RESEARCH ARTICLE

Animacy semantic network supports causal inferences about illness

Miriam Hauptman*, Marina Bedny

Department of Psychological & Brain Sciences, Johns Hopkins University, Baltimore, United States

eLife Assessment

This study investigates the neural basis of causal inference of illness, suggesting that it relies on semantic networks specific to living things in the absence of a generalized representation of causal inference across domains. The main hypothesis is **compelling**, and is supported by **solid** methods and data analysis. Overall, the findings make a **valuable** contribution to understanding the role of domain-specific semantic networks, particularly the precuneus, in implicit causal inference about illness.

Abstract Inferring the causes of illness is a culturally universal example of causal thinking. We tested the hypothesis that making causal inferences about biological processes (e.g. illness) depends on the animacy semantic network. Participants (n=20) undergoing fMRI read two-sentence vignettes that elicited implicit causal inferences across sentences, either about the emergence of illness or about the mechanical breakdown of inanimate objects, in addition to noncausal control vignettes. All vignettes were about people and were linguistically matched. The same participants performed localizer tasks: language, logical reasoning, and mentalizing. Inferring illness causes, relative to all control conditions, selectively engaged a portion of the precuneus (PC) previously implicated in the semantic representation of animates (e.g. people, animals). Neural responses to causal inferences about illness were adjacent to but distinct from responses to mental state inferences, suggesting a neural mind/body distinction. We failed to find evidence for domain-general responses to causal inference. Causal inference is supported by content-specific semantic networks that encode causal knowledge.

***For correspondence:** mhauptm1@jhu.edu

Competing interest: The authors declare that no competing interests exist.

Funding: See page 15

Preprint posted
13 July 2024
Sent for Review
20 August 2024
Reviewed preprint posted
19 December 2024
Reviewed preprint revised
02 May 2025
Version of Record published
12 November 2025

Reviewing Editor: Roberto Bottini, University of Trento, Italy

© Copyright Hauptman and Bedny. This article is distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use and redistribution provided that the original author and source are credited.

Introduction

A distinguishing feature of human cognition is our ability to reason about complex cause-effect relationships, particularly when causes are not directly perceptible (***Tooby and DeVore, 1987***; ***Lagnado et al., 2007***; ***Rottman et al., 2011***; ***Muentener and Schulz, 2014***; ***Sloman and Lagnado, 2015***; ***Goddu and Gopnik, 2024***). When reading something like, *Hugh sat by sneezing passengers on the subway. Now he has a case of COVID*, we naturally infer a causal relationship between crowded spaces and the invisible transmission of infectious disease. Here we investigate the neurocognitive mechanisms that support such automatic inferences by studying causal inferences about illness.

Adults have rich, culturally specific causal knowledge about the invisible forces that bring about illness, from pathogen transmission to divine retribution (***Notaro et al., 2001***; ***Raman and Winer, 2004***; ***Lynch and Medin, 2006***; ***Legare and Gelman, 2008***; ***Legare et al., 2012***; ***Legare and Shtulman, 2017***). In many societies, designated 'healers' become experts in diagnosing and treating disease (***Foster, 1976***; ***Ackerknecht, 1982***; ***Norman et al., 2009***; ***Lightner et al., 2021***). Nonexpert adults routinely infer the causes of illness in themselves and others (e.g. *how did my friend get COVID?*).

Even young children think about illness in systematic ways, reflecting their burgeoning commonsense understanding of the biological world (**Wellman and Gelman, 1992**; **Keil, 1992**; **Inagaki and Hatano, 2006**). Young children attribute illness to contaminated food, contact with a sick person, and parental inheritance (**Springer and Ruckel, 1992**; **Kalish, 1996**; **Kalish, 1997**; **Keil et al., 1999**; **Notaro et al., 2001**; **Raman and Winer, 2004**; **Raman and Gelman, 2005**; **Legare and Gelman, 2008**; **Legare et al., 2009**; **DeJesus et al., 2021**).

Illness affects living things (e.g. people and animals) rather than inanimate objects (e.g. rocks, machines, houses). Thinking about living things (animates) as opposed to nonliving things (inanimate objects/places) recruits partially distinct neural systems (e.g. **Warrington and Shallice, 1984**; **Hillis and Caramazza, 1991**; **Caramazza and Shelton, 1998**; **Farah and Rabinowitz, 2003**). The precuneus (PC) is part of the 'animacy semantic network' and responds preferentially to living things (i.e. people and animals), whether presented as images or words (**Devlin et al., 2002**; **Fairhall and Caramazza, 2013a**; **Fairhall et al., 2014**; **Peer et al., 2015**; **Wang et al., 2016**; **Silson et al., 2019**; **Rabini et al., 2021**; **Deen and Freiwald, 2022**; **Aglinskas and Fairhall, 2023**; **Hauptman et al., 2025**). By contrast, parts of the visual system (e.g. fusiform face area [FFA]) that respond preferentially to animates do so primarily for images (**Kanwisher et al., 1997**; **Grill-Spector et al., 2004**; **Noppeney et al., 2006**; **Mahon et al., 2009**; **Konkle and Caramazza, 2013**; **Connolly et al., 2016**; see **Bi et al., 2016**, for a review). We hypothesized that the PC represents causal knowledge relevant to animates and tested the prediction that it would be activated during causal inferences about illness, which rely on such knowledge (preregistration: https://osf.io/6pnqg).

We also compared neural responses to causal inferences about the body (i.e. illness) and inferences about the mind (i.e. mental states). Both types of inferences are about animate entities, and some developmental work suggests that children use the same set of causal principles to think about bodies and minds (**Carey, 1985**; **Carey, 1988**). Other evidence suggests that by early childhood, young children have distinct causal knowledge about the body and the mind (**Springer and Keil, 1991**; **Callanan and Oakes, 1992**; **Wellman and Gelman, 1992**; **Inagaki and Hatano, 1993**; **Inagaki and Hatano, 2004**; **Keil, 1994**; **Hickling and Wellman, 2001**; **Medin et al., 2010**). For instance, preschoolers are more likely to view illness as a consequence of biological causes, such as contagion, rather than psychological causes, such as malicious intent (**Springer and Ruckel, 1992**; **Raman and Winer, 2004**; see also **Legare and Gelman, 2008**). The neural relationship between inferences about bodies and minds has not been fully described. The 'mentalizing network', including the PC, is engaged when people reason about agents' beliefs (**Saxe and Kanwisher, 2003**; **Saxe et al., 2006**; **Saxe and Powell, 2006**; **Dodell-Feder et al., 2011**; **Dufour et al., 2013**). We localized this network in individual participants and measured its neuroanatomical relationship to the network activated by illness inferences.

An alternative hypothesis is that domain-general neural mechanisms, separate from semantic networks, support causal inferences across domains. Children and adults make causal inferences across a wide range of domains and use similar cognitive principles (e.g. 'screening off') when doing so (e.g. **Saxe and Carey, 2006**; **Tenenbaum et al., 2007**; **Carey, 2011**; **Cheng and Novick, 1992**; **Waldmann and Holyoak, 1992**; **Pearl, 2000**; **Gopnik et al., 2001**; **Steyvers et al., 2003**; **Gopnik et al., 2004**; **Schulz and Gopnik, 2004**; **Rehder and Burnett, 2005**; **Lagnado et al., 2007**; **Rottman and Hastie, 2014**; **Davis and Rehder, 2020**). Prior neuroscience work has hypothesized that the frontotemporal language network may support a broad range of causal inferences during comprehension (**Kuperberg et al., 2006**; **Mason and Just, 2011**; **Prat et al., 2011**; see also **Spelke, 2003**; **Spelke, 2022**; **Pinker, 2003**). Alternatively, causal inference could depend on frontoparietal mechanisms that also support other types of reasoning, such as logical deduction (**Goldvarg and Johnson-Laird, 2001**; **Barbey and Patterson, 2011**; **Khemlani et al., 2014**; **Operskalski and Barbey, 2017**). Finally, it has been suggested that causal inferences are supported by a dedicated 'causal engine' in prefrontal cortex that supports all and only causal inferences across domains (**Pramod et al., 2023**). We tested these alternative hypotheses in the specific case of implicit causal inferences that unfold naturally during language comprehension (**Black and Bern, 1981**; **Keenan et al., 1984**; **Trabasso and Sperry, 1985**; **Myers et al., 1987**; **Duffy et al., 1990**).

Most prior studies investigating causal inference used explicit causality judgment tasks (**Ferstl and von Cramon, 2001**; **Satpute et al., 2005**; **Fugelsang and Dunbar, 2005**; **Kuperberg et al., 2006**; **Fenker et al., 2010**; **Kranjec et al., 2012**; **Pramod et al., 2023**). For example, **Kuperberg et al.,**

2006 asked participants to rate the causal relatedness of three-sentence stories and observed higher responses to causally related stories in left frontotemporal cortex. Studies of implicit causal inference report frontotemporal and frontoparietal responses (*Chow et al., 2008*; *Mason and Just, 2011*; *Prat et al., 2011*). Across these prior studies, no consistent neural signature of causal inference has emerged. Importantly, in many studies, causal trials were more difficult, and/or linguistic variables were not matched across causal and noncausal conditions. As a result, some of the observed effects may reflect linguistic or executive load. In addition, almost no prior studies localized language or logical reasoning networks in individual participants, making it difficult to assess the involvement of these systems (e.g. *Fedorenko et al., 2010*; *Monti et al., 2009*; *Pramod et al., 2023*). Most prior work also did not distinguish between causal inferences about different semantic domains known to depend on partially distinct neural networks, e.g., biological, mechanical, or mental state inferences (cf. *Mason and Just, 2011*; *Pramod et al., 2023*). If such inferences recruit partially distinct neural systems, their neural signatures might have been missed.

In the current experiment, participants read two-sentence vignettes (e.g. 'Hugh sat by sneezing passengers on the subway. Now he has a case of COVID.'). The first sentence described a potential cause and the second sentence a potential effect. Such causally connected sentences arise frequently in naturalistic discourse (*Singer, 1994*; *Graesser et al., 1994*). Participants performed a covert task of detecting 'magical' catch trial vignettes that encouraged them to attend to the meaning of the critical vignettes while reading as naturally as possible. We chose an orthogonal foil detection task rather than an explicit causal judgment task to investigate automatic causal inferences during reading and to unconfound such processing as much as possible from explicit decision-making processes. Analogous foil detection paradigms have been used to study sentence processing and word recognition (e.g. *Pallier et al., 2011*; *Dehaene-Lambertz et al., 2018*).

Causal inferences about illness were compared to two control conditions: (i) causal inferences about mechanical breakdown (e.g. 'Jake dropped all of his things on the subway. Now he has a shattered phone.') and (ii) illness-related language that was not causally connected (e.g. 'Lynn dropped all of her things on the subway. Now she has a case of COVID.'). This combination of control conditions allowed us to test jointly for sensitivity to content domain and causality. In other words, this design enabled us to test the hypothesis that causal inferences about illness recruit the animacy semantic network. Critically, all vignettes, including mechanical ones, described events involving people, such that responses to causal inferences about illness in the animacy semantic network could not be explained by the presence of animate agents. As a further control, we included the number of people in each vignette as a covariate of no interest in our fMRI analysis. Noncausal vignettes were constructed by shuffling causes/effects across conditions and were therefore matched to the causal vignettes in linguistic content. A separate group of participants rated the causal relatedness of all vignettes prior to the experiment. In addition to the main causal inference experiment, we also localized language, logical reasoning, and mentalizing networks in each participant. Following prior work, we predicted that the neural systems that support causal inference would exhibit increased activity during such inferences. Thus, our primary neural prediction was that animacy-responsive PC would respond more to causal inferences about illness compared to all other control conditions. We also used multivariate methods to investigate differences between conditions.

Results

Behavioral results

Accuracy on the magic detection task was at ceiling (M=97.9% ± 2.2 SD), and there were no significant differences across the four main experimental conditions (*Illness-Causal*, *Mechanical-Causal*, *Noncausal-Illness First*, *Noncausal-Mechanical First*), $F_{(3,57)}$ = 2.39, p=0.08. A one-way repeated-measures ANOVA evaluating response time revealed a main effect of condition, $F_{(3,57)}$ = 32.63, p<0.001, whereby participants were faster on *Illness-Causal* trials (M=4.73 ± 0.81 SD) compared to *Noncausal-Illness First* (M=5.33 s±0.85 SD) and *Noncausal-Mechanical First* (M=5.27 s±0.89 SD) trials. There were no differences in response time between the *Mechanical-Causal* condition (M=5.15 s±0.88 SD) and any other conditions. Performance on the localizer tasks was similar to previously reported studies that used these paradigms (see Appendix 3 for full behavioral results).

Inferring illness causes recruits animacy-responsive PC

We found distinctly localized neural responses to causal inferences about illness relative to both mechanical causal inferences and noncausal vignettes. A bilateral PC region previously implicated in thinking about animate entities (i.e. people and animals) responded preferentially to causal inferences about illness over both mechanical causal inferences and causally unrelated sentences in whole-cortex analysis ($p<0.05$, corrected for multiple comparisons; ***Figure 1C***) and in individual-subject overlap maps (***Figure 1—figure supplement 1; Figure 1—figure supplement 2***). PC responses during illness inferences overlapped with previously reported responses to people-related concepts (***Fairhall and Caramazza, 2013b***; ***Figure 1—figure supplement 3***).

Relative to illness inferences and noncausal vignettes, inferring the causes of mechanical breakdown in inanimate entities activated bilateral anterior parahippocampal regions (i.e. anterior PPA), suggesting a double dissociation between illness and mechanical inferences (***Figure 2***; ***Epstein and Kanwisher, 1998***; ***Weiner et al., 2018***). This anterior PPA region is engaged during memory/verbal tasks about physical spaces (***Baldassano et al., 2013***; ***Fairhall et al., 2014***; ***Silson et al., 2019***; ***Steel et al., 2021***; ***Häusler et al., 2022***; ***Hauptman et al., 2025***).

