention Task, people were more likely to choose the causal item if it was more causally relevant and less similar to the target and if the similar item was less causally relevant and more similar to the target. For C-D, estimates indicate the change in log odds for each unit increase in the predictor, and ribbons indicate confidence intervals based on standard errors of the predicted probabilities.

Participants were randomly assigned to complete either a Trait Inference Task or an Intervention Task (Figure 2A). The Trait Inference Task asked participants to consider a hypothetical creature that has one target ability (e.g. the ability to get tired), and were asked to infer which other ability the creature would also have (the causal option: the ability to jump up and down, vs the similar option: the ability to feel scared).

In the Intervention Task, participants saw the same items and choices, but embedded in a different task context. Instead of reasoning about the abilities of a novel creature, participants were asked to plan an intervention. They considered an outcome (e.g. "make someone feel tired"), and selected which of two choices they thought would most likely bring about that outcome (the causal option: "make them jump up and down" vs the similar option: "make them feel scared").

In both tasks, participants saw one trial at a time. The order of trials and left-right positions of the choice items were randomized. Participants were excluded if they failed a single attention check question (1 participant met this criterion). All other aspects of the design and procedure (IRB approval, collecting informed consent, debriefing and demographics) were identical to Study 1.

Hypothesis and analysis

We hypothesized, in line with prior work, that patterns of latent causes between the 15 events will guide participants' expectations about which abilities tend to co-occur in the Trait Inference task Gray et al. (2007); Weisman et al. (2017, 2021), and that direct causal beliefs about specific pairs of events will guide participants' causal interventions Ho et al. (2022); Wu et al. (2024).

The dependent measure was participants' choice between the causal and the similar item, on each trial. We modeled these choices in a mixed effects logistic regression, implemented using the lme4 package Bates et al. (2014) in R Team (2021). First, we tested the simplest prediction of our hypothesis: that people would be more likely to choose the causal option (over the similar option) in the Intervention than the Trait Inference task. We did this by fitting a logistic regression model on choices as a function of task condition, including a random intercept for participants: `glmer(response ~ condition + (1|subject_id), family = binomial(link = 'logit'))`.

Second, we asked whether we could predict variability in people's choice across trials from distances in the two representational spaces between the target and each option. For each condition, we modeled participants' choices as a function of these variables (see SI §2.2).

Results and Discussion

People chose the causally relevant item above chance in the Intervention Task (60%, 95% CI [55.1, 64.7]), and below chance in the Trait Inference Task (15.3%, 95% CI [12.4,18.7]); these proportions differed across tasks (β = 2.11, SE = 0.16, $z = 13.18$, $p<0.001$). This trend appeared for all 15 items that we tested (Figure 2B), although the size of the effect varied substantially across trials. For example, for the target "see something" (causal option: "take a walk" vs similar option: "hear something"), people's probability of choosing the causal item was 9% in the Trait Inference Task and 69% in the Intervention Task. By contrast, for the target item "think about something" (causal option: "take a walk" vs similar option: "remember something"), this difference was much smaller: 0% in the Trait Inference Task and 4% in the Intervention Task.

What accounts for this variability across items? We found that the distances reported by participants in Study 1 the new set of participants' choices between the causal and similar options (Figure 2C-D), even though participants in this study were not asked explicitly to use similarity or causal relevance for either task. In the Trait Inference task, people were more likely to choose the similar option (that is, less likely to choose the causal option), if (i) it was more similar and less causally relevant to the target ($\beta = 10.55$, SE = 2.06, $z = 5.12$, $p<.001$), and (ii) if the causal option (the distractor, for this condition) was less similar and more causally relevant to the target ($\beta = 1.35$, SE = 0.66, $z = 2.04$, $p<0.001$). In the Intervention Task, people were more likely to choose the causal option if (i) it was more causally relevant and less similar to the target ($\beta = 13.27$, SE = 1.52, $z = 8.71$, $p<0.001$), and (ii) if the similar option was less causally relevant and more similar to the target ($\beta = 4.35$, SE = 0.53, $z = 8.16$, $p<0.001$). Calling back to the example from the last paragraph, one potential reason that we observed a small effect for the target "think about something" is that "remember something" (the similar option) is nearby in both causal and similarity space, leading people to prefer it in both tasks.

General Discussion

In this paper, we studied the structure of adults' representations of other people's mental states, actions, and physiological states. We found that adults possess two distinct representational spaces, one organized in terms of latent causes (for example, a mind that causes both remembering and seeing; a body that causes both hunger and pain), and a second organized in terms of causal relations connecting specific events to each other (for example, seeing something can cause someone to feel hungry). In Study 1, we replicated prior results that people think about certain events as "going together" in a way that separates events of the mind, actions, and events of the body.

Our novel contribution was the demonstration that adults also draw asymmetric causal connections between events of the mind, biological processes, and physical actions, that were not well explained by prior accounts about latent dimensions of mental life (Weisman et al., 2017, 2021, 2025), or by an intuitive separation between the mind versus the body (Berent et al., 2022). Instead, we found evidence for a causal theory of other people that crossed domain boundaries, both at a broader scale (that is, mental states tend to cause actions more than vice versa), and at a finer scale (that is, jumping up and down tends to make others tired, but not sick). The interpretation that these are two mutually compatible, causally coherent framework theories of other people predicts that people should flexibly prioritize them, depending on the task. We found evidence for this prediction in Study 2: Despite never being asked to do so, adults recruited the latent-cause representations captured by our Sorting Task to make trait inferences, and the direct-cause representations captured by our Causal Task for planning interventions on other people.

