Measuring adults’ representations of other people’s minds, physical actions, and biological bodies

Start Date

As of June 26, 2025, data from an exploratory pilot study (N = 50) using the same methods and analyses have been collected and analyzed.

Background

Prior work has shown that our knowledge of other people is organized into distinct domains, including representations of people as physical objects (solid bodies that exist continuously through space and time; Saxe et al. 2006; Spelke et al. 1995), mental beings (with distinct kinds of mental states such as desires and beliefs; Yu and Wellman 2024), and living systems (subject to states like hunger and pain, and processes like birth, growth, and death; Carey 1985; Hatano and Inagaki 1994).

Our causal understanding plays a central role in structuring representations of the domains of mind, physical action and biological body. One kind of causal organization is supported by studies examining the latent dimensions underlying our conceptions of other agents and their mental lives. For example, human adults organize the emotional and mental experiences of other people in terms of three latent factors corresponding to mind, body and heart (Weisman, Dweck, & Markman, 2017; Weisman et al., 2021). Even children rely on latent unobserved factors to group entities, classifying non-physical mental events (such as a thought about a cookie) as separate from tangible objects (such as an actual physical cookie), and expecting members of a natural kind, and not members who are perceptually similar, to share unseen properties (Gelman & Markman, 1987; Johnson & Wellman, 1982; Wellman & Estes, 1986). One implication of these results is that children and adults represent underlying factors, such as “the mind”, that causes experiences like remembering, seeing and thinking; and “the biological body”, which gives rise to events like feeling sick and feeling hungry. We shall call this set of beliefs **“Latent Factors”: Representations that structure the experiences of other people in terms of their underlying latent dimensions (e.g. the mind, the mechanical body, and the biological body)**. Yet, no experiment to date has tested the hypothesis that people think about mental events, bodily actions, and physiological events as similar to themselves, and distinct from each other. Therefore the first contribution of this experiment is to measure those intuitions and test them against alternative ways of organising these events (e.g. events about the body vs events about the mind).

The second contribution of this experiment is to contrast people’s intuitions about how latent factors organize events with intuitions about how causal beliefs connect specific events to each other (e.g. whether seeing can make someone feel hungry). One possibility is that the latent factors people ascribe to mental and bodily events impose domain-specific constraints on people’s causal representations about individual events: People may expect events with a shared latent factor (such as thinking and remembering, both events caused by “the mind”) to be more causally connected with each other, than events with different latent factors (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 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 eating strawberries) (Schulz et al., 2007; Schulz & Gopnik, 2004). A second possibility is that people’s causal expectations about individual events go beyond the domain boundaries implied by representations of latent factors. Theory of mind, for example, has been studied as a cognitive process that connects observable physical actions to their unobservable mental causes (Frith & Frith, 2003; Gopnik & Wellman, 1992; Leslie et al., 2004). However, less is known about the specific pattern of causal relationships within and across different domains. Do people expect actions to also cause actions, just as much as mental states cause actions in theory of mind? Do people expect all cognitive events to not cause biological events, or only some? The nature of these asymmetric relationships, if they exist, to our knowledge, has not been studied. We shall call this set of beliefs that allows people to causally connect individual events, “**Direct Causes”: Representations of other people that are related to but dissociable from representations of latent factors, that structures their experiences in terms of direct causal relationships between mental events, actions, and physiological events, connecting each domain of events to themselves, and to each other.**

Thus, prior work leaves several open questions. First, because prior research has tested just 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 larger-scale picture even of adults’ intuition about the causal connections across domains. Are representations of direct causes in adults redundant with representations of latent factors, such that causal expectations about individual events fall along domain boundaries? Or do people’s beliefs about the causal relationships between events from these three categories extend beyond domain boundaries? What kinds of causal symmetries and asymmetries do adults represent in this set of beliefs?

Therefore in the current work, we will conduct a larger-scale study to measure these beliefs in people’s minds. Do we represent other people by conceiving of the (i) mind, action and body as distinct domains with shared latent factors, and (ii) as causally connected domains?