In individual-subject functional ROI (fROI) analysis (leave-one-run-out), we similarly found that inferring illness causes activated the PC more than inferring causes of mechanical breakdown (repeated-measures ANOVA, condition (*Illness-Causal*, *Mechanical-Causal*) × hemisphere (left, right): main effect of condition, $F_{(1,19)} = 19.18$, $p<0.001$, main effect of hemisphere, $F_{(1,19)} = 0.3$, $p=0.59$, condition × hemisphere interaction, $F_{(1,19)} = 27.48$, $p < 0.001$; ***Figure 1A***). This effect was larger in the left than in the right PC (paired samples t-tests; left PC: $t_{(19)} = 5.36$, $p<0.001$, right PC: $t_{(19)} = 2.27$, $p=0.04$). Illness inferences also activated the PC more than illness-related language that was not causally connected (repeated-measures ANOVA, condition (*Illness-Causal*, *Noncausal-Illness First*) × hemisphere (left, right): main effect of condition, $F_{(1,19)} = 4.66$, $p=0.04$, main effect of hemisphere, $F_{(1,19)} = 2.51$, $p=0.13$, condition × hemisphere interaction, $F_{(1,19)} = 8.07$, $p=0.01$; repeated-measures ANOVA, condition (*Illness-Causal*, *Noncausal-Mechanical First*) × hemisphere left, right: main effect of condition, $F_{(1,19)} = 4.38$, $p=0.05$; main effect of hemisphere, $F_{(1,19)} = 1.17$, $p = 0.29$; condition × hemisphere interaction, $F_{(1,19)} = 17.89$, $p<0.001$; ***Figure 1A***). Both effects were significant only in the left PC (paired samples t-tests; *Illness-Causal* vs. *Noncausal-Illness First*, left PC: $t_{(19)} = 2.77$, $p=0.01$, right PC: $t_{(19)} = 1.28$, $p=0.22$; *Illness-Causal* vs. *Noncausal-Mechanical First*, left PC: $t_{(19)} = 3.21$, $p=0.005$, right PC: $t_{(19)} = 0.5$, $p = 0.62$).

We also observed increased activity for illness inferences compared to mechanical inferences in the temporoparietal junction (TPJ) (leave-one-run-out individual-subject fROI analysis; repeated-measures ANOVA, condition (*Illness-Causal*, *Mechanical-Causal*) × hemisphere (left, right): main effect of condition, $F_{(1,19)} = 5.33$, $p=0.03$, main effect of hemisphere, $F_{(1,19)} = 1.02$, $p=0.33$, condition × hemisphere interaction, $F_{(1,19)} = 4.24$, $p=0.05$; ***Figure 1—figure supplements 4 and 5***). This effect was significant only in the left TPJ (paired samples t-tests; left TPJ: $t_{(19)} = 2.64$, $p=0.02$, right TPJ: $t_{(19)} = 1.13$, $p=0.27$). Unlike the PC, the TPJ did not show a preference for illness inferences compared to illness-related language that was not causally connected (repeated-measures ANOVA, condition (*Illness-Causal*, *Noncausal-Illness First*) × hemisphere (left, right): main effect of condition, $F_{(1,19)} = 0.006$, $p=0.94$, main effect of hemisphere, $F_{(1,19)} = 2.19$, $p=0.16$, condition × hemisphere interaction, $F_{(1,19)} = 1.27$, $p=0.27$; repeated-measures ANOVA, condition (*Illness-Causal*, *Noncausal-Mechanical First*) × hemisphere (left, right): main effect of condition, $F_{(1,19)} = 0.73$, $p=0.41$; main effect of hemisphere, $F_{(1,19)} = 1.24$, $p=0.28$; condition × hemisphere interaction, $F_{(1,19)} = 3.34$, $p=0.08$; ***Figure 1—figure supplements 4 and 5***).

In contrast to animacy-responsive PC, the anterior PPA showed the opposite pattern, responding more to mechanical inferences than illness inferences (leave-one-run-out individual-subject fROI analysis; repeated-measures ANOVA, condition (*Mechanical-Causal*, *Illness-Causal*) × hemisphere (left, right): main effect of condition, $F_{(1,19)} = 17.93$, $p<0.001$, main effect of hemisphere, $F_{(1,19)} = 1.33$, $p=0.26$, condition × hemisphere interaction, $F_{(1,19)} = 7.8$, $p=0.01$; ***Figure 2***). This effect was significant only in the left anterior PPA (paired samples t-tests; left anterior PPA: $t_{(19)} = 4$, $p<0.001$, right anterior PPA: $t_{(19)} = 1.88$, $p=0.08$). The anterior PPA also showed a preference for mechanical inferences compared to mechanical-related language that was not causally connected (repeated-measures ANOVA, condition (*Mechanical-Causal*, *Noncausal-Illness First*) × hemisphere (left, right): main effect of condition, $F_{(1,19)} = 14.81$, $p=0.001$, main effect of hemisphere, $F_{(1,19)} = 1.81$, $p=0.2$, condition × hemisphere interaction, $F_{(1,19)} = 7.35$, $p=0.01$; repeated-measures ANOVA, condition (*Mechanical-Causal*,

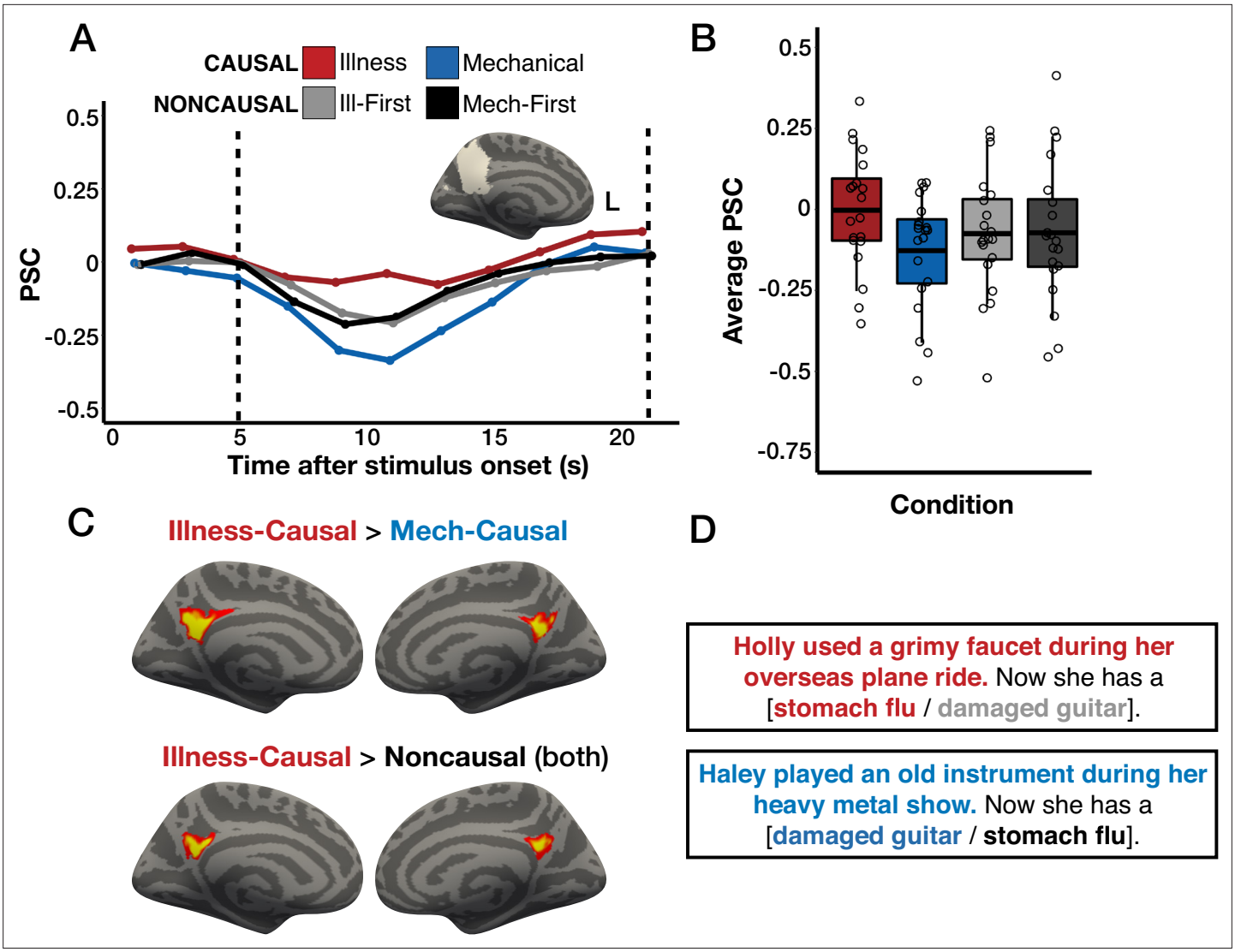

Figure 1. Responses to illness inferences in the precuneus (PC). (**A**) Percent signal change (PSC) for each condition among the top 5% *Illness-Causal>Mechanical-Causal* vertices in a left PC search space (***Dufour et al., 2013***) in individual participants, established via a leave-one-run-out analysis. (**B**) Average PSC in the critical window (marked by dotted lines in A) across participants. The horizontal line within each boxplot indicates the overall mean. (**C**) Whole-cortex results (one-tailed) for *Illness-Causal>Mechanical-Causal* and *Illness-Causal>Noncausal* (both versions of noncausal vignettes), corrected for multiple comparisons ($p<0.05$ family-wise error rate [FWER], cluster-forming threshold $p<0.01$ uncorrected). Vertices are color-coded on a scale from $p=0.01$ to $p=0.00001$. (**D**) Example stimuli. 'Magical' catch trials similar in meaning and structure (e.g. 'Sadie forgot to wash her face after she ran in the heat. Now she has a cucumber nose.') enabled the use of a semantic 'magic detection' task.

The online version of this article includes the following figure supplement(s) for figure 1:

Figure supplement 1. Group overlap in univariate contrasts comparing causal (*Illness-Causal*, *Mechanical-Causal*) and noncausal conditions (*Noncausal-Illness First + Noncausal-Mechanical First*) in the precuneus (PC), winner-take-all approach.

Figure supplement 2. Group overlap in univariate contrasts comparing causal (*Illness-Causal*, *Mechanical-Causal*) and noncausal conditions (*Noncausal-Illness First + Noncausal-Mechanical First*) in the precuneus (PC).

Figure supplement 3. Overlap between left precuneus (PC) responses to illness inferences in the current study and people-related stimuli in a separate study (***Fairhall and Caramazza, 2013b***).

Figure supplement 4. Responses to illness inferences in bilateral precuneus (PC) and temporoparietal junction (TPJ).

Figure supplement 5. Subject dispersion data for responses to illness inferences in bilateral precuneus (PC) and temporoparietal junction (TPJ) (see ***Figure 1—figure supplement 4***).

Figure supplement 6. Percent signal change (PSC) for each condition among the top 5% Illness-Causal>Mechanical-Causal vertices in a left precuneus (PC) search space (***Dufour et al., 2013***) in individual participants, established via a leave-one-run-out analysis.

Figure 1 continued on next page

Figure 1 continued

Figure supplement 7. Functional localization of language, logical reasoning, and mentalizing networks (see ***Monti et al., 2009***; ***Fedorenko et al., 2010***; ***Dodell-Feder et al., 2011***; ***Liu et al., 2020***).

Figure supplement 8. Full whole-cortex univariate results.

Figure supplement 9. Comparison of whole-cortex results for number of people in each vignette (left) and illness inferences (right) from the same generalized linear model (GLM).

Figure supplement 10. Searchlight MVPA group maps.

Figure supplement 11. Subject dispersion data for individual-subject MVPA performed in functional ROIs (fROIs).

Figure supplement 12. Subject dispersion data for individual-subject MVPA performed in functional ROIs (fROIs).

Figure supplement 13. Responses to illness inferences in the fusiform face area (FFA).

Figure supplement 14. Comparison of mentalizing localizers used in previous work and in the current study, in three pilot participants.

Noncausal-Mechanical First) × hemisphere (left, right): main effect of condition, $F_{(1,19)} = 11.31$, p=0.003; main effect of hemisphere, $F_{(1,19)} = 3.34$, p=0.08; condition × hemisphere interaction, $F_{(1,19)} = 4$, p=0.06; ***Figure 2***). Similar to the PC, both effects were larger in the left than in the right hemisphere (post hoc paired samples t-tests; *Illness-Causal* vs. *Noncausal-Illness First*, left anterior PPA: $t_{(19)} = 3.85$, p=0.001, right anterior PPA: $t_{(19)} = 2.22$, p=0.04; *Illness-Causal* vs. *Noncausal-Mechanical First,* left anterior PPA: $t_{(19)} = 3.59$, p=0.002, right anterior PPA: $t_{(19)} = 1.19$, p=0.25).

In summary, we found distinctly localized responses to illness and mechanical causal inferences. Inferring illness causes preferentially recruited the animacy semantic network, particularly the PC.

Illness inferences and mental state inferences elicit spatially dissociable responses

Illness inferences and mental state inferences elicited spatially dissociable responses. In whole-cortex analysis, illness inferences recruited the PC bilaterally, with larger responses observed in the left hemisphere (***Figure 1***, see also fROI analysis showing left-lateralization above). By contrast, and in accordance with prior work (e.g. ***Saxe and Kanwisher, 2003***), mental state inferences recruited a broader

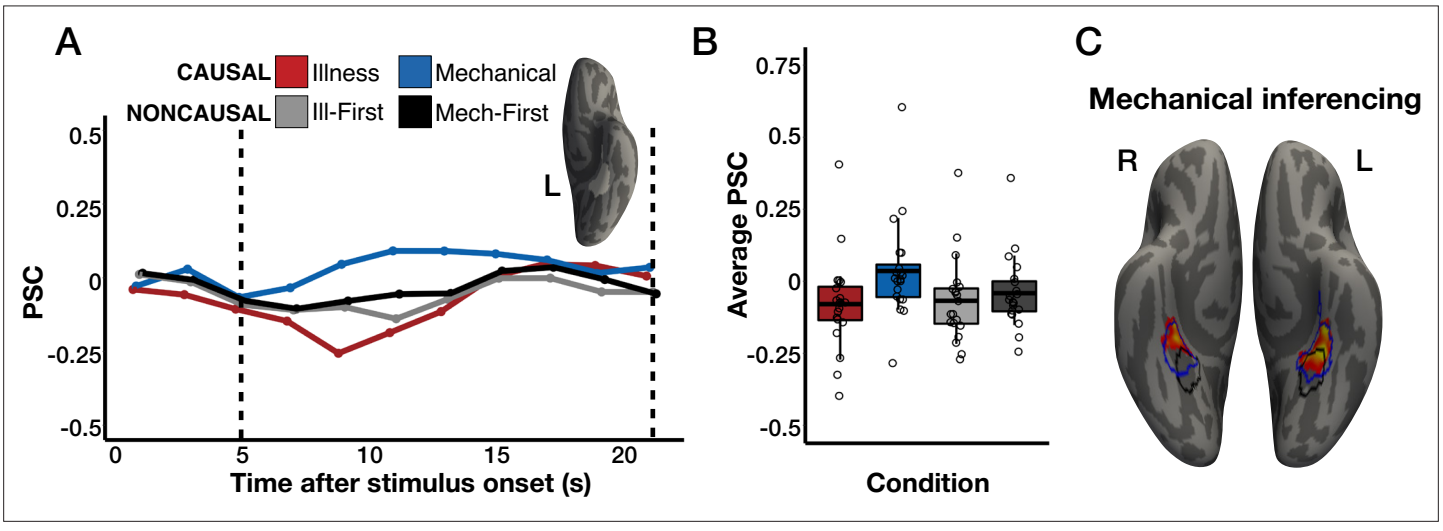

Figure 2. Responses to mechanical inferences in anterior parahippocampal regions (anterior PPA). (**A**) Percent signal change (PSC) for each condition among the top 5% *Mechanical-Causal>Illness-Causal* vertices in a left anterior PPA search space (***Hauptman et al., 2025***) in individual participants, established via a leave-one-run-out analysis. (**B**) Average PSC in the critical window (marked by dotted lines in A) across participants. The horizontal line within each boxplot indicates the overall mean. (**C**) The intersection of two whole-cortex contrasts (one-tailed), *Mechanical-Causal>Illness-Causal* and *Mechanical-Causal>Noncausal* that are corrected for multiple comparisons (p<0.05 family-wise error rate [FWER], cluster-forming threshold p<0.01 uncorrected). Vertices are color-coded on a scale from p=0.01 to p=0.00001. Similar to PC responses to illness inferences, anterior PPA is the only region to emerge across both mechanical inference contrasts. The average PPA location from a separate study involving perceptual place stimuli (***Weiner et al., 2018***) is overlaid in black. The average PPA location from a separate study involving verbal place stimuli (***Hauptman et al., 2025***) is overlaid in blue.

