Supplemental Information for the paper “Adults hold two parallel causal frameworks for reasoning about people's minds, actions and bodies”

Study 1

Methods: Representational Similarity Analysis

1.1 Generating Sorting Task RDM

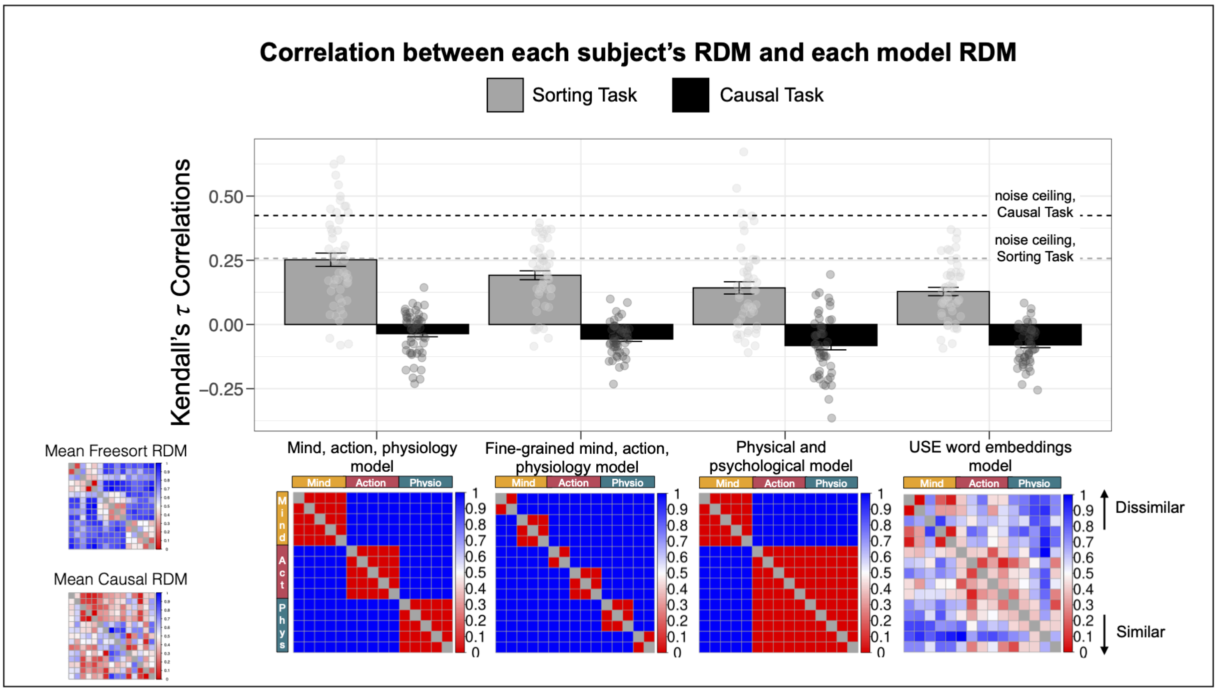
The output of the Sorting Task was a position vector with x and y coordinates for the final position of each of the 15 items on the canvas. We calculated the euclidean distance between each pair of items, using the formula $\sqrt{(x_{itemA} - x_{itemB})^2 + (y_{itemA} - y_{ItemB})^2}$. We then min-max normalized the resulting distance by subtracting the minimum euclidean distance and dividing this by the range, as shown in this formula: $\frac{(euclidean\ distance - min(euclidean\ distance)}{max(euclidean\ distance) - min(euclidean\ distance)}$. We called this final value the *freesort distance*. For each participant these values were 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 called this the *Freesort RDM.* A group-averaged RDM was created by averaging the RDMs across subjects; this is depicted in Figure 1B of the main manuscript.

1.2 Generating Causal Task RDM

The output of the Causal Task was an integer ranging from 0 (definitely not) to 100 (definitely yes) (corresponding on the final location of the slider knob). We inverted these response values, which mapped ratings of 100 to 0 and ratings of 0 to 100 using the formula $1 - response / 100$. This operation converted the measure into a dissimilarity measure (with values close to 0 meaning causally close, and values close to 1 meaning causally distant). Finally, we min-max normalized the values (as was done in the Sorting Task), and called the final value the *causal distance*. These were also organized into a 15 by 15 matrix to form the subject-specific *causal RDM*, which was averaged across participants to form the mean causal RDM, depicted in Figure 1B of the paper.

1.3 Comparing Observed RDMs to Model RDMs

Given these RDMs (one per participant per task) we compared them to different explanatory models, each expressed as an RDM with distinctive representational geometries for organizing the 15 items (see Supplementary Figure 1, x-axis). The 4 models were as follows:



Supplementary Figure 1: Kendall's τ correlations between the four models and average RDMs in the Sorting and Causal Task of 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. The Mean Freesort and Causal RDMs from subjects' responses are provided for reference in the bottom left.