Methods

Design

Participants will be shown 15 items spanning three a priori domain categories: 5 mental events (see something, hear something, choose what to do, remember something, think about something), 5 actions (reach for something, sit down, jump up and down, kick something, take a walk), and 5 bodily activities (get tired, become hungry, feel scared, experience pain, get sick). 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 valence, phrasing and thematic spread: among bodily items, rather than using the same transitive verb “feel”, we used variations like “get”, “experience” and “become”; among action items, we included both object-directed and non-object directed actions; and mental events spanned both perceptual and cognitive events.

We aim to measure and interpret people’s intuitions about (1) the similarities and differences in meaning and (2) the causal relations between all 15 events. Therefore, all participants will engage in a Sorting Task and a Causal Task, presented in a random order. In the Sorting Task participants will be presented with all fifteen items, and a circular sorting area. They will be asked to arrange all the items in the circle in whatever configuration they want, and to take as long as they want to complete the task.

The Causal Task will ask participants to provide causal ratings for all 210 pairings of the 15 items: this includes every item paired with every other item, both in forward and reverse causal order (such as item A causing item B and vice versa), but excludes items paired with themselves (such as item A causing item A). For each pair, participants will report on a continuous slider scale the extent to which experiencing one event (for example, jumping up and down) could make someone experience the other (for example, feel tired), with responses ranging from “definitely not” to “definitely yes”. Trials will be presented in groups of 15, either organized by the potential cause (for example, can jumping up and down make someone [...]?), or by the potential effect (for example, can [...] make someone jump up and down?) (randomized across participants), with the particular order of items within a group, and the order of groups, presented in a random order. Four attention check questions will be interspersed randomly through the experiment, and participants will be excluded if they failed at least two attention checks.

Tasks for both Studies 1 and 2 will be implemented using jsPsych version 7 (de Leeuw 2015) and hosted as a webpage on Github Pages. Participant responses will be collected on a private OSF repository using Datapipe (de Leeuw 2024). All anonymized data, and analysis scripts required to reproduce the results will be made available.

Procedure

After providing informed consent, participants will first see a general instruction page explaining that they are going to perform two tasks. Then participants will be presented with the Sorting Task and the Causal Task in a randomized order.

In the Sorting Task, participants will be told that they will see 15 cards describing experiences a person might have, and that their job is to organize the items into groups based on how similar in meaning they are. This will be followed by a short video demonstration, showing how to drag the items into an example configuration using placeholder items. During the main task, participants will see a circular sorting canvas with items placed in random positions outside the canvas. Participants will be asked to take as long as they need to drag and drop items in order to group phrases with a similar meaning inside the canvas. Participants will submit their responses by clicking a ‘continue’ button, which will only appear once all items are placed.

In the Causal Task, participants will be told they would see pairs of events a person might experience, and their job is to report whether experiencing one event could make someone experience the other. This will be followed by 15 pages, each with a question at the top asking to what extent experiencing one event can cause the second event. Below it will be a list of prompts asking about the causal relatedness between 15 pairs of items. Upon completing both tasks, participants will see a debriefing statement explaining the purpose of the study, and will be asked to optionally report their demographic information.

Hypotheses

Hypothesis Set 1: The structure of representations of latent factors

Our first set of alternative hypotheses concerns the structure of representations of latent factors, which reflects how adults group the capacities of other agents (tested in the Sorting Task), and is reflected in the number of grouping categories adults represent, and which specific items are included in each category. We consider four hypotheses, formalized as RDMs instantiating distinct representational geometries for organizing the 15 capacities.