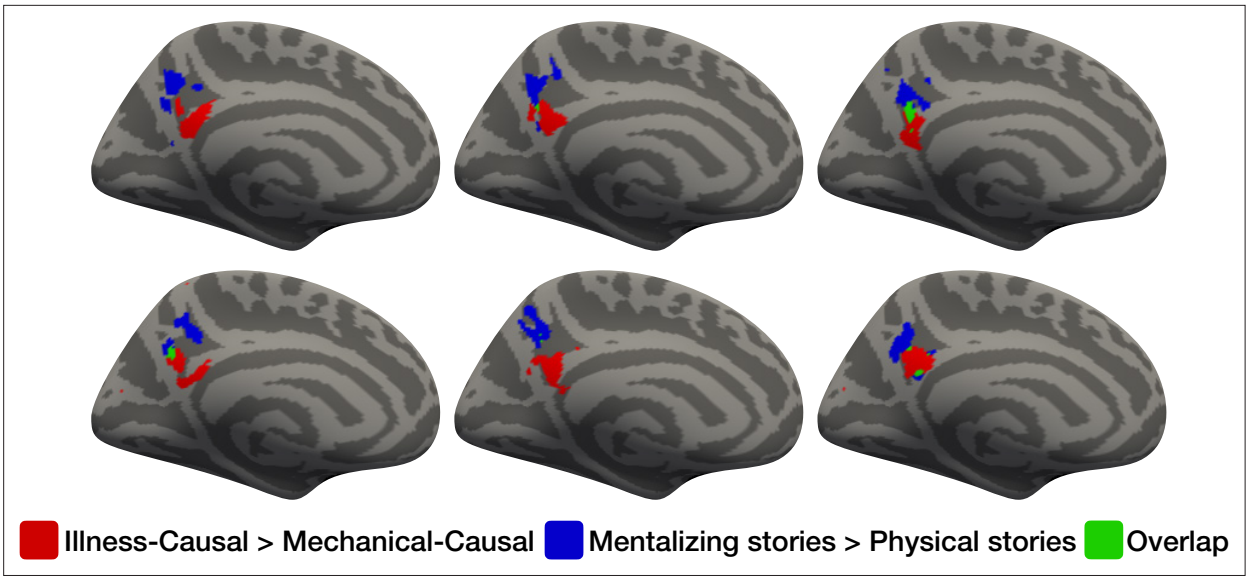

Figure 3. Spatial dissociation between univariate responses to illness inferences and mental state inferences in the precuneus (PC). The left medial surface of six individual participants were selected for visualization purposes. The locations of the top 10% most responsive vertices to *Illness-Causal>Mechanical-Causal* in a PC search space (***Dufour et al., 2013***) are shown in red. The locations of the top 10% most responsive vertices to *mentalizing stories >physical stories* (mentalizing localizer) in the same PC search space are shown in blue. Overlapping vertices are shown in green.

The online version of this article includes the following figure supplement(s) for figure 3:

Figure supplement 1. Spatial dissociation between responses to illness inferences and mental state inferences in left precuneus (PC).

network, including not only bilateral PC, but also bilateral TPJ, superior temporal sulcus, and medial and superior prefrontal cortex (***Figure 1—figure supplement 7***).

Within the left PC, responses to illness inferences were located ventrally to mental state inference responses (***Figure 3***, ***Figure 3—figure supplement 1***). The z-coordinates of individual-subject activation peaks for illness inferences and mental state inferences were significantly different (repeated-measures ANOVA, $F_{(1,19)}$ = 13.52, p=0.002). In addition, the size of the illness inference effect (*Illness-Causal >Mechanical-Causal*) was larger in illness-responsive vertices (leave-one-run-out individual-subject fROI analysis) than in mentalizing-responsive vertices in the left PC (individual-subject fROI analysis; repeated-measures ANOVA, $F_{(1,19)}$ = 24.72, p<0.001, ***Figure 1—figure supplements 4 and 5***). These results suggest that illness inferences and mental state inferences are carried out by neighboring but partially distinct subsets of the PC.

No univariate evidence for domain-general responses to implicit causal inference

Prior neuroscience studies hypothesizing the existence of a domain-general 'causal engine' have predicted that the language network and/or domain-general executive systems (e.g. the logic network) should show elevated activity during causal inference across domains. In the current study, neither the language nor the logic network exhibited elevated neural responses during causal inferences relative to linguistically matched sentence pairs that were not causally connected. Language regions in fronto-temporal cortex responded more to noncausal than causal vignettes (frontal search space: repeated-measures ANOVA, $F_{(1,19)}$ = 23.91, p<0.001; temporal search space: repeated-measures ANOVA, $F_{(1,19)}$ = 4.31, p=0.05; ***Figure 4***, ***Figure 4—figure supplement 1***). The logic network likewise responded marginally more to noncausal vignettes, likely reflecting greater difficulty associated with integrating unrelated sentences (repeated-measures ANOVA, $F_{(1,19)}$ = 3.88, p=0.07; ***Figure 4***).

In whole-cortex univariate analysis, no shared regions responded more to causal than noncausal vignettes across domains. Two whole-cortex univariate contrasts comparing causal and noncausal conditions (*Illness-Causal>Noncausal-Mechanical First*, *Mechanical-Causal>Noncausal-Mechanical First*) revealed increased activity for the noncausal condition in bilateral prefrontal cortex. The same prefrontal areas that responded more to noncausal than causal stimuli also responded more when

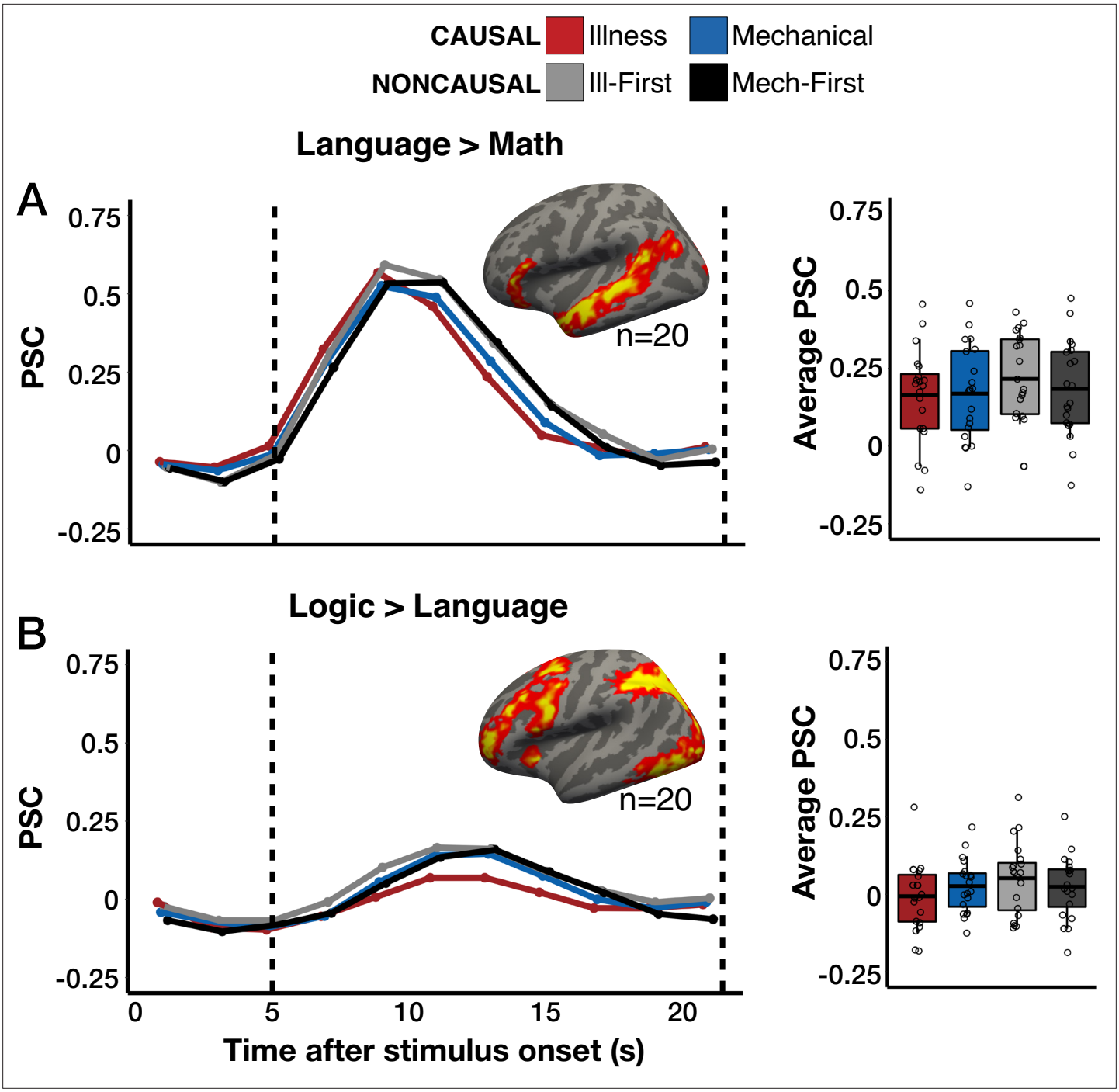

Figure 4. Individual-subject analysis of language- and logic-responsive vertices. (**A**) Percent signal change (PSC) for each condition among the top 5% most language-responsive vertices (language>*math*) in a temporal language network search space (***Fedorenko et al., 2010***). Results from a frontal language search space (***Fedorenko et al., 2010***) can be found in ***Figure 4—figure supplement 1***. (**B**) PSC among the top 5% most logic-responsive vertices (logic>*language*) in a logic network search space (***Liu et al., 2020***). Group maps for each contrast of interest (one-tailed) are corrected for multiple comparisons ($p<0.05$ family-wise error rate [FWER], cluster-forming threshold $p<0.01$ uncorrected). Vertices are color-coded on a scale from $p=0.01$ to $p=0.00001$. Boxplots display average PSC in the critical window (marked by dotted lines) across participants. The horizontal line within each boxplot indicates the overall mean.

The online version of this article includes the following figure supplement(s) for figure 4:

Figure supplement 1. Responses to causal inference in the language network.

participants were slower to complete the task, suggesting that these responses reflect a nonspecific difficulty effect (***Figure 1—figure supplement 8***).

In summary, none of the predicted networks nor any regions across the whole cortex exhibited the predicted domain-general causal inference pattern, i.e., larger responses to all causal than all noncausal vignettes. These results suggest that implicit causal inferences, which draw upon a person's existing knowledge of relevant causes and effects, do not depend on domain-general neural mechanisms. These results leave open the possibility that domain-general systems support the explicit search for causal connections (see Discussion section).

Multivariate analysis

In searchlight MVPA performed across the whole cortex, illness inferences and mechanical inferences produced spatially distinguishable neural patterns in the left PC extending dorsally into the superior parietal lobule, as well as in left anterior PPA and lateral occipitotemporal cortex. A whole-cortex searchlight analysis that tested whether each causal condition could be decoded from each noncausal condition found no shared regions that exhibited significant decoding across all causal vs. noncausal comparisons (***Figure 1—figure supplement 10***).

In individual-subject fROI decoding analyses, illness inferences and mechanical inferences produced spatially distinguishable neural patterns in the left PC, right PC, and left TPJ, as well as in language and logic networks (see ***Figure 1—figure supplement 12***, ***Supplementary file 2*** for full results). Note that these decoding results must be interpreted in light of the significant univariate differences observed across conditions that are reported above. Linear classifiers are highly sensitive to univariate differences (***Coutanche, 2013***; ***Kragel et al., 2012***; ***Hebart and Baker, 2018***; ***Woolgar et al., 2014***; ***Davis et al., 2014***; ***Pakravan et al., 2022***). Successful decoding may be driven by univariate differences in the predicted direction (e.g. causal>noncausal) or in the opposite direction (e.g. noncausal>causal). In particular, given that both the language and the logic networks exhibited higher univariate responses to noncausal compared to causal vignettes, decoding results observed in these networks may be driven by univariate differences.

Discussion

Causal knowledge is embedded in high-level semantic networks

We find that a semantic network previously implicated in thinking about animates, particularly the (PC), is preferentially engaged when people infer causes of illness compared to when they infer causes of mechanical breakdown or read causally unconnected sentences containing illness-related language. By contrast, mechanical inferences activate an anterior parahippocampal region previously implicated in thinking about and remembering places (***Baldassano et al., 2013***; ***Fairhall et al., 2014***; ***Silson et al., 2019***; ***Steel et al., 2021***; ***Häusler et al., 2022***; ***Hauptman et al., 2025***). This finding points to a neural double dissociation between biological and mechanical causal knowledge.

Previous work has implicated the PC in the representation of animate entities, i.e., people and animals (***Fairhall and Caramazza, 2013a***; ***Fairhall et al., 2014***; ***Peer et al., 2015***; ***Wang et al., 2016***; ***Silson et al., 2019***; ***Rabini et al., 2021***; ***Deen and Freiwald, 2022***; ***Aglinskas and Fairhall, 2023***; ***Hauptman et al., 2025***). Here, we show that the PC exhibits sensitivity to causal inferences about biological processes specific to animates, such as illness. These findings are consistent with our preregistered hypotheses and suggest that causal knowledge about animate and inanimate entities is distributed across multiple distinct semantic networks. Further, our results suggest that the animacy semantic network supports biological causal knowledge. Future work should test whether the animacy network is sensitive to causal information beyond illness, including about growth, nourishment, and death. We hypothesize that changes in biological causal knowledge during development, as well as cultural expertise in causal reasoning about illness (e.g. medical expertise), influence activity in the animacy network (***Legare et al., 2012***; ***Norman et al., 2009***).