(1) *Physical and psychological mode*l: 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).
(2) *Mind, action, 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.
(3) *Fine-grained mind, action, and physiology*: 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) *Word embeddings*: This model is based on 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: $cos\theta = \frac{A \cdot B}{|A||B|}$. 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 $\frac{1 - cosine\ similarity}{2}$. This made the values range from 0 (very similar) to 1 (very dissimilar).

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. We plotted the participant-level correspondences with the model RDMs alongside their respective noise ceilings (Supplementary Figure 1). Then, we compared the distribution of Kendall's τ across participants to 0 (no relationship between the model and the data), to measure how well a given model accounts for responses across participants in each task.

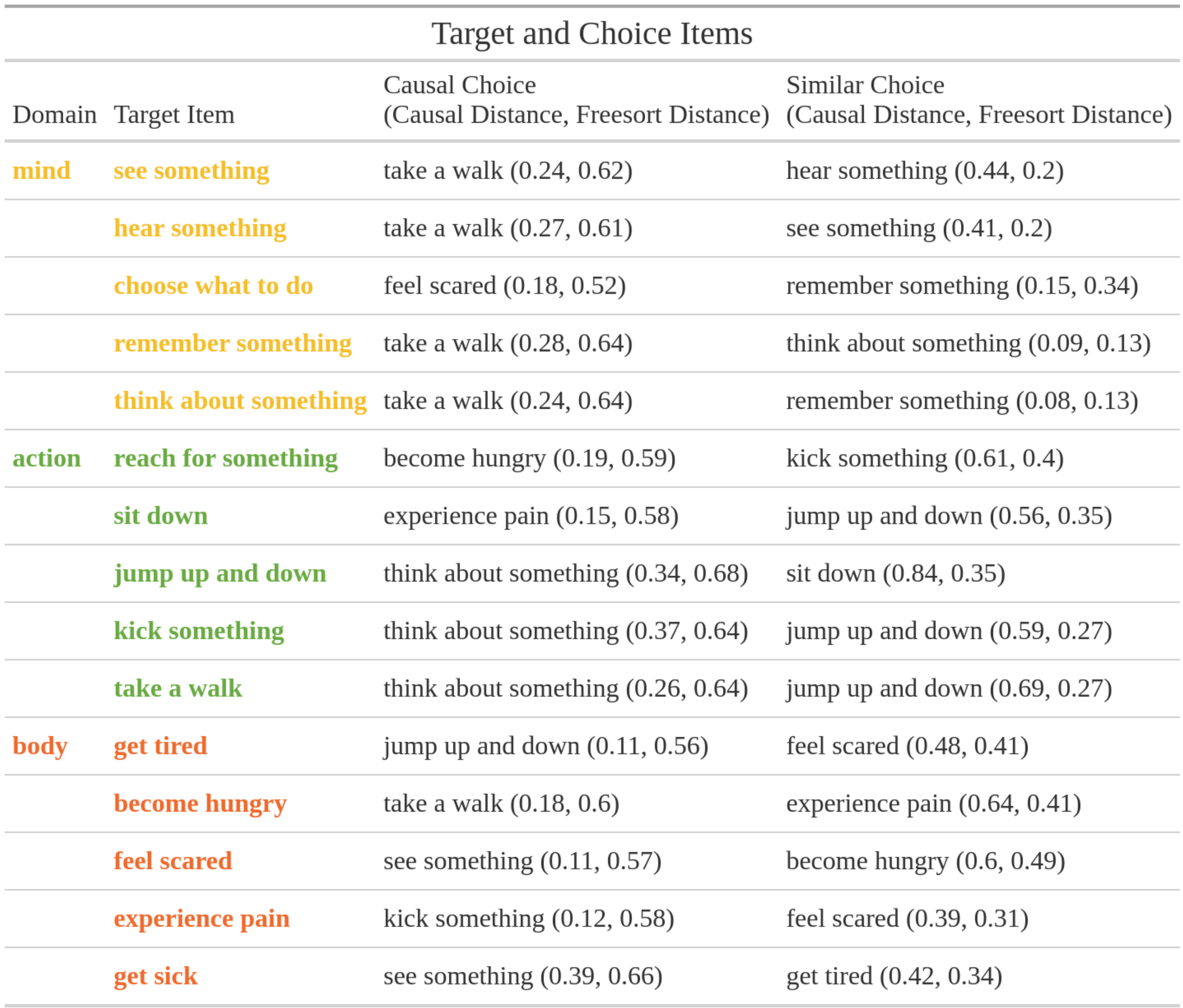
Study 2

2.1 Item-selection Procedure

Supplementary Table 1 (below) shows the 15 target items, and the choices they were paired with in both the Inference and Intervention Tasks. For each target item, we designed a choice set consisting of a "similar option" and a "causal option".