The 4 hypotheses for how the 15 items might be sorted, expressed as RDM models are as follows:

1. *Physical and psychological model*: Organizes the items into 2 categories: physical items (10 items, events of the body) and ethereal items (5 items, events of the mind). This is in line with the predictions of Intuitive Dualism, which argues that we intuitively draw a clear boundary between events of the physical world and events of the mental world (Berent et al., 2022). The latent factors proposed by this model are “the body” (responsible for actions and physiological processes) and “the mind” (responsible for mental events).
2. *Mind, action, and physiology model*: Organizes the items into 3 categories with 5 items per category: mental items, bodily items, and actions. This is similar to (1), except that we further distinguish between physical overt behaviors (actions), and physiological processes internal to the body. The latent factors proposed by this model are “the mechanical body” which causes the actions, “the biological body” which causes the physiological processes and “the mind” which causes the mental events.
3. *Fine-grained mind, action, and physiology model*: We built this model to make one further distinction within each domain proposed in (2). This RDM organizes the items into 6 categories: perceptual events (2 items: seeing and hearing), cognitive events (3 items: remembering, thinking, and choosing), object-directed actions (2 items: reaching and kicking), non-object-directed actions (3 items: walking, sitting and jumping), phasic changes that are more drawn out in time (3 items: sick, tired, hungry), versus ones which begin and end more abruptly (2 items: scared, pain).
4. *USE phrase embeddings model*: This model is based on phrase embeddings from the Universal Sentence Encoder model, which captures the similarity in semantic meaning between all 15 items in terms of similar shared features, and graded differences between the items (Cer et al., 2018). This is based on the hypothesis that phrases which appear in similar contexts have similar meanings, and can thus be captured by vector based representations derived from large scale corpora using machine learning techniques (McClelland & Rogers, 2003). To generate model

predictions for this model, we generated phrase embeddings for each of the 15 items using the USE model. Each embedding was a 512-dimensional vector capturing the semantic content of the phrase. To get a similarity measure, we computed the cosine similarity between each pair of phrase embeddings by taking the inner product of the two vectors and dividing by the product of their magnitudes. This produced a value ranging from -1 (completely dissimilar) to 1 (completely similar). Since our analysis required a dissimilarity measure, we transformed the cosine similarity by rescaling it using the formula (1 - cosine similarity) / 2. This made the values range from 0 (very similar) to 1 (very dissimilar).

Hypothesis Set 2: The relationship between representations of latent factors vs and direct causes

The second set of alternative hypotheses concerns the relationship between latent factors and direct causes. Are people's beliefs about the categories that structure these events, and the causal connections between these events, distinct or related?

1. **The two representational spaces are related or redundant**: One possibility is that direct cause representations are redundant with latent factors representations. Adults may expect events that share a latent factor to be more directly causally related to each other, than events that do not share a latent factor. This could arise from domain-specific constraints on causal expectations: adults may expect events that do not share a latent factor to have properties that make them less likely to engage in causal interactions with each other.
2. **The two representations are distinct**: Another possibility is that US adults have two dissociable, causal-explanatory frameworks of other people, one that organizes events of the mind, actions, and events of the body into their latent factors, and the other that specifies causal connections (or lack thereof) between events, both within and across domains.

This first experiment only begins to address these questions. For example, we will be able to measure and describe the structure of people's judgments, and test whether the two representational spaces from the two tasks (sorting and causal judgments) are similar vs distinct in structure. But, this study does not address what sorts of representations articulate people's judgments. That will be the focus of subsequent studies.

Analysis Plan

We will use representational similarity analysis (RSA; Kriegeskorte et al. 2008) to visualize, compare, and interpret participants' responses for both tasks. In brief, for each participant we will compute the normalized Euclidean distances between pairs of items for the Sorting Task, and the normalized "causal distance" (i.e. the slider response) between pairs of items in the Causal Task, and arrange them into representational dissimilarity matrices (RDMs) (see below). We will then measure the average agreement across individuals in each task by computing a group-level noise ceiling. This noise ceiling will indicate the maximum possible agreement an explanatory model could achieve with the data, given the natural variability in participant's responses (see below). Next, we will test both sets of our proposed hypotheses.