Our findings are consistent with prior evidence from naturalistic paradigms showing that the PC is sensitive to discourse-level processes across sentences (e.g. ***Hasson et al., 2008***; ***Lerner et al., 2011***; ***Lee and Chen, 2022***). We hypothesize that PC responses observed during naturalistic narrative comprehension are driven by causal inferences about animate agents, who are often the focus of narratives. Likewise, PC involvement in episodic memory could be related to animacy-related inferential

processes (**DiNicola et al., 2020**; **Ritchey and Cooper, 2020**). Future work can test this hypothesis by comparing causal inferences about animate and inanimate entities in naturalistic contexts, such as films and verbal narratives (see **Chen and Bornstein, 2024**, for a review on causal inference in narrative comprehension).

We find that neural responses during inferences about biological and mental properties of animates are linked yet separable. Inferring illness causes recruits neural circuits that are adjacent to but distinct from responses to mental state inferences in the PC (**Saxe and Kanwisher, 2003**; **Saxe et al., 2006**). Even young children provide different causal explanations for biological vs. psychological processes (**Springer and Keil, 1991**; **Callanan and Oakes, 1992**; **Wellman and Gelman, 1992**; **Inagaki and Hatano, 1993**; **Inagaki and Hatano, 2004**; **Keil, 1994**; **Hickling and Wellman, 2001**; **Medin et al., 2010**; cf. **Carey, 1985**; see also **Medin and Atran, 2004**). For example, when asked why blood flows to different parts of the body, 6-year-old endorse explanations referring to bodily function, e.g., 'because it provides energy to the body,' and not to mental states, e.g., 'because we want it to flow' (**Inagaki and Hatano, 1993**). At the same time, animate entities have a dual nature: they have both bodies and minds (**Opfer and Gelman, 2011**; **Spelke, 2022**). The current findings point to the existence of distinct but related neural systems for biological and mentalistic knowledge.

Our neuroimaging findings are consistent with evidence from developmental psychology suggesting that causal knowledge is central to human concepts starting early in development (**Keil, 1992**; **Wellman and Gelman, 1992**; **Hatano and Inagaki, 1994**; **Springer and Keil, 1991**; **Simons and Keil, 1995**; **Atran, 1998**; **Keil et al., 1999**; **Coley et al., 2002**; **Medin and Atran, 2004**). According to the 'intuitive theories' account, semantic knowledge is organized into causal frameworks that serve as 'grammars for causal inference' (**Tenenbaum et al., 2007**; **Wellman and Gelman, 1992**; **Gopnik and Meltzoff, 1997**; **Gopnik and Wellman, 2012**; **Gerstenberg and Tenenbaum, 2017**; see also **Boyer, 1995**; **Barrett et al., 2007**; **Cosmides and Tooby, 2013**; **Bender et al., 2017**). For example, preschoolers intuit that animates but not inanimate objects get sick and need nourishment to grow and live (e.g. **Rosengren et al., 1991**; **Kalish, 1996**; **Gutheil et al., 1998**; **Raman and Gelman, 2005**; see **Inagaki and Hatano, 2004**; **Opfer and Gelman, 2011**, for reviews). The present results suggest that such knowledge is encoded in high-level semantic brain networks. By contrast, we failed to find sensitivity to causal inference in portions of the ventral stream previously associated with the perception of animate agents (see Appendix 4, **Figure 1—figure supplement 13** for details). Sensitivity to causal information may be a distinguishing characteristic of high-level, amodal semantic networks, as opposed to perceptual regions that are activated during semantic tasks (e.g. **Martin and Chao, 2001**; **Thompson-Schill, 2003**; **Barsalou et al., 2003**; **Binder and Desai, 2011**; **Bi, 2021**).

No evidence for domain-general neural responses during implicit causal inference

In the current study, participants read two sentence vignettes that either elicited causal inferences or were not causally connected. No brain regions responded more to causal inferences across domains compared to noncausal vignettes in this task. The language network responded more to noncausal than causal vignettes, possibly due to greater difficulty associated with processing the meaning of a sentence that does not follow from the prior context. Prior studies find that the language network is specialized primarily for sentence-internal processing (**Fedorenko and Varley, 2016**; **Jacoby and Fedorenko, 2020**; **Blank and Fedorenko, 2020**) and patients with agrammatic aphasia can make causal inferences about pictorial stimuli (**Varley and Siegal, 2000**; **Varley, 2014**). Together, these results suggest that the language system itself is unlikely to support causal inference. Rather, during language comprehension, the language system interacts with semantic networks to enable causal inference (**Simony et al., 2016**; **Yeshurun et al., 2017**; **Chang et al., 2022**). Notably, in the current study, responses to causal inference in semantic networks were stronger in the left hemisphere. The left lateralization of such responses may enable efficient interfacing with the language system during comprehension.

We also failed to find evidence for the claim that the frontoparietal logical reasoning network, a domain-general executive system, supports implicit causal inferences. By contrast, the frontoparietal network responded more to noncausal than causal vignettes. Finally, we failed to observe elevated responses to causal inference across domains anywhere in the brain in whole-cortex analysis. A large swath of prefrontal cortex responded more to one noncausal condition (*Noncausal-Mechanical First*)

compared to both causal conditions. The same prefrontal regions also exhibited increased activity when participants were slower to respond to the task. Thus, this 'reverse causality effect' likely reflects processing demands rather than causal inference per se. An alternative interpretation of the elevated prefrontal activity observed for one of the noncausal conditions is that it reflects the effortful search for a causal connection between sentences when such a connection is difficult to find. This interpretation would suggest that domain-general executive mechanisms become engaged when causal inferences are effortful and explicit. By contrast, semantic systems are engaged when we implicitly infer a known causal relationship.

Causal inferences are a highly varied class, and domain-general systems likely play an important role in many causal inferences not tested in the current study. The vignettes used in the current study stipulate illness causes, allowing participants to reason from causes to effects. By contrast, illness reasoning performed by medical experts proceeds from effects to causes and can involve searching for potential causes within highly complicated and interconnected causal systems (***Schmidt et al., 1990***; ***Norman et al., 2009***; ***Meder and Mayrhofer, 2017***). The discovery of novel causal relationships (e.g. 'blicket detectors'; ***Gopnik et al., 2001***) and the identification of complex causes, even in the case of illness, may depend in part on domain-general neural mechanisms. The present results suggest, however, that causal knowledge is embedded within high-level semantic systems, and that biological causal knowledge is embedded with a semantic system relevant to animacy.

Materials and methods

Open science practices

The methods and analysis of this experiment were preregistered prior to data collection (https://osf.io/6pnqg).

Participants

Twenty adults (7 women, 13 men, 25–37 years of age, *M*=28.7 years±3.2 SD) participated in the study. Participants either had or were pursuing graduate degrees (*M*=8.8 years of post-secondary education). Two additional participants were excluded from the final dataset due to excessive head motion (>2 mm) and an image artifact. One participant in the final dataset exhibited excessive head motion (>2 mm) during one run of the language/logic localizer task that was excluded from analysis. All participants were screened for cognitive and neurological disabilities (self-report). Participants gave written informed consent and were compensated $30 per hour. The study was reviewed and approved by the Johns Hopkins Medicine Institutional Review Boards (IRB00270868).

Causal inference experiment

Stimuli

Participants read two-sentence vignettes in four conditions, two causal and two noncausal (***Figure 1D***). Each vignette focused on a single agent, specified by a proper name in the initial sentence and by a pronoun in the second sentence. The first sentence described something the agent did or experienced and served as the potential cause. The second sentence described the potential effect (e.g. 'Kelly shared plastic toys with a sick toddler at her preschool. Now she has a case of chickenpox.'). *Illness-Causal* vignettes elicited inferences about biological causes of illness, including pathogen transmission, exposure to environmental toxins, and genetic mutations (see ***Supplementary file 1*** for a full list of the types of illnesses included in our stimuli). *Mechanical-Causal* vignettes elicited inferences about physical causes of structural damage to personally valuable inanimate objects (e.g. houses, jewelry). Two noncausal conditions used the same sentences as in the *Illness-Causal* and *Mechanical-Causal* conditions but in a shuffled order: illness cause with mechanical effect (*Noncausal-Illness First*) or mechanical cause with illness effect (*Noncausal-Mechanical First*). Explicit causality judgments collected from a separate group of online participants (*n*=26) verified that both causal conditions *Illness-Causal*, *Mechanical-Causal* were more causally related than both noncausal conditions, $t(25) = 36.97$, $p<0.001$. In addition, *Illness-Causal* and *Mechanical-Causal* items received equally high causality ratings, $t(25) = -0.64$, $p=0.53$ (see Appendix 1 for details).

Illness-Causal and *Mechanical-Causal* vignettes were constructed in pairs, such that each member of a given pair shared parallel or near-parallel phrase structure. All conditions were also matched

(pairwise t-tests, all ps>0.3, no statistical correction) on multiple linguistic variables known to modulate neural activity in language regions (e.g. ***Pallier et al., 2011***; ***Shain et al., 2020***). These included number of characters, number of words, average number of characters per word, average word frequency, average bigram surprisal (Google Books Ngram Viewer, https://books.google.com/ngrams/), and average syntactic dependency length (Stanford Parser; ***Marneffe et al., 2006***). Word frequency was calculated as the negative log of a word's frequency in the Google corpus between the years 2017 and 2019. Bigram surprisal was calculated as the negative log of the frequency of a given two-word phrase in the Google corpus divided by the frequency of the first word of the phrase (see Appendix 2 for details). All conditions were matched for all linguistic variables across the first sentence, second sentence, and the entire vignette.

Procedure

We used a 'magic detection' task to encourage participants to process the meaning of the vignettes without making explicit causality judgments. Participants saw 'magical' catch trials that closely resembled the experimental trials but were fantastical (e.g. 'Sadie forgot to wash her face after she ran in the heat. Now she has a cucumber nose.'). On each trial, participants indicated via button press whether 'something magical' occurred in the vignette (Yes/No). This semantic foil detection task encouraged participants to attend to the meaning of the critical vignettes while reading as naturally as possible. We required participants to press a button on every trial to ensure they were attending to the stimuli. Both sentences in a given vignette were presented simultaneously for 7 s, one above the other, followed by a 12 s inter-trial interval. Each participant saw 38 trials per condition (152 trials) plus 36 'magical' catch trials (188 total trials) in one of two versions, counterbalanced across participants, such that individual participants did not see the same sentence in both causal and noncausal vignettes. The two stimulus versions had similar meanings but different surface forms (e.g. 'Luna stood by coughing travelers on the train…' vs. 'Hugh sat by sneezing passengers on the subway…').

The experiment was divided into six 10 min runs containing six to seven trials per condition per run presented in a pseudorandom order. Vignettes from the same experimental condition repeated no more than twice consecutively, vignettes that shared similar phrase structure never repeated within a run, vignettes that referred to the same illness never repeated consecutively, and vignettes from each condition, including catch trials, were equally distributed in time across the course of the experiment.

Mentalizing localizer experiment

To test the relationship between neural responses to inferences about the body and the mind, and to localize animacy regions, we used a localizer task to identify the mentalizing network in each participant (***Saxe and Kanwisher, 2003***; ***Dodell-Feder et al., 2011***; https://saxelab.mit.edu/use-our-efficient-false-belief-localizer/). In this task, participants read 10 mentalizing stories (e.g. a protagonist has a false belief about an object's location) and 10 physical stories (physical representations depicting outdated scenes, e.g., a photograph showing an object that has since been removed) before answering a true/false comprehension question. We used the mentalizing stories from the original localizer but created new stimuli for the physical stories condition. Our physical stories incorporated more vivid descriptions of physical interactions and did not make any references to human agents, enabling us to use the mentalizing localizer as a localizer for animacy. The new physical stories were also linguistically matched to the mentalizing stories to reduce linguistic confounds (see ***Shain et al., 2023***). Specifically, we matched physical and mentalizing stories (pairwise t-tests, all ps >0.3, no statistical correction) for number of characters, number of words, average number of characters per word, average syntactic dependency length, average word frequency, and average bigram surprisal, as was done for the causal inference vignettes. A comparison of both localizer versions in three pilot participants can be found in ***Figure 1—figure supplement 14***.

Trials were presented in an event-related design, with each one lasting 16 s (12 s stories + 4 s comprehension question) followed by a 12 s inter-trial interval. Participants completed 2 5 min runs of the task, with trial order counterbalanced across runs and participants. The mentalizing network was identified in individual participants by contrasting *mentalizing stories* > *physical stories* (***Saxe and Kanwisher, 2003***; ***Dodell-Feder et al., 2011***).

Language/logic localizer experiment

To test for the presence of domain-general responses to causal inference in the language and logic networks (e.g. ***Kuperberg et al., 2006***; ***Operskalski and Barbey, 2017***), we used an additional localizer task. The task had three conditions: language, logic, and math. In the language condition, participants judged whether two visually presented sentences, one in active and one in passive voice, shared the same meaning. In the logic condition, participants judged whether two logical statements were consistent (e.g. *If either not Z or not Y then X* vs. *If not X then both Z and Y*). In the math condition, participants judged whether the variable *X* had the same value across two equations (for details, see ***Liu et al., 2020***). Trials lasted 20 s (1 s fixation + 19 s display of stimuli) and were presented in an event-related design. Participants completed two 9 min runs of the task, with trial order counterbalanced across runs and participants. Following prior studies, the language network was identified in individual participants by contrasting *language > math* and the logic network by contrasting *logic > language* (***Monti et al., 2009***; ***Kanjlia et al., 2016***; ***Liu et al., 2020***).

Data acquisition

Whole-brain fMRI data was acquired at the F.M. Kirby Research Center for Functional Brain Imaging on a 3T Phillips Achieva Multix X-Series scanner. T1-weighted structural images were collected in 150 axial slices with 1 mm isotropic voxels using the magnetization-prepared rapid gradient-echo (MP-RAGE) sequence. Functional T2*-weighted BOLD scans were collected using a gradient echo planar imaging (EPI) sequence with the following parameters: 36 sequential ascending axial slices, repetition time (TR)=2 s, echo time (TE)=0.03 s, flip angle = 70°, field of view (FOV) matrix = 76 × 70, slice thickness = 2.5 mm, inter-slice gap = 0.5, slice-coverage FH = 107.5, voxel size = 2.4×2.4×3 mm^3, PE direction = L/R, first order shimming. Data were acquired in one experimental session lasting approximately 120 min. All stimuli were visually presented on a rear projection screen with a Cambridge Research Systems BOLDscreen 32 UHD LCD display (image resolution = 1920 × 1080) using custom scripts written in PsychoPy3 (https://www.psychopy.org/, ***Peirce et al., 2019***). Participants viewed the screen via a front-silvered, 45° inclined mirror attached to the top of the head coil.

fMRI data preprocessing and GLM analysis

Preprocessing included motion correction, high-pass filtering (128 s), mapping to the cortical surface (Freesurfer), spatially smoothing on the surface (6 mm FWHM Gaussian kernel), and prewhitening to remove temporal autocorrelation. Covariates of no interest included signal from white matter, cerebral spinal fluid, and motion spikes.

