Deductive and Inductive Processing Dissociate in the Human Brain

Hope H Kean
(hopekean@mit.edu)
BCS, MIT, 43 Vassar Street
Cambridge, MA 02139 USA

Alexander Fung
(alexfung@mit.edu)
BCS, MIT, 43 Vassar Street
Cambridge, MA 02139 USA

Josh Rule
(rule@berkeley.edu)
2121 Berkeley Way
Berkeley, CA 94704

Josh Tenenbaum
(josh.tenenbaum@gmail.com)
BCS, MIT, 43 Vassar Street
Cambridge, MA 02139 USA

Steve Piantadosi
(spiantado@gmail.com)
2121 Berkeley Way
Berkeley, CA 94704

Evelina Fedorenko
(evelina9@mit.edu)
BCS, MIT, 43 Vassar Street
Cambridge, MA 02139 USA

Logical deduction and induction are fundamental components of human reasoning that have been argued to be distinct on theoretical grounds. However, it is unclear whether empirically, these are dissociable cognitive processes, or instead, instances of the same underlying cognitive operation. The representational format of logical reasoning has also been debated, with some arguing for linguistic representations, but others advocating for a symbolic but non-linguistic 'language of thought (LOT)' in which all logical reasoning is performed. Here, we use brain imaging (fMRI) to address both these questions and find that i) deduction and induction are neurally dissociable, and ii) neither form of reasoning relies on linguistic representations.

Introduction

Analytic and pragmatic theories of rational thought have postulated a divide between inductive and deductive reasoning (Carnap, 1950, 1952; Pólya, 1954; Priest, 1999; Rips, 2003; etc.). However, other proposals have posited a single domain-general processor that carries out hypothesis generation, confirmation, and elimination. In these latter proposals, inductive and deductive schemas both reduce to a general format, such as a mental model (Johnson-Laird, 1994) or a set of world descriptions attached to corresponding conditional probabilities (e.g. Lake et al. 2016; Wong & Grand et al. 2024), or to an ultimately domain-specific set of heuristics for navigating the world (e.g. Cosmides 1989; Cheng & Holyoak 1985). Critically, inducing novel hypotheses and deductive elimination of falsified hypotheses occur in the same model under all these accounts.

Some evidence for a dissociation between inductive and deductive reasoning comes from different developmental trajectories. In particular, inductive theory generation appears to emerge in infancy (Gopnik, Meltzoff, & Kohl, 1999). In contrast, deductive reasoning is much slower to develop: the ability to reason by denying the consequent (Modus Tollens) emerges around age 3 years (Mody & Carey 2016), and we remain quite poor at disjunctive syllogistic reasoning even in adulthood (Wason 1966). However, previous attempts to distinguish inductive and deductive reasoning using brain imaging have not produced a clear answer (Osherson et al., 1998; Goel & Dolan, 2004).

In the current study, we build on advances in our understanding of the neural substrates of abstract reasoning and fluid intelligence (e.g., Duncan et al., 2020) and use a state-of-the-art precision fMRI approach (Gratton & Braga, 2021), where all critical comparisons are performed within individual participants, to ask whether inductive and deductive reasoning dissociate in their neural substrates. In particular, the goal is to disambiguate among i) a **monolithic** account of human reasoning, in which inductive and deductive cognitive operations are examples of another more fundamental cognitive operation underlying all abstract reasoning (and perhaps goal-directed behaviors more generally), ii) accounts where **deduction is a special case of induction** or **induction is a special case of deduction** (e.g., higher complexity or higher uncertainty), and iii) an account where induction and deduction are **distinct** processes.

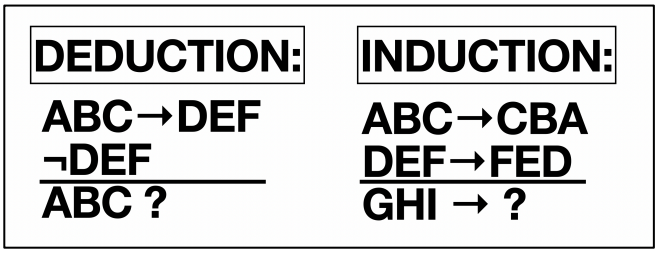

Figure 1. Illustration of the deduction and induction tasks used in the fMRI study.