The similar option was the item closest in similarity space to the target item, and farthest in causal space to the target item. This was captured by a variable called $sorting\_minus\_causal\_distance$, and the similar option was the item with the lowest value on this variable (typically, a large negative value; for example, for the target item "jump up and down", the candidate "sit down" had a similarity distance of 0.35 and a causal distance of 0.84, leading to a $sorting\_minus\_causal\_distance$ of -0.49, versus the candidate "think about something" which had a $sorting\_minus\_causal\_distance$ of 0.34. Thus, "sit down" became the similar option item). We imposed a constraint that the most similar item be in the same domain as the target item; this was already true for 13 of 15 of the target items, that is, all items except for "choose what to do" (i.e. an item in the "mind" domain) whose lowest $sorting\_minus\_causal\_distance$ item was "jump up and down" (i.e. an item in the "action" domain); and "get sick" (i.e. an item in the "physiological" domain) whose most similar item was "sit down" (i.e. an item in the "action" domain).

The causal option was the item closest in causal space to the target item and farthest in similarity space to the target item; thus, items high on $sorting\_minus\_causal\_distance$. We imposed a constraint that the most causal item be in a different domain from the target item, though this was already true for 14 of the 15 items, that is, all except for “choose what to do”, whose causal items were a tie between “hear something” and “feel scared”.



Target and Choice Items

		Causal Choice	Similar Choice
Domain	Target Item	(Causal Distance, Freesort Distance)	(Causal Distance, Freesort Distance)
mind	see something	take a walk (0.24, 0.62)	hear something (0.44, 0.2)
	hear something	take a walk (0.27, 0.61)	see something (0.41, 0.2)
	choose what to do	feel scared (0.18, 0.52)	remember something (0.15, 0.34)
	remember something	take a walk (0.28, 0.64)	think about something (0.09, 0.13)
	think about something	take a walk (0.24, 0.64)	remember something (0.08, 0.13)
action	reach for something	become hungry (0.19, 0.59)	kick something (0.61, 0.4)
	sit down	experience pain (0.15, 0.58)	jump up and down (0.56, 0.35)
	jump up and down	think about something (0.34, 0.68)	sit down (0.84, 0.35)
	kick something	think about something (0.37, 0.64)	jump up and down (0.59, 0.27)
	take a walk	think about something (0.26, 0.64)	jump up and down (0.69, 0.27)
body	get tired	jump up and down (0.11, 0.56)	feel scared (0.48, 0.41)
	become hungry	take a walk (0.18, 0.6)	experience pain (0.64, 0.41)
	feel scared	see something (0.11, 0.57)	become hungry (0.6, 0.49)
	experience pain	kick something (0.12, 0.58)	feel scared (0.39, 0.31)
	get sick	see something (0.39, 0.66)	get tired (0.42, 0.34)

Supplementary Table 1: Items used in Study 2. 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. These distances were computed from group-averaged responses from Study 1.

Thus, both the similar item and the causal item were maximally distinct in terms of both similarity and causal relevance.

2.2 Predicting variability in choices as function of the causal the sorting distance

In Study 2, we found that people chose the causally relevant item more in the Intervention Task than in the Trait Inference task. However, the size of this effect varied substantially across trials, thus we explored the sources of this variability.

One potential source of variability is the causal relevance and similarity between the target, the causal choice, and the similar choice. While our item selection procedure maximized the *difference* between these two items, both options still vary in causal and similarity distance to the target items. Thus, in a secondary set of analyses, we asked whether the distance of the two options in causal space and similarity space modulated the effect of our task manipulation.

For each condition, we ran a model to predict people's choices across trials from the causal relevance and the similarity of the items in the choice set, relative to the target. The predictor in this model was the difference between the sorting and causal distances of the choice items, each computed relative to the target. This predictor captured the extent to which each choice item was causally close yet dissimilar, and causally distant yet similar.

In the Trait Inference condition, the model formula was as follows:

$$glm(responses) \sim causal\_item\_sorting\_minus\_causal\_distance \; + \\ similar\_item\_sorting\_minus\_causal\_distance, \; family \; = \; binomial(link \; = \; "logit)$$

This model originally had a random intercept for subject id, but this term was removed after the model failed to converge. The results of this model are visualized in Figure 2C of the main manuscript.

In the Intervention condition, the model formula was as follows:

$$glm(responses) \sim causal\_item\_sorting\_minus\_causal\_distance \; + \\ similar\_item\_sorting\_minus\_causal\_distance \; + \; (1|subject\_id), \; family \; = \\ binomial(link \; = \; "logit)$$

The results of this model are visualized in Figure 2D of the main manuscript.

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

Cer, D., Yang, Y., Kong, S.-Y., Hua, N., Limtiaco, N., John, R. S., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., Sung, Y.-H., Strope, B., & Kurzweil, R. (2018). Universal Sentence Encoder. In *arXiv [cs.CL]*. arXiv. http://arxiv.org/abs/1803.11175

McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews. Neuroscience*, *4*(4), 310–322. https://doi.org/10.1038/nrn1076