For the main causal inference experiment, the generalized linear model (GLM) modeled the four main conditions (*Illness-Causal*, *Mechanical-Causal*, *Noncausal-Illness First, Noncausal-Mechanical First*) and the 'magical' catch trials during the 7 s display of the vignettes after convolving with a canonical hemodynamic response function and its first temporal derivative. The GLM additionally included participant response time and number of people in each vignette as covariates of no interest. For the mentalizing localizer experiment, a separate predictor was included for each condition (*mentalizing stories, physical stories*), modeling the 16 s display of each story and corresponding comprehension question. For the language/logic localizer experiment, a separate predictor was included for each of the three conditions (*language, logic, math*), modeling the 20 s duration of each trial.

For each task, runs were modeled separately and combined within-subject using a fixed-effects model (***Dale et al., 1999***; ***Smith et al., 2004***). Group-level random-effects analyses were corrected for multiple comparisons across the whole cortex at $p<0.05$ family-wise error rate (FWER) using a nonparametric permutation test (cluster-forming threshold $p<0.01$ uncorrected) (***Winkler et al., 2014***; ***Eklund et al., 2016***; ***Eklund et al., 2019***).

Individual-subject fROI analysis: univariate

We defined individual-subject fROIs in the PC and TPJ, as well as in the language (frontal and temporal search spaces) and logic networks. In an exploratory analysis, we defined individual-subject fROIs in an anterior parahippocampal region (i.e. anterior PPA). For all analyses, percent signal change (PSC) was extracted and averaged over the entire duration of the trial (17 s total), starting at 4 s to account for hemodynamic lag.

Illness inference fROIs were created in bilateral PC and TPJ group search spaces (**Dufour et al., 2013**) using an iterated leave-one-run-out procedure, which allowed us to perform sensitive individual-subjects analysis while avoiding statistical nonindependence (**Vul and Kanwisher, 2011**). In each participant, we identified the most illness inference-responsive vertices in bilateral PC and TPJ search spaces separately in five of the six runs (top 5% of vertices, *Illness-Causal>Mechanical-Causal*). We then extracted PSC for each condition compared to rest in the held-out run (*Illness-Causal*, *Mechanical-Causal*, *Noncausal-Illness First, Noncausal-Mechanical First*), averaging the results across all iterations. We used the same approach to create mechanical inference fROIs in bilateral anterior PPA search spaces from a previous study on place word representations (**Hauptman et al., 2025**). All aspects of this analysis were the same as those described above, except that the most mechanical inference-responsive vertices (top 5%, *Mechanical-Causal>Illness-Causal*) were selected.

Mentalizing fROIs were created by selecting the most mentalizing-responsive vertices (top 5%) in bilateral PC and TPJ search spaces (**Dufour et al., 2013**) using the *mentalizing stories>physical stories* contrast from the mentalizing localizer. Language fROIs were identified by selecting the most language-responsive vertices (top 5%) in left frontal and temporal language areas (search spaces: **Fedorenko et al., 2010**) using the language>*math* contrast from the language/logic localizer. A logic-responsive fROI was identified by selecting the most logic-responsive vertices (top 5%) in a left frontoparietal network (search space: **Liu et al., 2020**) using the *logic>language* contrast. In each fROI, we extracted PSC for all conditions in the causal inference experiment.

Individual-subject fROI analysis: multivariate

We performed MVPA (PyMVPA toolbox; **Hanke et al., 2009**) to test whether patterns of activity in the PC, TPJ, language network, and logic network distinguished illness inferences from mechanical inferences. In each participant, we identified the top 300 vertices most responsive to the mentalizing localizer (*mentalizing stories>physical stories*) in bilateral PC and TPJ search spaces (**Dufour et al., 2013**). We also identified the top 300 vertices most responsive to language (*language>math*) in a left language network search space (**Fedorenko et al., 2010**) and the top 300 vertices most responsive to logical reasoning (*logic>language*) in a left logic network search space (**Liu et al., 2020**).

In an exploratory analysis, we performed MVPA to test whether patterns of activity in the left PC and in the language and logic networks distinguished causal from noncausal vignettes. To avoid statistical nonindependence, we defined additional fROIs in the left PC for the purposes of this analysis. In each participant, we identified the top 300 vertices most responsive to the critical conditions over rest (*Illness-Causal+Mechanical-Causal+Noncausal-Illness First +Noncausal-Mechanical First>Rest*) in a left PC search space (**Dufour et al., 2013**).

For each vertex in each participant's fROIs, we obtained one observation per condition per run (z-scored beta parameter estimate of the GLM). A linear support vector machine (SVM) was then trained on data all but one of the runs and tested on the left-out run in a cross-validation procedure. Classification accuracy was averaged across all permutations of the training/test splits. We compared classifier performance within each fROI to chance (50%; one-tailed test). Significance was evaluated against an empirically generated null distribution using a combined permutation and bootstrap approach (**Schreiber and Krekelberg, 2013**; **Stelzer et al., 2013**). In this approach, t-statistics obtained for the observed data are compared against an empirically generated null distribution. We report the t-values obtained for the observed data and the nonparametric p-values, where p corresponds to the proportion of the shuffled analyses that generated a comparable or higher t-value.

The null distribution was generated using a balanced block permutation test by shuffling condition labels within run 1000 times for each subject (**Schreiber and Krekelberg, 2013**). Then, a bootstrapping procedure was used to generate an empirical null distribution for each statistical test across participants by sampling one permuted accuracy value from each participant's null distribution 15,000 times (with replacement) and running each statistical test on these permuted samples, thus generating a null distribution of 15,000 statistical values for each test (**Stelzer et al., 2013**).

Searchlight MVPA

We used a linear SVM classifier to test decoding between all pairs of causal and noncausal conditions (i.e. *Illness-Causal* vs. *Mechanical-Causal*, *Illness-Causal* vs. *Noncausal-Mechanical First*, *Illness-Causal* vs. *Noncausal-Illness First*, *Mechanical-Causal* vs. *Noncausal-Mechanical First*, and *Mechanical-Causal*

vs. *Noncausal-Illness First*) across the whole cortex using a 10 mm radius spherical searchlight (according to geodesic distance, to better respect cortical anatomy over Euclidean distance; ***Glasser et al., 2013***). This yielded for each participant five classification maps, indicating the classifier's accuracy in a neighborhood surrounding every vertex. Individual subject searchlight accuracy maps were then averaged within analysis, and the resulting group-wise maps were thresholded using the PyMVPA implementation of the two-step cluster-thresholding procedure described in ***Stelzer et al., 2013*** (***Hanke et al., 2009***). This procedure permutes block labels within participant to generate a null distribution within subject (100 times) and then samples from these (10,000) to generate a group-wise null distribution (as in the fROI analysis). The whole-brain searchlight maps are then thresholded using a combination of vertex-wise threshold ($p<0.001$ uncorrected) and cluster size threshold (FWER $p<0.05$, corrected for multiple comparisons across the entire cortical surface).

Data availability statement

Custom lab software for fMRI analysis is available via GitHub (https://github.com/NPDL/NPDL-scripts copy archived at ***Lane et al., 2025***). Stimuli and code specific to this project are accessible via OSF (https://osf.io/cx9n2/). fMRI and behavioral data are accessible via OpenICPSR (10.3886/E237324V1).

Acknowledgements

We thank the F.M. Kirby Research Center for Functional Brain Imaging at the Kennedy Krieger Institute for their assistance with data collection, and the participants for making this research possible. This work was supported by a grant from the National Science Foundation (BCS-2318685 to MB).

Additional information

Funding

Funder	Grant reference number	Author
National Science Foundation	BCS-2318685	Marina Bedny

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Miriam Hauptman, Conceptualization, Data curation, Software, Formal analysis, Investigation, Visualization, Methodology, Writing – original draft, Writing – review and editing; Marina Bedny, Conceptualization, Resources, Supervision, Funding acquisition, Methodology, Writing – review and editing

Author ORCIDs

Miriam Hauptman https://orcid.org/0000-0002-5903-1552

Ethics

All participants gave written informed consent. The study was reviewed and approved by the Johns Hopkins Medicine Institutional Review Boards.

Peer review material

Reviewer #1 (Public review): https://doi.org/10.7554/eLife.101944.3.sa1
Reviewer #2 (Public review): https://doi.org/10.7554/eLife.101944.3.sa2
Reviewer #3 (Public review): https://doi.org/10.7554/eLife.101944.3.sa3
Author response https://doi.org/10.7554/eLife.101944.3.sa4

Additional files

Supplementary files

Supplementary file 1. Illness types present in the stimulus set.

Supplementary file 2. Results of preregistered MVPA for *Illness-Causal* vs. *Mechanical-Causal* in individual-subject functional ROIs (fROI). Each fROI was created by selecting the top 300 vertices for each contrast (see 'Contrast') in each search space. Accuracy refers to classifier performance against chance (50%) for *Illness-Causal* vs. *Mechanical-Causal*. Permuted and Bonferroni-corrected (across fROIs) p-values are reported. Ment_vs_phys: *mentalizing stories>physical stories* (mentalizing localizer). Caus_vs_rest: *Illness-Causal+Illness-Mechanical>Rest*. Logic_vs_lang: *logic >language* (language/logic localizer). Lang_vs_math: *language>math* (language/logic localizer). Visualizations of these results are displayed in ***Figure 1—figure supplement 11***.

Supplementary file 3. MVPA results for all tests in select individual-subject functional ROIs (fROI). Each fROI was created by selecting the top 300 vertices for each contrast in each search space: left PC (LPC)=top *main experimental conditions>rest*, language = top *language>math* (language/logic localizer), logic = top *logic >language* (language/logic localizer). Accuracy refers to classifier performance against chance (50%) for each test. Permuted and Bonferroni-corrected (across fROIs) p-values are reported. Visualizations of these results are displayed in ***Figure 1—figure supplement 12***.

MDAR checklist

Data availability

fMRI and behavioral data are publicly available via OpenICPSR (https://doi.org/10.3886/E237324V1).

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Hauptman M, Bedny M	2025	The neural basis of causal inferences about biological and physical processes	https://doi.org/10.3886/E237324V1	OpenICPSR, 10.3886/E237324V1

References

Ackerknecht EH. 1982. *A Short History of Medicine*. Johns Hopkins University Press.

Aglinskas A, Fairhall SL. 2023. Similar representation of names and faces in the network for person perception. *NeuroImage* **274**:120100. DOI: https://doi.org/10.1016/j.neuroimage.2023.120100, PMID: 37080346

Atran S. 1998. Folk biology and the anthropology of science: cognitive universals and cultural particulars. *The Behavioral and Brain Sciences* **21**:547–569. DOI: https://doi.org/10.1017/s0140525x98001277, PMID: 10097021

Baldassano C, Beck DM, Fei-Fei L. 2013. Differential connectivity within the Parahippocampal place area. *NeuroImage* **75**:228–237. DOI: https://doi.org/10.1016/j.neuroimage.2013.02.073, PMID: 23507385

Barbey AK, Patterson R. 2011. Architecture of explanatory inference in the human prefrontal cortex. *Frontiers in Psychology* **2**:162. DOI: https://doi.org/10.3389/fpsyg.2011.00162, PMID: 21845182

Barrett HC, Cosmides L, Tooby J. 2007. The hominid entry into the cognitive niche. In: Gangestad SW, Simpson JA (Eds). *The Evolution of Mind: Fundamental Questions and Controversies*. The Guilford Press. p. 241–248.

Barsalou LW, Kyle Simmons W, Barbey AK, Wilson CD. 2003. Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences* **7**:84–91. DOI: https://doi.org/10.1016/S1364-6613(02)00029-3

Bender A, Beller S, Medin DL. 2017. Causal cognition and culture. In: Waldmann M (Ed). *The Oxford Handbook of Causal Reasoning*. Oxford University Press. p. 717–738. DOI: https://doi.org/10.1093/oxfordhb/9780199399550.013.34

Bi Y, Wang X, Caramazza A. 2016. Object domain and modality in the ventral visual pathway. *Trends in Cognitive Sciences* **20**:282–290. DOI: https://doi.org/10.1016/j.tics.2016.02.002

Bi Y. 2021. Dual coding of knowledge in the human brain. *Trends in Cognitive Sciences* **25**:883–895. DOI: https://doi.org/10.1016/j.tics.2021.07.006

Binder JR, Desai RH. 2011. The neurobiology of semantic memory. *Trends in Cognitive Sciences* **15**:527–536. DOI: https://doi.org/10.1016/j.tics.2011.10.001

Black JB, Bern H. 1981. Causal coherence and memory for events in narratives. *Journal of Verbal Learning and Verbal Behavior* **20**:267–275. DOI: https://doi.org/10.1016/S0022-5371(81)90417-5

Blank IA, Fedorenko E. 2020. No evidence for differences among language regions in their temporal receptive windows. *NeuroImage* **219**:116925. DOI: https://doi.org/10.1016/j.neuroimage.2020.116925, PMID: 32407994

Boyer P. 1995. Causal understandings in cultural representations. In: Sperber D, Premack D, Premack AJ (Eds). *Causal Cognition: A Multidisciplinary Debate*. Oxford University Press. p. 615–649. DOI: https://doi.org/10.1093/acprof:oso/9780198524021.003.0020

Callanan MA, Oakes LM. 1992. Preschoolers' questions and parents' explanations: causal thinking in everyday activity. *Cognitive Development* **7**:213–233. DOI: https://doi.org/10.1016/0885-2014(92)90012-G