In our study (n=16 participants), we aimed to study deduction and induction in their native formats (**Fig. 1**). For ***deductive reasoning***, we adopted a paradigm which taps deduction-specific cognitive load by contrasting mental operations for disjunctive syllogisms (Modus Tollens), which consist of denying the consequent—arguably the most fundamental deductive operation required for all hypothesis elimination, and a simpler operation (Modus Ponens), which is easier for humans because it falls naturally from affirming the antecedent (Monti et al., 2007; Coetzee & Monti, 2018). For ***inductive reasoning***, we adopted a paradigm from Rule, Piantadosi, and Tenenbaum (2020), where participants are provided with an input list and an output list and have to guess (induce) the rule that led to the input→output transformation. They then have a chance to test their hypothesis on a new input list, and so on until they guess the correct rule. In the control condition, they know what the rule is and simply have to apply it to a list. This rule-guessing paradigm plausibly taps similar cognitive operations as the Abstract Reasoning Corpus (ARC, Chollet 2019) and its sequels (ConceptARC, Moskvichev et al., 2023), as well as a standard tests of fluid intelligence (e.g., Ravens Progressive Matrices; RAPM, Raven et al., 1998), and the Number Game (Tenenbaum 1999), as well as perhaps more concrete novel concept learning tasks (Gauthier & Tart, 1997; Xu & Tenenbaum 2007).

Results

First, we asked **whether deduction and induction are neurally separable**. In the first analysis, we examined responses in the domain-general Multiple Demand (MD) network, which has been implicated in abstract reasoning and general fluid intelligence (Duncan et al., 2020) to the two critical tasks. The MD network was functionally defined using a standard 'localizer' based on a spatial working memory (WM) task (Fedorenko et al., 2013; Assem et al., 2020). We found that the MD network was engaged during inductive, but not deductive reasoning (**Fig. 2A**).

Next, we searched across the brain for regions that are sensitive to deductive reasoning load (i.e., respond more to Modus Tollens than Modus Ponens). We found 20 regions that responded strongly to deductive load (the effects were estimated in left-out runs of data, ensuring no circularity). When we searched across the brain for ROIs within which there was not any responsiveness to spatial WM, we found only 2 left frontal lobe ROIs (Fig. 2B), and these deduction-responsive regions were distinct from the MD network (also as evidenced by the lack of responsiveness during the spatial WM task conditions), and they also responded weakly to inductive reasoning (the response for the critical inductive condition is the same as to the Modus Ponens condition).