Visualizing Sorting data: Generating Sorting Task RDM

The output of the Sorting Task will be a position vector with x and y coordinates for the final position of each of the 15 items on the canvas. We will calculate the euclidean distance between each pair of items, using the formula $\sqrt{(x_{itemA} - x_{itemB})^2 + (y_{itemA} - y_{ItemB})^2}$. We will then min-max normalize the resulting distance by subtracting the minimum euclidean distance and dividing this by the range, using this formula: $\frac{(euclidean\ distance - min(euclidean\ distance)}{max(euclidean\ distance) - min(euclidean\ distance)}$. We shall call this final value the *freesort distance*.

For each participant these values will be organized into a 15 by 15 matrix—a representational dissimilarity matrix (RDM)—with each value in the matrix corresponding to the freesort distance between pairs of items. We shall call this the Freesort RDM. We shall then create a group-averaged RDM by averaging the RDMs across subjects.

Visualizing Causal data: Generating Causal Task RDM

The output of the Causal Task will be an integer ranging from 0 (definitely not) to 100 (definitely yes) (corresponding on the final location of the slider knob). We shall invert these response values to map ratings of 100 to 0 and ratings of 0 to 100 using the formula $1 - response / 100$. This operation will convert the measure into a dissimilarity measure (with values close to 0 meaning causally close, and values close to 1 meaning causally distant). Finally, we shall min-max normalize the values (as in the Sorting Task), and call the final value the *causal distance*.

We shall also organize each participant's responses into a 15 by 15 matrix to form the subject-specific causal RDM, which will be averaged across participants to form the mean causal RDM.

Computing noise ceiling

We will compute the group-level upper and lower bound noise ceiling for each task.

For the Sorting Task, we will compute Kendall's τ between each participant's upper diagonal RDM and the average of all the other participants' upper diagonal RDMs, iterating through participants, and averaging across folds. The upper-bound of the noise ceiling will be computed by including the held-out subject's upper diagonal RDM in the average of all other participant's RDMs (i.e. the grand average), whereas the lower-bound noise ceiling will be computed by omitting the held-out subject in the average of all other participant's RDMs.

For the Causal Task, we will use the same procedure to compute an upper and lower bound noise ceiling for the upper and lower diagonal of people's responses, since the measurements are asymmetric. We will average the estimates across the two halves of the data to compute the final noise ceilings for the task.

Confirmatory Analysis 1: The structure of the Latent Factors

To analyze the structure of people's sorting judgments, we will compare people's empirical sorting task RDMs to 4 alternative hypotheses for organizing the 15 items, each expressed as an RDM with distinctive representational geometries (see Hypotheses for full list). The model RDMs are as follows:

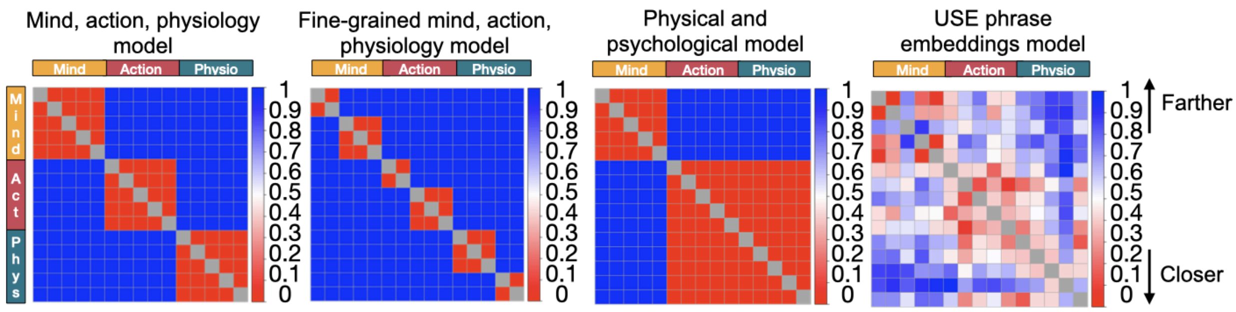

Figure 1: The four model RDMs. The axes are labelled with a priori domains of mind (yellow), action (red) and physiology/ biological body (blue-green). There are 5 items within each domain and 15 total items. The *Mind, action, physiology model* organizes the items into 3 categories; the *Fine-grained mind, action, and physiology model* organizes items into 6 categories; the *Physical and psychological model* organizes items into 2 categories and the *USE phrase embeddings model* organizes items based on the cosine similarities between their embedding vectors.

