The Language of Thought is not Language: Evidence from Formal Induction

Hope Kean, Alex Fung*, Paris Jaggers*, Josh Rule, Yael Benn, Josh Tenenbaum, Steve Piantadosi, Rosemary Varley, Evelina Fedorenko

Contact: hopekean@mit.edu

MIT · UCL · UC Berkeley

BACKGROUND

Domain General — Abstracted from Content
Domain Specific — Domain-specific Content

Reasoning & Thinking

- Induction λ ? □→□ □→□ formal non-deterministic
- Deduction ¬ & ∨ → ∃ ∀ formal deterministic
- Abduction common-sense plausible non-deterministic — Pierce 1878
- Simulation — Data → Model — mental model theory — Johnson-Laird 1983
- Analogical — Source → Target — structure-mapping — Gentner 1983
- Physical "Intuitive Physics" — Battaglia et al. 2013
- Social "Theory of Mind" — Saxe & Kanwisher 2003
- "Core" Domains — Object, Place, Agent, Space, Num, Form — Spelke & Kinzler 2007
- Arithmetic Math
- Spatial WM
- Attention Cog Control

Research Questions:
1. What is the format of logical reasoning and abstract thought?
2. Do they use fundamentally linguistic representations?
3. Where are they processed in the human brain?
4. Is there a single universal symbolic 'language of thought' (LOT) in which all logical reasoning is performed?
5. What is the relationship between LOT and Natural Language (NL)?

Language of Thought Hypothesis (LOTH) Variants

LOTH Strong Form	LOT = NL	✗ Falsified by these data
LOTH Moderate Form	LOT ↔ MR* ↔ NL *Mental Representations	Not addressed by these data
LOTH Weak Form	LOT has analogies to NL	Trivial

HYPOTHESES

Linguistic
- Hypothesis 1: **Linguistic** format used for Logical Reasoning and the **Language System** is responsible for representing and executing LOT.

Non-linguistic
- Hypothesis 2a: Logical Reasoning is **non-linguistic** and accomplished by the **Multiple Demand (MD) System**.
- Hypothesis 2b: Logical Circuitry is non-linguistic and **Distinct** from both MD & Language.

RESULTS

The Language System

Language — Sentences > Nonwords
Spatial Working Memory — Hard (8 items) > Easy (4 items)
Induction — Pre 'Got It' > Post 'Got It'
Deduction — Modus Tollens > Modus Ponens
*see Methods

The Multiple Demand System

Neurally Dissociable from the MD activations
Spatial Correlations Between Task Representations
(within SpWM, between SpWM vs. Induction, within Induction)

Hypothetical *Deductive System*

Logic fROIs' non-responsiveness to Arithmetic & Physics
(Logic-related, Non-logic cognitive tasks)

Global Aphasics Induce Rules Requiring LOT Representations

(MIT Students, Controls Avg 55yo, GS late 40s, SA late 70s) — Proportion Rules Solved
Inferred Got It Trial

Global Aphasic MRI — Global Aphasic Rule Description

DISCUSSION

Logical Reasoning is Neurally Distinct from Linguistic Processing

i) Inductive and Deductive Logical Reasoning **do not rely on linguistic representations.**

ii) Instead, **inductive reasoning** recruits the domain-general network for cognitive demand—the Multiple Demand network.

iii) Furthermore, **deductive reasoning** recruits a distinct set of functional regions housed in the inferior frontal and middle frontal gyri.

iv) The format of reasoning representations (i.e. LOT) remains an important open question, however, **the LOT ≠ Language.**

Multiple Demand System — Induction — Deductive System — Language System

METHODS

Induction
Pre 'Got It' > Post 'Got It'
ABC→CBA DEF→FED GHI → ?
ABC→CBA DEF→FED GHI → ? λ rev

Deduction
Modus Tollens > Modus Ponens
ABC→DEF ¬DEF ABC ?
ABC→DEF ABC DEF ?

Group-constrained Subject-Specific Analysis ('GcSS')
individual activation maps ... n → % overlap → watershed → Atlas → Parcels
Julian et al., 2012 web.mit.edu/bcs/nklab/GSS.shtml

Individual fROI response profile Analysis ('mROI')
Fedorenko et al. 2010 spm_ss: evlab.mit.edu/resources

Task Topography Similarity Analysis ('spcorr')
Task 1 — Task 2

REFERENCES

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.
Pólya, G. (1954). Mathematics and plausible reasoning. I. Induction and analogy in mathematics. Princeton University Press
Osherson, D., et al. (1998) Distinct brain loci in deductive versus probabilistic reasoning. Neuropsychologia
Goel, V., & Dolan, RJ. (2004) Differential involvement of left prefrontal cortex in inductive and deductive reasoning. Cognition
Monti, MM., et al. (2007) Functional neuroanatomy of deductive inference. NeuroImage
Coetzee, JP., & Monti, MM. (2018). At the core of reasoning: Dissociating deductive and non-deductive load. Human Brain Mapping.
Rule, J., et al. (2024) Efficient learning of rule-based concepts via metaprogram search. Nature Comm.
Assem, M., et al. (2020) Domain-general cognitive core defined in multimodally parcellated human cortex. Cerebral Cortex.
Fedorenko, E., et al. (2010) New method for fMRI investigations of language. Journal of Neurophysiology.
Fedorenko, E., Duncan, J., and Kanwisher, K. (2013) Broad domain generality in focal regions of frontal and parietal cortex. PNAS
Fedorenko, E., Ivanova, A., and Regev, T. (2024) The language network as a natural kind within the broader landscape of the human brain. NRN
Tenenbaum, J. (1999) A Bayesian framework for concept learning. Doctoral Dissertation.
Carruthers, P. (2002) The cognitive functions of language. Behavioral and Brain Sciences.