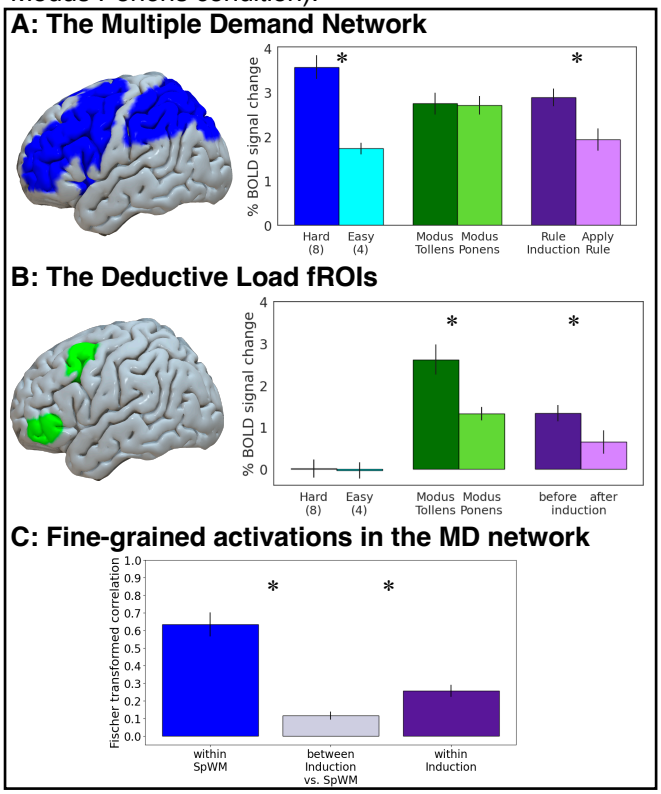

Figure 2. **A**: Responses in the Multiple Demand (MD) network (across regions; individual regions' profiles look similar) to hard and easy spatial WM (blue bars), deduction (green bars), and induction (purple bars; lighter bar is the easier, control condition in all cases). **B**: Responses in the Deductive load regions to spatial WM, deduction (estimated using independent runs of data), and induction. **C**: Spatial Correlations between task representations within the MD parcels (computed by Fischer transformed correlation coefficients).

Finally, for the induction task and the spatial WM task, both of which recruited the MD network, we asked whether the fine-grained neural representations are dissociable. To do so, we examined the similarity of the activation patterns within the MD areas i) across the runs within a task, vs. ii) between tasks, and found a robust dissociation (**Fig. 2C**).

To ask **whether deductive or inductive reasoning rely on natural language representations**, we examined responses in the language-selective fronto-temporal network, which has been implicated in linguistic comprehension and production (Fedorenko et al., 2024). The language network was defined using a standard 'localizer' based on a contrast of reading sentences vs. perceptually similar meaningless stimuli—nonword sequences (Fedorenko et al., 2010). This network showed no response during inductive reasoning or during deductive reasoning (note that both of the deductive task conditions use linguistic stimuli so elicit a positive response, but critically, the more demanding condition does not elicit a stronger response; cf. Fig. 2B) (**Fig. 3**).

In summary, deductive and inductive reasoning appear to be dissociable in the human brain: inductive reasoning recruits the domain-general network for abstract reasoning—the Multiple Demand network (although it shows a distinct fine-grained pattern within this network relative to a demanding working memory task). However, deductive reasoning recruits a distinct set of brain areas that respond only weakly during inductive reasoning. Moreover, although the format of reasoning representations remains an important open question, we can rule out the hypothesis that they rely on linguistic representations (cf. Carruthers, 2002)

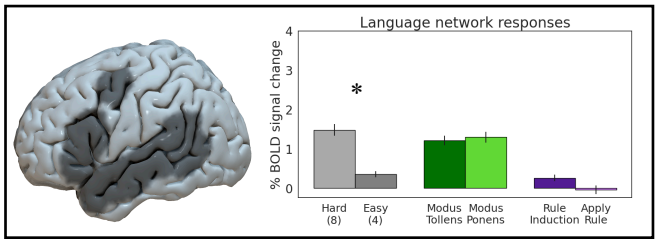

Figure 3. Responses in the language network (across regions; individual regions' profiles look similar) to sentence and nonword reading (grey bars), deduction (green bars), and induction (purple bars).

References

Carnap, R. (1950). Logical foundations of probability. *University of Chicago Press.*

Carnap, R. (1951). The Continuum of Inductive Methods. *University of Chicago Press.*

Pólya, G. (1954). Mathematics and plausible reasoning. I. Induction and analogy in mathematics. II. Patterns of plausible inference. *Princeton University Press.*

Priest, G. (1999). Validity. In Varzi A.C. (Ed.), *The nature of logic* (pp. 183–206). CSLI Publications.

Rips, LJ. (2003). The Psychology of Proofs: Deductive Reasoning in Human Thinking. *MIT Press.*

Johnson-Laird P. (1994). Mental Models and probabilistic thinking. *Cognition*, *50*, 189-209.

Lake, B., Salakhutdinov, R., Tenenbaum, JB. (2015). Human-level concept learning through probabilistic program induction. *Science.* 350 (6266), 1332-1338.