Caramazza A, Shelton JR. 1998. Domain-specific knowledge systems in the brain the animate-inanimate distinction. *Journal of Cognitive Neuroscience* **10**:1–34. DOI: https://doi.org/10.1162/089892998563752, PMID: 9526080
Carey S. 1985. *Conceptual Change in Childhood*. Bradford Books.
Carey S. 1988. Conceptual differences between children and adults. *Mind & Language* **3**:167–181. DOI: https://doi.org/10.1111/j.1468-0017.1988.tb00141.x
Carey S. 2011. *The Origin of Concepts*. Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780195367638.001.0001
Chang CHC, Nastase SA, Hasson U. 2022. Information flow across the cortical timescale hierarchy during narrative construction. *PNAS* **119**:e2209307119. DOI: https://doi.org/10.1073/pnas.2209307119, PMID: 36508677
Chen J, Bornstein AM. 2024. The causal structure and computational value of narratives. *Trends in Cognitive Sciences* **28**:769–781. DOI: https://doi.org/10.1016/j.tics.2024.04.003
Cheng PW, Novick LR. 1992. Covariation in natural causal induction. *Psychological Review* **99**:365–382. DOI: https://doi.org/10.1037/0033-295x.99.2.365, PMID: 1594730
Chow HM, Kaup B, Raabe M, Greenlee MW. 2008. Evidence of fronto-temporal interactions for strategic inference processes during language comprehension. *NeuroImage* **40**:940–954. DOI: https://doi.org/10.1016/j.neuroimage.2007.11.044, PMID: 18201911
Coley JD, Solomon GEA, Shafto P. 2002. The development of folkbiology: a cognitive science perspective on children's understanding of the biological world. In: Kahn P, Kellert S (Eds). *Children and Nature: Psychological, Sociocultural and Evolutionary Investigations*. MIT Press. p. 65–91. DOI: https://doi.org/10.7551/mitpress/1807.003.0004
Connolly AC, Sha L, Guntupalli JS, Oosterhof N, Halchenko YO, Nastase SA, di Oleggio Castello MV, Abdi H, Jobst BC, Gobbini MI, Haxby JV. 2016. How the human brain represents perceived dangerousness or "Predacity" of Animals. *The Journal of Neuroscience* **36**:5373–5384. DOI: https://doi.org/10.1523/JNEUROSCI.3395-15.2016, PMID: 27170133
Cosmides L, Tooby J. 2013. Unraveling the enigma of human intelligence: evolutionary psychology and the multimodular mind. In: Cosmides L (Ed). *In The Evolution of Intelligence*. Psychology Press. p. 145–198.
Coutanche MN. 2013. Distinguishing multi-voxel patterns and mean activation: why, how, and what does it tell us? *Cognitive, Affective, & Behavioral Neuroscience* **13**:667–673. DOI: https://doi.org/10.3758/s13415-013-0186-2
Dale AM, Fischl B, Sereno MI. 1999. Cortical Surface-Based Analysis. *NeuroImage* **9**:179–194. DOI: https://doi.org/10.1006/nimg.1998.0395
Davis T, LaRocque KF, Mumford JA, Norman KA, Wagner AD, Poldrack RA. 2014. What do differences between multi-voxel and univariate analysis mean? How subject-, voxel-, and trial-level variance impact fMRI analysis. *NeuroImage* **97**:271–283. DOI: https://doi.org/10.1016/j.neuroimage.2014.04.037, PMID: 24768930
Davis ZJ, Rehder B. 2020. A process model of causal reasoning. *Cognitive Science* **44**:e12839. DOI: https://doi.org/10.1111/cogs.12839, PMID: 32419205
Deen B, Freiwald WA. 2022. Parallel systems for social and spatial reasoning within the cortical apex. *bioRxiv*. DOI: https://doi.org/10.1101/2021.09.23.461550
Dehaene-Lambertz G, Monzalvo K, Dehaene S. 2018. The emergence of the visual word form: Longitudinal evolution of category-specific ventral visual areas during reading acquisition. *PLOS Biology* **16**:e2004103. DOI: https://doi.org/10.1371/journal.pbio.2004103, PMID: 29509766
DeJesus JM, Venkatesh S, Kinzler KD. 2021. Young children's ability to make predictions about novel illnesses. *Child Development* **92**:e817–e831. DOI: https://doi.org/10.1111/cdev.13655, PMID: 34463345
Devlin JT, Russell RP, Davis MH, Price CJ, Moss HE, Fadili MJ, Tyler LK. 2002. Is there an anatomical basis for category-specificity? Semantic memory studies in PET and fMRI. *Neuropsychologia* **40**:54–75. DOI: https://doi.org/10.1016/s0028-3932(01)00066-5, PMID: 11595262
DiNicola LM, Braga RM, Buckner RL. 2020. Parallel distributed networks dissociate episodic and social functions within the individual. *Journal of Neurophysiology* **123**:1144–1179. DOI: https://doi.org/10.1152/jn.00529.2019, PMID: 32049593
Dodell-Feder D, Koster-Hale J, Bedny M, Saxe R. 2011. fMRI item analysis in a theory of mind task. *NeuroImage* **55**:705–712. DOI: https://doi.org/10.1016/j.neuroimage.2010.12.040, PMID: 21182967
Duffy SA, Shinjo M, Myers JL. 1990. The effect of encoding task on memory for sentence pairs varying in causal relatedness. *Journal of Memory and Language* **29**:27–42. DOI: https://doi.org/10.1016/0749-596X(90)90008-N
Dufour N, Redcay E, Young L, Mavros PL, Moran JM, Triantafyllou C, Gabrieli JDE, Saxe R. 2013. Similar brain activation during false belief tasks in a large sample of adults with and without autism. *PLOS ONE* **8**:e75468. DOI: https://doi.org/10.1371/journal.pone.0075468
Eklund A, Nichols TE, Knutsson H. 2016. Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *PNAS* **113**:7900–7905. DOI: https://doi.org/10.1073/pnas.1602413113
Eklund A, Knutsson H, Nichols TE. 2019. Cluster failure revisited: Impact of first level design and physiological noise on cluster false positive rates. *Human Brain Mapping* **40**:2017–2032. DOI: https://doi.org/10.1002/hbm.24350, PMID: 30318709
Epstein R, Kanwisher N. 1998. A cortical representation of the local visual environment. *Nature* **392**:598–601. DOI: https://doi.org/10.1038/33402, PMID: 9560155
Fairhall SL, Caramazza A. 2013a. Brain regions that represent amodal conceptual knowledge. *The Journal of Neuroscience* **33**:10552–10558. DOI: https://doi.org/10.1523/JNEUROSCI.0051-13.2013, PMID: 23785167

Fairhall SL, Caramazza A. 2013b. Category-selective neural substrates for person- and place-related concepts. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior* **49**:2748–2757. DOI: https://doi.org/10.1016/j.cortex.2013.05.010, PMID: 23831433

Fairhall SL, Anzellotti S, Ubaldi S, Caramazza A. 2014. Person- and place-selective neural substrates for entity-specific semantic access. *Cerebral Cortex* **24**:1687–1696. DOI: https://doi.org/10.1093/cercor/bht039, PMID: 23425892

Farah MJ, Rabinowitz C. 2003. Genetic and environmental influences on the organisation of semantic memory in the brain:is "living things" an innate category? *Cognitive Neuropsychology* **20**:401–408. DOI: https://doi.org/10.1080/02643290244000293, PMID: 20957577

Fedorenko E, Hsieh PJ, Nieto-Castañón A, Whitfield-Gabrieli S, Kanwisher N. 2010. New method for fMRI investigations of language: defining ROIs functionally in individual subjects. *Journal of Neurophysiology* **104**:1177–1194. DOI: https://doi.org/10.1152/jn.00032.2010, PMID: 20410363

Fedorenko E, Varley R. 2016. Language and thought are not the same thing: evidence from neuroimaging and neurological patients. *Annals of the New York Academy of Sciences* **1369**:132–153. DOI: https://doi.org/10.1111/nyas.13046, PMID: 27096882

Fenker DB, Schoenfeld MA, Waldmann MR, Schuetze H, Heinze HJ, Duezel E. 2010. "Virus and epidemic": causal knowledge activates prediction error circuitry. *Journal of Cognitive Neuroscience* **22**:2151–2163. DOI: https://doi.org/10.1162/jocn.2009.21387, PMID: 19925196

Ferstl EC, von Cramon DY. 2001. The role of coherence and cohesion in text comprehension: an event-related fMRI study. *Brain Research. Cognitive Brain Research* **11**:325–340. DOI: https://doi.org/10.1016/s0926-6410(01)00007-6, PMID: 11339984

Foster GM. 1976. Disease etiologies in non-western medical systems. *American Anthropologist* **78**:773–782. DOI: https://doi.org/10.1525/aa.1976.78.4.02a00030

Fugelsang JA, Dunbar KN. 2005. Brain-based mechanisms underlying complex causal thinking. *Neuropsychologia* **43**:1204–1213. DOI: https://doi.org/10.1016/j.neuropsychologia.2004.10.012, PMID: 15817178

Gerstenberg T, Tenenbaum JB. 2017. Intuitive theories. In: Waldmann M (Ed). *The Oxford Handbook of Causal Reasoning*. Oxford University Press.

Glasser MF, Sotiropoulos SN, Wilson JA, Coalson TS, Fischl B, Andersson JL, Xu J, Jbabdi S, Webster M, Polimeni JR, Van Essen DC, Jenkinson M, WU-Minn HCP Consortium. 2013. The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage* **80**:105–124. DOI: https://doi.org/10.1016/j.neuroimage.2013.04.127, PMID: 23668970

Goddu MK, Gopnik A. 2024. The development of human causal learning and reasoning. *Nature Reviews Psychology* **3**:319–339. DOI: https://doi.org/10.1038/s44159-024-00300-5

Goldvarg E, Johnson-Laird PN. 2001. Naive causality: a mental model theory of causal meaning and reasoning. *Cognitive Science* **25**:565–610. DOI: https://doi.org/10.1207/s15516709cog2504_3

Gopnik A, Meltzoff AN. 1997. *Words, Thoughts, and Theories*. The MIT Press.

Gopnik A, Sobel DM, Schulz LE, Glymour C. 2001. Causal learning mechanisms in very young children: two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology* **37**:620–629. DOI: https://doi.org/10.1037/0012-1649.37.5.620, PMID: 11552758

Gopnik A, Glymour C, Sobel DM, Schulz LE, Kushnir T, Danks D. 2004. A theory of causal learning in children: causal maps and Bayes nets. *Psychological Review* **111**:3–32. DOI: https://doi.org/10.1037/0033-295X.111.1.3, PMID: 14756583

Gopnik A, Wellman HM. 2012. Reconstructing constructivism: causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin* **138**:1085–1108. DOI: https://doi.org/10.1037/a0028044, PMID: 22582739

Graesser AC, Singer M, Trabasso T. 1994. Constructing inferences during narrative text comprehension. *Psychological Review* **101**:371–395. DOI: https://doi.org/10.1037/0033-295x.101.3.371, PMID: 7938337

Grill-Spector K, Knouf N, Kanwisher N. 2004. The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience* **7**:555–562. DOI: https://doi.org/10.1038/nn1224, PMID: 15077112

Gutheil G, Vera A, Keil FC. 1998. Do houseflies think? Patterns of induction and biological beliefs in development. *Cognition* **66**:33–49. DOI: https://doi.org/10.1016/s0010-0277(97)00049-8, PMID: 9675977

Hanke M, Halchenko YO, Sederberg PB, Hanson SJ, Haxby JV, Pollmann S. 2009. PyMVPA: A python toolbox for multivariate pattern analysis of fMRI data. *Neuroinformatics* **7**:37–53. DOI: https://doi.org/10.1007/s12021-008-9041-y, PMID: 19184561

Hasson U, Yang E, Vallines I, Heeger DJ, Rubin N. 2008. A hierarchy of temporal receptive windows in human cortex. *The Journal of Neuroscience* **28**:2539–2550. DOI: https://doi.org/10.1523/JNEUROSCI.5487-07.2008, PMID: 18322098

Hatano G, Inagaki K. 1994. Young children's naive theory of biology. *Cognition* **50**:171–188. DOI: https://doi.org/10.1016/0010-0277(94)90027-2, PMID: 8039360

Hauptman M, Elli G, Pant R, Bedny M. 2025. Neural specialization for "visual" concepts emerges in the absence of vision. *Cognition* **257**:106058. DOI: https://doi.org/10.1016/j.cognition.2024.106058, PMID: 39827755

Häusler CO, Eickhoff SB, Hanke M. 2022. Processing of visual and non-visual naturalistic spatial information in the "parahippocampal place area". *Scientific Data* **9**:147. DOI: https://doi.org/10.1038/s41597-022-01250-4, PMID: 35365659

Hebart MN, Baker CI. 2018. Deconstructing multivariate decoding for the study of brain function. *NeuroImage* **180**:4–18. DOI: https://doi.org/10.1016/j.neuroimage.2017.08.005, PMID: 28782682

Hickling AK, Wellman HM. 2001. The emergence of children's causal explanations and theories: evidence from everyday conversation. *Developmental Psychology* **37**:668–683. DOI: https://doi.org/10.1037//0012-1649.37.5.668, PMID: 11552762

Hillis AE, Caramazza A. 1991. Category-specific naming and comprehension impairment: a double dissociation. *Brain* **114 (Pt 5)**:2081–2094. DOI: https://doi.org/10.1093/brain/114.5.2081, PMID: 1933235

Inagaki K, Hatano G. 1993. Young children's understanding of the mind-body distinction. *Child Development* **64**:1534–1549 PMID: 8222887.

Inagaki K, Hatano G. 2004. Vitalistic causality in young children's naive biology. *Trends in Cognitive Sciences* **8**:356–362. DOI: https://doi.org/10.1016/j.tics.2004.06.004

Inagaki K, Hatano G. 2006. Young children's conception of the biological world. *Current Directions in Psychological Science* **15**:177–181. DOI: https://doi.org/10.1111/j.1467-8721.2006.00431.x

Jacoby N, Fedorenko E. 2020. Discourse-level comprehension engages medial frontal Theory of Mind brain regions even for expository texts. *Language, Cognition and Neuroscience* **35**:780–796. DOI: https://doi.org/10.1080/23273798.2018.1525494, PMID: 32984430

Julian JB, Fedorenko E, Webster J, Kanwisher N. 2012. An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *NeuroImage* **60**:2357–2364. DOI: https://doi.org/10.1016/j.neuroimage.2012.02.055, PMID: 22398396

Kalish CW. 1996. Preschoolers' understanding of germs as invisible mechanisms. *Cognitive Development* **11**:83–106. DOI: https://doi.org/10.1016/S0885-2014(96)90029-5

Kalish C. 1997. Preschoolers' understanding of mental and bodily reactions to contamination: what you don't know can hurt you, but cannot sadden you. *Developmental Psychology* **33**:79–91. DOI: https://doi.org/10.1037//0012-1649.33.1.79, PMID: 9050393

Kanjlia S, Lane C, Feigenson L, Bedny M. 2016. Absence of visual experience modifies the neural basis of numerical thinking. *PNAS* **113**:11172–11177. DOI: https://doi.org/10.1073/pnas.1524982113

Kanwisher N, McDermott J, Chun MM. 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience* **17**:4302–4311. DOI: https://doi.org/10.1523/JNEUROSCI.17-11-04302.1997, PMID: 9151747

Keenan JM, Baillet SD, Brown P. 1984. The effects of causal cohesion on comprehension and memory. *Journal of Verbal Learning and Verbal Behavior* **23**:115–126. DOI: https://doi.org/10.1016/S0022-5371(84)90082-3

Keil FC. 1992. The origins of an autonomous biology. In: Gunnar MR, Maratsos M (Eds). *Modularity and Constraints in Language and Cognition*. Psychology Press. p. 103–137.