A priori, it is unclear which of these models will best explain adults' data from the Sorting Task. All of them have the potential to capture latent factors that support adults' grouping of the items. Some prior work suggests, in line with the *Physical and psychological model*, that one of the earliest categorical distinctions may involve dichotomies between mental and physical events (Opfer and Gelman 2010; Wellman and Estes 1986). However, adults and even young children can entertain more granular distinctions within the physical and mental domain (Weisman et al. 2021; Berent et al. 2022). At the same time, people may not spontaneously generate the finer grained distinctions proposed by the *Fine-grained mind, action, and physiology model*: the granularity of categorization may be shaped by relevance of the items to our causal-explanatory theories of the world (Murphy and Medin 1985). This principle is best captured by the *Mind, action, and physiology model*, which distinguishes among mental processes, physiological states, and physical actions—three broad yet functionally meaningful categories that support our common sense explanation and prediction of other people's behaviors. Finally, while the *Phrase embeddings model* may capture some semantic relationships among the items, these meanings may or may not capture people's intuitions about what defines the semantic meaning of these models (Ullman 2023; Mahowald et al. 2024). By contrast, none of these models contain representations of direct causal relations within or across domains, measured during the Causal Task. In future work, we will build towards this RDM in a data-driven way. For now, we will remain agnostic about the specific causal structure this framework represents.

Based on these theoretical considerations, and empirical evidence from a prior exploratory study (see **Pilot Data**), we predict that:

(1) **Prediction 1:** Participants' responses in the Sorting Task will be best explained by the *Mind, action and physiology model*.

(2) **Prediction 2:** Whichever model best accounts for people's responses in the Sorting Task, that RDM will do a poorer job in accounting for people's responses in the Causal Task.

To test **Predictions 1-2,** we will compute Kendall's τ between each theoretical RDM and participant's empirical RDM for the Sorting Task and Causal Task. For RDMs from the Sorting Task, we will compute Kendall's τ between the normalized values from the upper off-diagonal values and all 4 theoretical models, since the responses in this task will be necessarily symmetrical; for RDMs from the Causal Task, we will compute Kendall's τ between the normalized values from both halves of the off-diagonal matrix and all 4 theoretical models, since people will give ratings about both whether one item could cause a second, and vice versa.

To test **Prediction 1**, we will compare the distribution of Kendall's τ comparing the 4 theoretical RDMs to the empirical data from the Sorting Task by regressing τ on RDM type. We will fit a linear mixed effects model of the form: **tau ~ RDM_type + (1|subj)**. RDM_type is a dummy-coded categorical predictor with 4 levels, and we will set the reference level to the *Mind, action and physiology* RDM*.* The coefficients from the model will therefore express how much better that the *Mind, action and physiology* RDM describes the data than each of the other 3 RDMs, and that all models may account for some people's data better than others.

Confirmatory Analysis 2: Comparing Latent Factors with Direct Cause Representations

To test **Prediction 2**, we will take the theoretical RDM that numerically or statistically best accounts for people's responses in the Sorting Task, and compare the distribution of Kendall's τ from this task with the distribution of Kendall's τ comparing that RDM to responses from the Causal Task. We will fit a linear mixed effects model of the form: **tau ~ task + (1|subj)**, where 'task' is a dummy coded categorical predictor (reference level = Sorting, vs Causal). The coefficient on that predictor will therefore represent how much better the "best" theoretical RDM is at accounting for the data from the Sorting task, relative to the Causal Task.

For both analyses, 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.