Wong, L., Grand, G., Lew, AK., Goodman, ND., Mansinghka, VK., Andreas, J., and Tenenbaum, JT. (2023) From word models to world models: Translating from natural language to the probabilistic language of thought. *arXiv:2306.12672*

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition, 31*(3), 187–276.

Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology,* 17(4), 391–416.

Gopnik, A., Meltzoff, A., and Kuhl P. (2000) The scientist in the crib: minds, brains, and how children learn. *Harper Collins*.

Mody, S., and Carey, S. (2016) The emergence of reasoning by the disjunctive syllogism in early childhood. *Cognition* 154:40-48

Wason, PC. (1966) Reasoning In B. Foss (ed.), *New Horizons in Psychology.* Harmondsworth: Penguin Books. pp. 135-151

Osherson, D., Perani, D., Cappa, S., Schnur, T., Grassi, F., Fazio, F. (1998) Distinct brain loci in deductive versus probabilistic reasoning. *Neuropsychologia* 36 (4) 369-376.

Goel, V., and Dolan, RJ. (2004) Differential involvement of left prefrontal cortex in inductive and deductive reasoning. *Cognition* 93(3):B109-21

Duncan, J., Assem, M., and Shashidhara, S. (2020) Integrated Intelligence from Distributed Brain Activity. *Trends in Cognitive Sciences* 24 (10):P838-852

Gratton, C., and Braga, RM. (2021) Deep imaging of the individual brain: past, practice, and promise. *Current Opinions in Behavioral Sciences* 40: 1-212

Monti, MM., Osherson, DN., Martinez, M., and Parsons, L. (2007) Functional neuroanatomy of deductive inference: A language-independent distributed network. *NeuroImage* 37 (3), 1005-1016.

Coetzee, JP., & Monti, MM. (2018). At the core of reasoning: Dissociating deductive and non-deductive load. *Human Brain Mapping*, 39(4), 1850.

Griffiths, T., Sobel, D., Tenenbaum, J., Gopnik, A. (2011) Bayes and Blickets: Effects of Knowledge on Causal Induction in Children and Adults. *Cognitive Science* 35: 1407-1455

Rule, J., Tenenbaum, J., and Piantadosi, S. (2020) The Child as Hacker. *Trends in Cognitive Sciences* 2073

Chollet, F. (2019) On the Measure of Intelligence. arXiv.

Moskvichev, A., Odouard V., and Mitchell M. (2023) The ConceptARC Benchmark: Evaluating Understanding and Generalization in the ARC Domain. arXiv.

Raven, R. (1998) Raven's progressive matrices and vocabulary scales. *John Hugh Court.*

Tenenbaum, J. (1999) A Bayesian framework for concept learning. *Doctoral Dissertation.*

Gauthier, I., and Tart, MJ. (1997) Becoming a "Greeble" expert: exploring mechanisms for face recognition

Xu, F., and Tenenbaum, J. (2007) Word learning as Bayesian inference. *Psychological Review* 114(2):245-72.

Fedorenko, E., Duncan, J., and Kanwisher, K. (2013) Broad domain generality in focal regions of frontal and parietal cortex. *Proceedings to the National Academy of Sciences.* 110(41):16616-21.

Assem, M., Glasser, M., Van Essen, D., and Duncan, J. (2020) Domain-general cognitive core defined in multimodally parcellated human cortex. *Cerebral Cortex*. 30(8): 4361-4380.

Fedorenko, E., Ivanova, A., and Regev, T. (2024) The language network as a natural kind within the broader landscape of the human brain. *Nature Reviews Neuroscience.*

Fedorenko, E., Hsieh, P., Nieto-Castañón A., Whitfield-Gabrieli, S., and Kanwisher, N. (2010) New method for fMRI investigations of language: defining ROIs functionally in individual subjects. *Journal of Neurophysiology.* 104(2):1177-94

Carruthers, P. (2002) The cognitive functions of language. *Behavioral and Brain Sciences.* 25 (6):657-674.