Keil FC. 1994. The birth and nurturance of concepts by domains: the origins of concepts of living things. In: Hirschfeld LA, Gelman SA (Eds). *Mapping the Mind: Domain Specificity in Cognition and Culture*. Cambridge University Press. p. 234–254. DOI: https://doi.org/10.1017/CBO9780511752902.010

Keil FC, Levin DT, Richman BA, Gutheil G. 1999. Mechanism and explanation in the development of biological thought: the case of disease. In: Medin DL, Atran S (Eds). *Folkbiology*. MIT Press. p. 285–320. DOI: https://doi.org/10.7551/mitpress/3042.003.0010

Khemlani SS, Barbey AK, Johnson-Laird PN. 2014. Causal reasoning with mental models. *Frontiers in Human Neuroscience* **8**:849. DOI: https://doi.org/10.3389/fnhum.2014.00849, PMID: 25389398

Konkle T, Caramazza A. 2013. Tripartite organization of the ventral stream by animacy and object size. *The Journal of Neuroscience* **33**:10235–10242. DOI: https://doi.org/10.1523/JNEUROSCI.0983-13.2013, PMID: 23785139

Kragel PA, Carter RM, Huettel SA. 2012. What makes a pattern? Matching decoding methods to data in multivariate pattern analysis. *Frontiers in Neuroscience* **6**:162. DOI: https://doi.org/10.3389/fnins.2012.00162, PMID: 23189035

Kranjec A, Cardillo ER, Schmidt GL, Lehet M, Chatterjee A. 2012. Deconstructing events: the neural bases for space, time, and causality. *Journal of Cognitive Neuroscience* **24**:1–16. DOI: https://doi.org/10.1162/jocn_a_00124, PMID: 21861674

Kuperberg GR, Lakshmanan BM, Caplan DN, Holcomb PJ. 2006. Making sense of discourse: an fMRI study of causal inferencing across sentences. *NeuroImage* **33**:343–361. DOI: https://doi.org/10.1016/j.neuroimage.2006.06.001, PMID: 16876436

Lagnado DA, Waldmann MR, Hagmaye Y, Sloman SA. 2007. Beyond covariation: cues to causal structure. In: Gopnik A, Schulz L (Eds). *Causal Learning*. Oxford University Press. p. 154–172. DOI: https://doi.org/10.1093/acprof:oso/9780195176803.003.0011

Lane C, Kim J, Kanjlia S. 2025. NPDL-scripts. swh:1:rev:d5a07b43f145951c35ca48632657fd79ebdab61a. Software Heritage. https://archive.softwareheritage.org/swh:1:dir:7428c8f96b5af84bf023699d7ac865b9a200e4e4;origin=https://github.com/NPDL/NPDL-scripts;visit=swh:1:snp:1e3cd75e38372361aa6a8d82c0c4b32af8a2f7d6;anchor=swh:1:rev:d5a07b43f145951c35ca48632657fd79ebdab61a

Lee H, Chen J. 2022. Predicting memory from the network structure of naturalistic events. *Nature Communications* **13**:4235. DOI: https://doi.org/10.1038/s41467-022-31965-2

Legare CH, Gelman SA. 2008. Bewitchment, biology, or both: the co-existence of natural and supernatural explanatory frameworks across development. *Cognitive Science* **32**:607–642. DOI: https://doi.org/10.1080/03640210802066766, PMID: 21635349

Legare CH, Wellman HM, Gelman SA. 2009. Evidence for an explanation advantage in naïve biological reasoning. *Cognitive Psychology* **58**:177–194. DOI: https://doi.org/10.1016/j.cogpsych.2008.06.002, PMID: 18710700

Legare CH, Evans EM, Rosengren KS, Harris PL. 2012. The coexistence of natural and supernatural explanations across cultures and development. *Child Development* **83**:779–793. DOI: https://doi.org/10.1111/j.1467-8624.2012.01743.x, PMID: 22417318

Legare C, Shtulman A. 2017. Explanatory pluralism across cultures and development. In: Legare C, Shtulman A (Eds). *In Metacognitive Diversity: An Interdisciplinary Approach*. Oxford University Press. p. 415–432. DOI: https://doi.org/10.1093/Oso/9780198789710.003.0019

Lerner Y, Honey CJ, Silbert LJ, Hasson U. 2011. Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *The Journal of Neuroscience* **31**:2906–2915. DOI: https://doi.org/10.1523/JNEUROSCI.3684-10.2011

Lightner AD, Heckelsmiller C, Hagen EH. 2021. Ethnoscientific expertise and knowledge specialisation in 55 traditional cultures. *Evolutionary Human Sciences* **3**:e37. DOI: https://doi.org/10.1017/ehs.2021.31, PMID: 37588549

Liu YF, Kim J, Wilson C, Bedny M. 2020. Computer code comprehension shares neural resources with formal logical inference in the fronto-parietal network. *eLife* **9**:e59340. DOI: https://doi.org/10.7554/eLife.59340, PMID: 33319745

Lynch E, Medin D. 2006. Explanatory models of illness: a study of within-culture variation. *Cognitive Psychology* **53**:285–309. DOI: https://doi.org/10.1016/j.cogpsych.2006.02.001, PMID: 16624275

Mahon BZ, Anzellotti S, Schwarzbach J, Zampini M, Caramazza A. 2009. Category-specific organization in the human brain does not require visual experience. *Neuron* **63**:397–405. DOI: https://doi.org/10.1016/j.neuron.2009.07.012, PMID: 19679078

Marneffe MC, MacCartney B, Manning C. 2006. Generating typed dependency parses from phrase structure parses. Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06) Genoa.

Martin A, Chao LL. 2001. Semantic memory and the brain: structure and processes. *Current Opinion in Neurobiology* **11**:194–201. DOI: https://doi.org/10.1016/s0959-4388(00)00196-3, PMID: 11301239

Mason RA, Just MA. 2011. Differentiable cortical networks for inferences concerning people's intentions versus physical causality. *Human Brain Mapping* **32**:313–329. DOI: https://doi.org/10.1002/hbm.21021, PMID: 21229617

Meder B, Mayrhofer R. 2017. Diagnostic reasoning. In: Waldmann M (Ed). *The Oxford Handbook of Causal Reasoning*. Oxford University Press. p. 433–458. DOI: https://doi.org/10.1093/oxfordhb/9780199399550.013.23

Medin DL, Atran S. 2004. The native mind: biological categorization and reasoning in development and across cultures. *Psychological Review* **111**:960–983. DOI: https://doi.org/10.1037/0033-295X.111.4.960, PMID: 15482069

Medin D, Waxman S, Woodring J, Washinawatok K. 2010. Human-centeredness is not a universal feature of young children's reasoning: culture and experience matter when reasoning about biological entities. *Cognitive Development* **25**:197–207. DOI: https://doi.org/10.1016/j.cogdev.2010.02.001, PMID: 20824197

Monti MM, Parsons LM, Osherson DN. 2009. The boundaries of language and thought in deductive inference. *PNAS* **106**:12554–12559. DOI: https://doi.org/10.1073/pnas.0902422106

Muentener P, Schulz L. 2014. Toddlers infer unobserved causes for spontaneous events. *Frontiers in Psychology* **5**:1496. DOI: https://doi.org/10.3389/fpsyg.2014.01496, PMID: 25566161

Myers JL, Shinjo M, Duffy SA. 1987. Degree of causal relatedness and memory. *Journal of Memory and Language* **26**:453–465. DOI: https://doi.org/10.1016/0749-596X(87)90101-X

Noppeney U, Price CJ, Penny WD, Friston KJ. 2006. Two distinct neural mechanisms for category-selective responses. *Cerebral Cortex* **16**:437–445. DOI: https://doi.org/10.1093/cercor/bhi123, PMID: 15944370

Norman GR, Grierson LEM, Sherbino J, Hamstra SJ, Schmidt HG, Mamede S. 2009. Expertise in medicine and surgery. In: Ericsson KA, Hoffman RR, Kozbelt A, Williams AM (Eds). *The Cambridge Handbook of Expertise and Expert Performance*. Cambridge University Press. p. 339–354. DOI: https://doi.org/10.1017/CBO9780511816796.019

Notaro PC, Gelman SA, Zimmerman MA. 2001. Children's understanding of psychogenic bodily reactions. *Child Development* **72**:444–459. DOI: https://doi.org/10.1111/1467-8624.00289, PMID: 11333077

Operskalski JT, Barbey AK. 2017. Cognitive neuroscience of causal reasoning. In: Waldmann M (Ed). *The Oxford Handbook of Causal Reasoning*. Oxford University Press. p. 217–242. DOI: https://doi.org/10.1093/oxfordhb/9780199399550.013.16

Opfer JE, Gelman SA. 2011. Development of the animate-inanimate distinction. In: Goswami U (Ed). *The Wiley-Blackwell Handbook of Childhood Cognitive Development*. Wiley-Blackwell. p. 213–238. DOI: https://doi.org/10.1002/9781444325485

Pakravan M, Abbaszadeh M, Ghazizadeh A. 2022. Coordinated multivoxel coding beyond univariate effects is not likely to be observable in fMRI data. *NeuroImage* **247**:118825. DOI: https://doi.org/10.1016/j.neuroimage.2021.118825, PMID: 34942362

Pallier C, Devauchelle AD, Dehaene S. 2011. Cortical representation of the constituent structure of sentences. *PNAS* **108**:2522–2527. DOI: https://doi.org/10.1073/pnas.1018711108

Pearl J. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge University Press.

Peer M, Salomon R, Goldberg I, Blanke O, Arzy S. 2015. Brain system for mental orientation in space, time, and person. *PNAS* **112**:11072–11077. DOI: https://doi.org/10.1073/pnas.1504242112
Peirce J, Gray JR, Simpson S, MacAskill M, Höchenberger R, Sogo H, Kastman E, Lindeløv JK. 2019. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods* **51**:195–203. DOI: https://doi.org/10.3758/s13428-018-01193-y, PMID: 30734206
Pinker S. 2003. Language as an adaptation to the cognitive niche. In: Christiansen MH, Kirby S (Eds). *Language Evolution*. Oxford University Press. p. 16–37. DOI: https://doi.org/10.1093/acprof:oso/9780199244843.003.0002
Pramod RT, Chomik J, Schulz L, Kanwisher N. 2023. A region in human left prefrontal cortex selectively engaged in causal reasoning. In Proceedings of the Annual Meeting of the Cognitive Science Society.
Prat CS, Mason RA, Just MA. 2011. Individual differences in the neural basis of causal inferencing. *Brain and Language* **116**:1–13. DOI: https://doi.org/10.1016/j.bandl.2010.08.004, PMID: 21051084
Rabini G, Ubaldi S, Fairhall SL. 2021. Combining concepts across categorical domains: a linking role of the precuneus. *Neurobiology of Language* **2**:354–371. DOI: https://doi.org/10.1162/nol_a_00039, PMID: 34595480
Raman L, Winer GA. 2004. Evidence of more immanent justice responding in adults than children: A challenge to traditional developmental theories. *British Journal of Developmental Psychology* **22**:255–274. DOI: https://doi.org/10.1348/026151004323044609
Raman L, Gelman SA. 2005. Children's understanding of the transmission of genetic disorders and contagious illnesses. *Developmental Psychology* **41**:171–182. DOI: https://doi.org/10.1037/0012-1649.41.1.171, PMID: 15656747
Rehder B, Burnett RC. 2005. Feature inference and the causal structure of categories. *Cognitive Psychology* **50**:264–314. DOI: https://doi.org/10.1016/j.cogpsych.2004.09.002, PMID: 15826612
Ritchey M, Cooper RA. 2020. Deconstructing the posterior medial episodic network. *Trends in Cognitive Sciences* **24**:451–465. DOI: https://doi.org/10.1016/j.tics.2020.03.006
Rosengren KS, Gelman SA, Kalish CW, McCormick M. 1991. As time goes by: children's early understanding of growth in animals. *Child Development* **62**:1302–1320. DOI: https://doi.org/10.2307/1130808, PMID: 1786717
Rottman BM, Ahn WK, Luhmann CC. 2011. When and how do people reason about unobserved causes. In: Illari P, Russo F, Williamson J (Eds). *Causality in the Sciences*. Oxford University Press. p. 150–183. DOI: https://doi.org/10.1093/acprof:oso/9780199574131.003.0008
Rottman BM, Hastie R. 2014. Reasoning about causal relationships: Inferences on causal networks. *Psychological Bulletin* **140**:109–139. DOI: https://doi.org/10.1037/a0031903, PMID: 23544658
Satpute AB, Fenker DB, Waldmann MR, Tabibnia G, Holyoak KJ, Lieberman MD. 2005. An fMRI study of causal judgments. *The European Journal of Neuroscience* **22**:1233–1238. DOI: https://doi.org/10.1111/j.1460-9568.2005.04292.x, PMID: 16176366
Saxe R, Kanwisher N. 2003. People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind". *NeuroImage* **19**:1835–1842. DOI: https://doi.org/10.1016/s1053-8119(03)00230-1, PMID: 12948738
Saxe R, Carey S. 2006. The perception of causality in infancy. *Acta Psychologica* **123**:144–165. DOI: https://doi.org/10.1016/j.actpsy.2006.05.005
Saxe R, Moran JM, Scholz J, Gabrieli J. 2006. Overlapping and non-overlapping brain regions for theory of mind and self reflection in individual subjects. *Social Cognitive and Affective Neuroscience* **1**:229–234. DOI: https://doi.org/10.1093/scan/nsl034, PMID: 18985110
Saxe R, Powell LJ. 2006. It's the thought that counts: specific brain regions for one component of theory of mind. *Psychological Science* **17**:692–699. DOI: https://doi.org/10.1111/j.1467-9280.2006.01768.x, PMID: 16913952
Schmidt HG, Norman GR, Boshuizen HP. 1990. A cognitive perspective on medical expertise: theory and implication. *Academic Medicine* **65**:611–621. DOI: https://doi.org/10.1097/00001888-199010000-00001, PMID: 2261032
Schreiber K, Krekelberg B. 2013. The statistical analysis of multi-voxel patterns in functional imaging. *PLOS ONE* **8**:e69328. DOI: https://doi.org/10.1371/journal.pone.0069328, PMID: 23861966
Schulz LE, Gopnik A. 2004. Causal learning across domains. *Developmental Psychology* **40**:162–176. DOI: https://doi.org/10.1037/0012-1649.40.2.162, PMID: 14979758
Shain C, Blank IA, van Schijndel M, Schuler W, Fedorenko E. 2020. fMRI reveals language-specific predictive coding during naturalistic sentence comprehension. *Neuropsychologia* **138**:107307. DOI: https://doi.org/10.1016/j.neuropsychologia.2019.107307, PMID: 31874149
Shain C, Paunov A, Chen X, Lipkin B, Fedorenko E. 2023. No evidence of theory of mind reasoning in the human language network. *Cerebral Cortex* **33**:6299–6319. DOI: https://doi.org/10.1093/cercor/bhac505, PMID: 36585774
Silson EH, Steel A, Kidder A, Gilmore AW, Baker CI. 2019. Distinct subdivisions of human medial parietal cortex support recollection of people and places. *eLife* **8**:e47391. DOI: https://doi.org/10.7554/eLife.47391, PMID: 31305238
Simons DJ, Keil FC. 1995. An abstract to concrete shift in the development of biological thought: the insides story. *Cognition* **56**:129–163. DOI: https://doi.org/10.1016/0010-0277(94)00660-d, PMID: 7554792
Simony E, Honey CJ, Chen J, Lositsky O, Yeshurun Y, Wiesel A, Hasson U. 2016. Dynamic reconfiguration of the default mode network during narrative comprehension. *Nature Communications* **7**:12141. DOI: https://doi.org/10.1038/ncomms12141, PMID: 27424918
Singer M. 1994. Discourse inference processes. In: Gernsbacher MA (Ed). *Handbook of Psycholinguistics*. Academic Press. p. 479–515.