Existing data

We ran an exploratory initial study (N = 50) of the Sorting Task and Causal Task in a within-subjects design. We found the following:

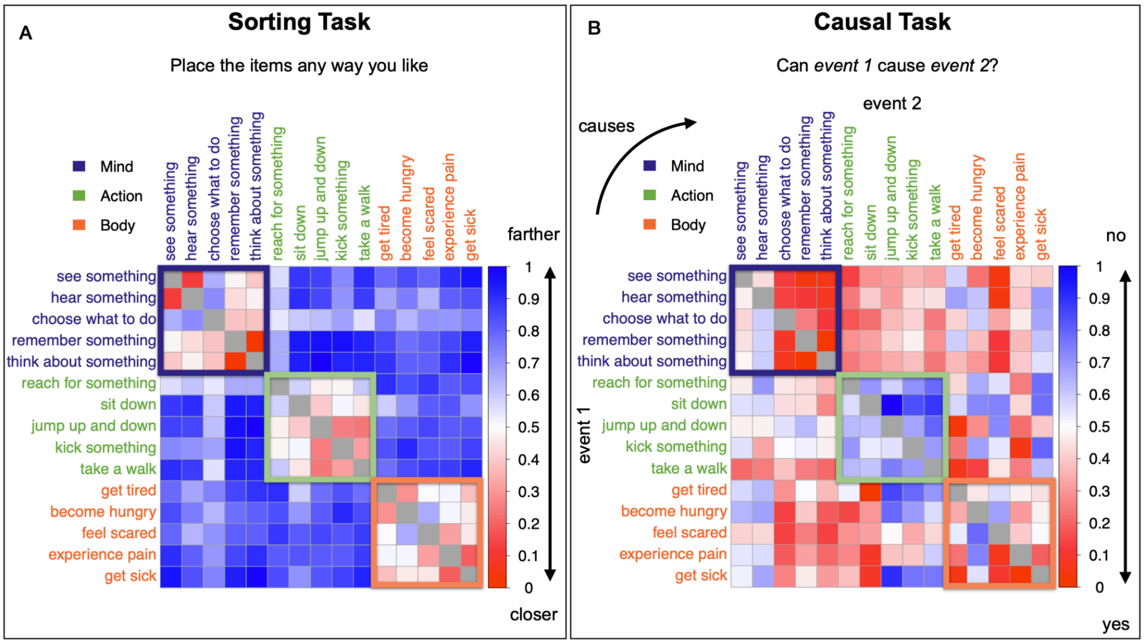

Figure 2: (A-B) Results for pilot study 1. (A) Group-averaged representational dissimilarity matrix (RDM) for responses in the Sorting Task, achieved by computing the normalized euclidean distance between the sorted items for each participant, and averaging across participants. (B) Group-averaged RDM 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 symmetrical, while the Causal Task RDM is asymmetrical.