Implications for theories of domains

There are two senses in which "domain" is used in cognitive science. One sense, reflected in the Latent Causes framework, treats domains as clusters of phenomena that share an underlying cause (such as events of the mind, or natural kinds in intuitive biology). A second sense, reflected in the Direct Cause framework, treats domains as intuitive theories (such as theory of mind) that specify direct causal relationships between events (such as mental states to actions), regardless of domain boundaries. Our results suggest that people simultaneously hold both frameworks in mind, including beliefs expressing interactions between domain-specific systems of knowledge.

Implications for cognitive development

How do adults acquire causal beliefs that span domains? One possibility is that they are built on top of an early-emerging ontological commitment that the physical and physiological body, and the immaterial mind, are distinct. This account predicts that children begin to work out cross-domains connections by accumulating evidence about themselves and other people. Yet, prior literature complicates this simple story: by some accounts, children's intuitive understanding within domains undergoes conceptual change. For example, children learn to distinguish being alive with being animate (Carey, 1985) and to robustly represent beliefs as propositional attitudes (Wellman & Woolley, 1990; Wellman et al., 2001). Therefore, it is plausible children face an even more difficult learning problem than previously proposed: they are tasked with constructing both intuitive theories of physics, psychology, and biology; and a meta-theory that connects these domains. Alternatively, it is possible that young children may already have some basic elements of this meta-theory, which accommodates more complex representations as each domain develops. We intend to explore these possibilities in future work.

Acknowledgments

Many thanks to members of the LiuLab for feedback. Shari Liu was supported by National Institutes of Health NRSA Postdoc Fellowship (F32HD103363). Joseph Outa was supported by Robert S. Waldrop & Dorothy L. Waldrop Graduate Fellowship.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv [stat.CO]*.

Berent, I., Theodore, R. M., & Valencia, E. (2022). Autism attenuates the perception of the mind-body divide. *Proc. Natl. Acad. Sci. U. S. A.*, *119*(49), e2211628119.

Carey, S. (1985). *Conceptual change in childhood*. London, England: MIT Press.

Carey, S. (2011). *The origin of concepts*. New York, NY: Oxford University Press.

Cer, D., Yang, Y., Kong, S.-Y., Hua, N., Limtiaco, N., John, R. S., ... Kurzweil, R. (2018). Universal sentence encoder. *arXiv [cs.CL]*.

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

de Leeuw, J. R. (2024). DataPipe: Born-open data collection for online experiments. *Behav. Res. Methods*, *56*(3), 2499–2506.

Frith, C., & Frith, U. (2005). Theory of mind. *Curr. Biol.*, *15*(17), R644–6.

Gopnik, A., & Wellman, H. M. (1992). Why the child's theory of mind really is a theory. *Mind Lang.*, *7*(1-2), 145–171.

Gopnik, A., & Wellman, H. M. (1994). The theory theory. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind* (pp. 257–293). Cambridge: Cambridge University Press.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619.

Hatano, G., & Inagaki, K. (1994). Young children's naive theory of biology. *Cognition*, *50*, 171–188.

Ho, M. K., Saxe, R., & Cushman, F. (2022). Planning with theory of mind. *Trends Cogn. Sci.*, *26*(11), 959–971.

Johnson, C. N., & Wellman, H. M. (1982). Children's developing conceptions of the mind and brain. *Child Dev.*, *53*(1), 222–234.

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Front. Syst. Neurosci.*, *2*, 4.

Laurence, S., & Margolis, E. (1999). Concepts and cognitive science. *Concepts: core readings.*.

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in "theory of mind". *Trends Cogn. Sci.*, *8*(12), 528–533.

Malle, B. F. (2019). How many dimensions of mind perception really are there? In *Proceedings of the annual meeting of the cognitive science society* (Vol. 41).

McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nat. Rev. Neurosci.*, *4*(4), 310–322.

Schulz, L. E., Bonawitz, E. B., & Griffiths, T. L. (2007). Can being scared cause tummy aches? naive theories, ambiguous evidence, and preschoolers' causal inferences. *Dev. Psychol.*, *43*(5), 1124–1139.

Schulz, L. E., & Gopnik, A. (2004). Causal learning across domains. *Dev. Psychol.*, *40*(2), 162–176.

Team, R. C. (2021). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Weisman, K., Dweck, C. S., & Markman, E. M. (2017). Rethinking people's conceptions of mental life. *Proc. Natl. Acad. Sci. U. S. A.*, *114*(43), 11374–11379.

Weisman, K., King, L. S., & Humphreys, K. (2025). Beliefs about the development of mental life. *Open Mind*, *9*, 515–539.

Weisman, K., Legare, C. H., Smith, R. E., Dzokoto, V. A., Aulino, F., Ng, E., ... Luhrmann, T. M. (2021). Similarities and differences in concepts of mental life among adults and children in five cultures. *Nat. Hum. Behav.*, *5*(10), 1358–1368.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Dev.*, *72*(3), 655–684.

Wellman, H. M., & Estes, D. (1986). Early understanding of mental entities: A reexamination of childhood realism. *Child development*, 910–923.

Wellman, H. M., & Gelman, S. A. (1998). Knowledge acquisition in foundational domains. *Handbook of child psychology: Volume 2: Cognition, perception, and language.*, *2*(1998), 523–573.

Wellman, H. M., & Woolley, J. D. (1990). From simple desires to ordinary beliefs: the early development of everyday psychology. *Cognition*, *35*(3), 245–275.

Wu, S., Schulz, L., & Saxe, R. (2024). How to change a mind: Adults and children use the causal structure of theory of mind to intervene on others' behaviors. In *Proceedings of the annual meeting of the cognitive science society, vol 46.*