Sloman SA, Lagnado D. 2015. Causality in thought. *Annual Review of Psychology* **66**:223–247. DOI: https://doi.org/10.1146/annurev-psych-010814-015135, PMID: 25061673

Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE, Niazy RK, Saunders J, Vickers J, Zhang Y, De Stefano N, Brady JM, Matthews PM. 2004. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* **23 Suppl 1**:S208–S19. DOI: https://doi.org/10.1016/j.neuroimage.2004.07.051, PMID: 15501092

Spelke ES. 2003. What makes us smart? core knowledge and natural language. In: Gentner D, Goldin-Meadow S (Eds). *Language in Mind*. MIT Press. p. 277–311. DOI: https://doi.org/10.7551/mitpress/4117.003.0017

Spelke ES. 2022. *What Babies Know: Core Knowledge and Composition*. Oxford University Press.

Springer K, Keil FC. 1991. Early differentiation of causal mechanisms appropriate to biological and nonbiological kinds. *Child Development* **62**:767–781 PMID: 1935342.

Springer K, Ruckel J. 1992. Early beliefs about the cause of illness: evidence against immanent justice. *Cognitive Development* **7**:429–443. DOI: https://doi.org/10.1016/0885-2014(92)80002-W

Steel A, Billings MM, Silson EH, Robertson CE. 2021. A network linking scene perception and spatial memory systems in posterior cerebral cortex. *Nature Communications* **12**:2632. DOI: https://doi.org/10.1038/s41467-021-22848-z, PMID: 33976141

Stelzer J, Chen Y, Turner R. 2013. Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): random permutations and cluster size control. *NeuroImage* **65**:69–82. DOI: https://doi.org/10.1016/j.neuroimage.2012.09.063, PMID: 23041526

Steyvers M, Tenenbaum JB, Wagenmakers EJ, Blum B. 2003. Inferring causal networks from observations and interventions. *Cognitive Science* **27**:453–489. DOI: https://doi.org/10.1207/s15516709cog2703_6

Tenenbaum JB, Griffiths TL, Niyogi S. 2007. Intuitive theories as grammars for causal inference. In: Gopnik A, Schulz L (Eds). *Causal Learning*. Oxford University Press. p. 301–322. DOI: https://doi.org/10.1093/acprof:oso/9780195176803.003.0020

Thompson-Schill SL. 2003. Neuroimaging studies of semantic memory: inferring "how" from "where". *Neuropsychologia* **41**:280–292. DOI: https://doi.org/10.1016/s0028-3932(02)00161-6, PMID: 12457754

Tooby J, DeVore I. 1987. The reconstruction of hominid behavioral evolution through strategic modeling. In: Kinzey WG (Ed). *The Evolution of Human Behavior: Primate Models*. SUNY Press.

Trabasso T, Sperry LL. 1985. Causal relatedness and importance of story events. *Journal of Memory and Language* **24**:595–611. DOI: https://doi.org/10.1016/0749-596X(85)90048-8

Varley R, Siegal M. 2000. Evidence for cognition without grammar from causal reasoning and "theory of mind" in an agrammatic aphasic patient. *Current Biology* **10**:723–726. DOI: https://doi.org/10.1016/s0960-9822(00)00538-8, PMID: 10873809

Varley R. 2014. Reason without much language. *Language Sciences* **46**:232–244. DOI: https://doi.org/10.1016/j.langsci.2014.06.012

Vul E, Kanwisher N. 2011. Begging the question: the non-independence error in fmri. In: Hanson SJ, Bunzl M (Eds). *Foundational Issues in Human Brain Mapping*. MIT Press. DOI: https://doi.org/10.7551/mitpress/9780262014021.003.0007

Waldmann MR, Holyoak KJ. 1992. Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology* **121**:222–236. DOI: https://doi.org/10.1037//0096-3445.121.2.222

Wang X, Peelen MV, Han Z, Caramazza A, Bi Y. 2016. The role of vision in the neural representation of unique entities. *Neuropsychologia* **87**:144–156. DOI: https://doi.org/10.1016/j.neuropsychologia.2016.05.007, PMID: 27174518

Warrington EK, Shallice T. 1984. Category specific semantic impairments. *Brain* **107 (Pt 3)**:829–854. DOI: https://doi.org/10.1093/brain/107.3.829, PMID: 6206910

Weiner KS, Barnett MA, Witthoft N, Golarai G, Stigliani A, Kay KN, Gomez J, Natu VS, Amunts K, Zilles K, Grill-Spector K. 2018. Defining the most probable location of the parahippocampal place area using cortex-based alignment and cross-validation. *NeuroImage* **170**:373–384. DOI: https://doi.org/10.1016/j.neuroimage.2017.04.040, PMID: 28435097

Wellman HM, Gelman SA. 1992. Cognitive development: foundational theories of core domains. *Annual Review of Psychology* **43**:337–375. DOI: https://doi.org/10.1146/annurev.ps.43.020192.002005, PMID: 1539946

Willems RM, Frank SL, Nijhof AD, Hagoort P, van den Bosch A. 2016. Prediction during natural language comprehension. *Cerebral Cortex* **26**:2506–2516. DOI: https://doi.org/10.1093/cercor/bhv075, PMID: 25903464

Winkler AM, Ridgway GR, Webster MA, Smith SM, Nichols TE. 2014. Permutation inference for the general linear model. *NeuroImage* **92**:381–397. DOI: https://doi.org/10.1016/j.neuroimage.2014.01.060, PMID: 24530839

Woolgar A, Golland P, Bode S. 2014. Coping with confounds in multivoxel pattern analysis: what should we do about reaction time differences? A comment on Todd, Nystrom & Cohen 2013. *NeuroImage* **98**:506–512. DOI: https://doi.org/10.1016/j.neuroimage.2014.04.059, PMID: 24793832

Yeshurun Y, Nguyen M, Hasson U. 2017. Amplification of local changes along the timescale processing hierarchy. *PNAS* **114**:9475–9480. DOI: https://doi.org/10.1073/pnas.1701652114, PMID: 28811367

Appendix 1

Online experiment protocol

Prior to the fMRI experiment, we collected explicit causality judgments from a separate group of online participants (n=30). Each online participant read all vignettes from the causal inference experiment (152 vignettes) in addition to 12 filler vignettes that were designed to be either maximally causally related or unrelated (164 vignettes total), one vignette at a time. Their task was to judge the extent to which it was possible that the event described in the first sentence of each vignette caused the event described in the second sentence on a four-point scale (1=not possible; 4=very possible). Four participants were excluded on the basis of inaccurate responses on the filler trials (i.e. difference between average ratings for maximally causally related and maximally causally unrelated vignettes <2). Among the 26 remaining participants, 12 read vignettes from Version A and 14 read vignettes from Version B of the experiment. To eliminate erroneous responses, we first excluded trials with RTs 2.5 SD outside their respective condition means within participants and then excluded trials with outlier RTs (more than 1.5 IQR below Q1 or more than 1.5 IQR above Q3) across participants (approximately 5% of all trials excluded in total). We found that both causal conditions (*Illness-Causal*, *Mechanical-Causal*) were more causally connected than both noncausal conditions, $t(25) = 36.97$, $p<0.001$ (causal: M=3.51 ± 0.78 SD, noncausal: M=1.10 ± 0.45 SD). In addition, *Illness-Causal* and *Mechanical-Causal* items received equally high causality ratings, $t(25) = -0.64$, $p=0.53$ (*Illness-Causal:* M=3.49 ± 0.77 SD, *Mechanical-Causal:* M=3.53 ± 0.79 SD).

Appendix 2

Details on measuring linguistic variables

All conditions were matched (pairwise t-tests, all ps>0.3, no statistical correction) on multiple linguistic variables known to modulate neural activity in language regions (e.g. ***Pallier et al., 2011***; ***Shain et al., 2020***). These included number of characters, number of words, average number of characters per word, average word frequency, average bigram surprisal (Google Books Ngram Viewer, https://books.google.com/ngrams/), and average syntactic dependency length (Stanford Parser; ***Marneffe et al., 2006***). Sentences that were incorrectly parsed by the automatic syntactic parser (i.e. past participle adjectives parsed as verbs) were corrected by hand. Word frequency was calculated as the negative log of a word's occurrence rate in the Google corpus between the years 2017 the 2019. Bigram surprisal was calculated as the negative log of the frequency of a given two-word phrase in the Google corpus divided by the frequency of the first word of the phrase.

This calculation uses a log base of 2 in order to express surprisal in terms of 'bits' that the first word provides in the context of the phrase. We used *bigram* surprisal as our surprisal measure to maximize the number of n-grams that had an entry in the corpus. Even so, 64 out of the 1515 total bigrams (4%) did not have an entry in the corpus and were therefore assigned the highest surprisal value among the rest of the bigrams (see ***Willems et al., 2016***).

Appendix 3

Full behavioral results

Accuracy on the magic detection task was at ceiling (*M*=97.9% ± 2.2 SD). There were no significant differences across the four main experimental conditions (*Illness-Causal*, *Mechanical-Causal*, *Noncausal-Illness First, Noncausal-Mechanical First*), but participants were more accurate on *Illness-Causal* trials compared to 'magical' catch trials ($F_{(4,76)}$ = 2.81, p=0.03; *Illness-Causal:* *M*=98.8% ± 2.2 SD; 'magical' catch trials*:* *M*=96.4% ± 3.8 SD).

A one-way repeated-measures ANOVA evaluating response time revealed a main effect of condition, $F_{(4,76)}$ = 8.17, p<0.001, whereby participants were faster on *Illness-Causal* trials (*M*=4.73 ± 0.81 SD) compared to *Noncausal-Illness First* (*M*=5.33 s±0.85 SD), *Noncausal-Mechanical First* (*M*=5.27 s±0.89 SD) trials, and 'magical' catch trials (*M*=5.34 s±0.89 SD). There were no differences in response time between *Mechanical-Causal* (*M*=5.15 s±0.88 SD) and any other conditions.

Accuracy on the language/logic localizer task was significantly lower for the logic task compared to both the language and math tasks (logic: *M*=67.5% ± 14.0 SD, math: *M*=93.8% ± 6.4 SD, language: *M*=98.1% ± 5.8 SD; $F_{(2,38)}$ = 60.38, p<0.0001). Similarly, response time was slowest on the logic task, followed by math and then language (logic: *M*=8.78 s±1.88 SD, math: *M*=6.20 s±1.37 SD, language: *M*=5.18 s±1.53 SD; $F_{(2,38)}$ = 44.28, p<0.001).

Accuracy on the mentalizing localizer task was not different across the mentalizing stories and physical stories conditions (mentalizing: 83.50%±15.7 SD, physical: 90.50%±12.3 SD; $F_{(1,19)}$ = 2.73, p=0.12). However, response time for the mentalizing stories was significantly slower (mentalizing: 3.46 s±0.55 SD, physical: 3.11 s±0.56 SD; $F_{(1,19)}$ = 16.59, p<0.001).

Appendix 4

Individual-subject univariate fROI analysis in the (FFA)

In an exploratory analysis, we defined individual-subject fROIs in the FFA.

Illness inference fROIs were created in left and right FFA search spaces from a previous study on responses to images of faces in the ventral stream (***Julian et al., 2012***) using an iterated leave-one-run-out procedure. In each participant, we identified the most illness inference-responsive vertices in left and right FFA search spaces in five of the six runs (top 5% of vertices, *Illness-Causal>Mechanical-Causal*). We then extracted PSC for each condition compared to rest in the held-out run (*Illness-Causal*, *Mechanical-Causal*, *Noncausal-Illness First, Noncausal-Mechanical First*), averaging the results across all iterations.

In contrast to the PC, the FFA did not show a preference for illness inferences compared to mechanical inferences (leave-one-run-out individual-subject fROI analysis; repeated-measures ANOVA, condition (*Illness-Causal*, *Mechanical-Causal*) × hemisphere (left, right): main effect of condition, $F_{(1,19)}$ = 0.04, p=0.84, main effect of hemisphere, $F_{(1,19)}$ = 9.46, p=0.006, condition × hemisphere interaction, $F_{(1,19)}$ = 1.34, p=0.26; ***Figure 1—figure supplement 13***). Additionally, the FFA did not show a preference for illness inferences compared to noncausal vignettes, which contained illness-related language but were not causally connected (repeated-measures ANOVA, condition (*Illness-Causal*, *Noncausal-Illness First*) × hemisphere (left, right): main effect of condition, $F_{(1,19)}$ = 0.94, p=0.34, main effect of hemisphere, $F_{(1,19)}$ = 4.47, p=0.05, condition × hemisphere interaction, $F_{(1,19)}$ = 0.06, p=0.82; repeated-measures ANOVA, condition (*Illness-Causal*, *Noncausal-Mechanical First*) × hemisphere (left, right): main effect of condition, $F_{(1,19)}$ = 0.07, p=0.8; main effect of hemisphere, $F_{(1,19)}$ = 7.59, p=0.01; condition × hemisphere interaction, $F_{(1,19)}$ = 2.72, p=0.12; ***Figure 1—figure supplement 13***). Thus, although the FFA exhibits a preference for images of animates (e.g. ***Kanwisher et al., 1997***), the current evidence suggests that this region is not sensitive to abstract causal knowledge about animacy-specific processes (i.e. illness).