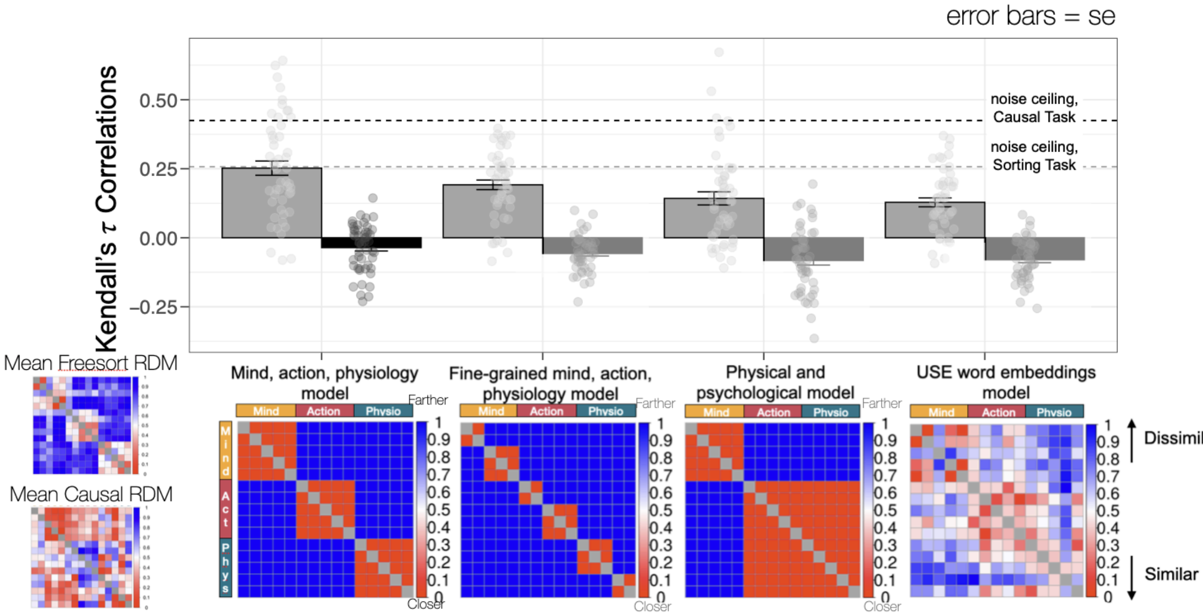

Figure 3: Kendall's τ correlations between the four models and average RDMs in the Sorting and Causal Task of pilot Study 1. Most of the variance in the Sorting Task RDM was captured by the Mind, action and physiology model (this correlation reached the Task's noise ceiling), followed by the Fine-grained mind, action and physiology model, the Physical and psychological model, and finally the

Universal Sentence Encoder (USE) embedding model. The Causal Task was negatively correlated with all four models, and was least captured by the Mind, action and physiology model.

Sample size and Stopping Rule

Prior studies examining the latent dimensions that underlie our conceptions of other agents and their mental lives have used sample sizes ranging from 200 to 700 (Weisman et al., 2017, 2021) and up to 2000 (Gray et al., 2007). However, these studies employ different methods from ours, so we based our sample size estimate on results from our pilot study.

Power analysis for Confirmatory Analysis 1 and 2

Based on a pilot study (N = 50), we used the R package *simr* to estimate the sample size needed to achieve 80% power to replicate all four model coefficients for the tau ~ RDM_type + (1|subj) model: the intercept (i.e. the reference model; Mind, action and physiology, τ = 0.252), and three coefficients, each expressing the difference in fit between the reference model and the fine-grained model ($\Delta\tau$ = –0.060), physical and psychological model ($\Delta\tau$= –0.110) and phrase-embeddings ($\Delta\tau$ = –0.125) model respectively. The sample size needed for 80% power to detect the effect of the intercept (the mind, action, physiology model) was <20, and the sample sizes needed to distinguish the mind, action and physiology model from the rest of the models was 45 for the fine-grained mind, action, physiology model (see figure below for this model coefficient's power curve), 18 for the physical and psychological model and 10 for the embeddings model. For the second confirmatory analysis the sample size required to detect how much better the mind, action and physiology model explains the Sorting Task RDM than the Causal Task RDM (i.e. to replicate the fixed effect of task for the model tau ~ task + (1|subj)) was also <20.

Considering the largest sample from all simulations (N = 45), 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 ~5% exclusion rate, we will collect data from 105 participants, prior to exclusions.

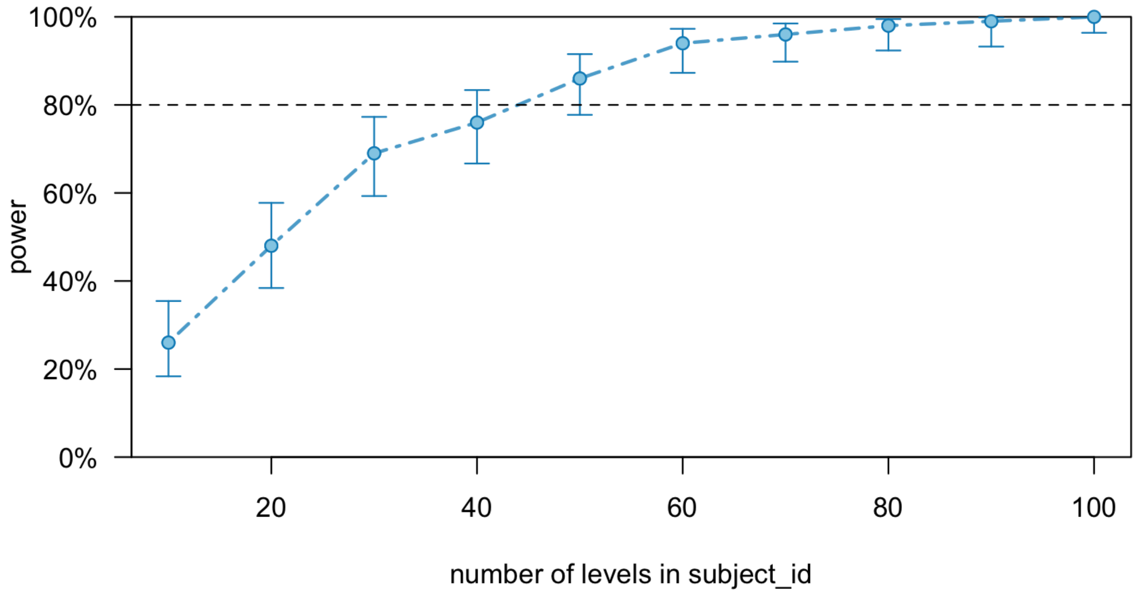

Figure 4: Power curve showing the power that would be achieved (y-axis) using different sample sizes (x-axis) to replicate the "fine-grained mind, action and physiology" model coefficient of the tau ~ RDM_type + (1|subj) model. This coefficient models the difference in fit to participant data between the mind, action and physiology model and the fine-grained mind, action and physiology model.

References

Berent, I., Theodore, R. M., & Valencia, E. (2022). Autism attenuates the perception of the mind-body divide. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(49), e2211628119. https://doi.org/10.1073/pnas.2211628119

Carey, S. (1985). *Conceptual Change in Childhood*. MIT Press.

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *358*(1431), 459–473. https://doi.org/10.1098/rstb.2002.1218

Gopnik, A., & Wellman, H. M. (1992). Why the child's theory of mind really is a theory. *Mind & Language*, *7*(1-2), 145–171. https://doi.org/10.1111/j.1468-0017.1992.tb00202.x

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

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science (New York, N.Y.)*, *315*(5812), 619. https://doi.org/10.1126/science.1134475

Hatano, G., & Inagaki, K. (1994). Young children's naive theory of biology. *Cognition*, *50*, 171–188. https://doi.org/10.1016/0010-0277(94)90027-2

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in "theory of mind." *Trends in Cognitive Sciences*, *8*(12), 528–533. https://doi.org/10.1016/j.tics.2004.10.001

Schulz, L. E., Bonawitz, E. B., & Griffiths, T. L. (2007). Can being scared cause tummy aches? Naive theories, ambiguous evidence, and preschoolers' causal inferences. *Developmental Psychology*, *43*(5), 1124–1139. https://doi.org/10.1037/0012-1649.43.5.1124

Schulz, L. E., & Gopnik, A. (2004). Causal learning across domains. *Developmental Psychology*, *40*(2), 162–176. https://doi.org/10.1037/0012-1649.40.2.162

Weisman, K., Dweck, C. S., & Markman, E. M. (2017). Rethinking people's conceptions of mental life. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(43), 11374–11379. https://doi.org/10.1073/pnas.1704347114

Weisman, K., Legare, C. H., Smith, R. E., Dzokoto, V. A., Aulino, F., Ng, E., Dulin, J. C., Ross-Zehnder, N., Brahinsky, J. D., & Luhrmann, T. M. (2021). Similarities and differences in concepts of mental life among adults and children in five cultures. *Nature Human Behaviour*, *5*(10), 1358–1368. https://doi.org/10.1038/s41562-021-01184-8

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. https://psycnet.apa.org/fulltext/2005-01927-010.pdf