

Essay

Not peer-reviewed version

The Concept of Generalized Reasoning and Its Underlying Circuitry

[Robert Friedman](#) *

Posted Date: 29 November 2024

doi: 10.20944/preprints202411.2410.v1

Keywords: Computation; circuitry; neural network; general reasoning; deep learning



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Essay

The Concept of Generalized Reasoning and Its Underlying Circuitry

Robert Friedman *

Department of Biological Sciences, University of South Carolina, Columbia, SC 29208, USA;
bobf111@zohomail.com

* Retired

Abstract: The properties and processes of general reasoning are often undefined in literature, leading to a perspective relegated at best to the boundaries of pure philosophy. It is therefore of interest to confine the term as a physical process of information flow instead. This dependence on the elements of matter and energy is the link between the human brain and the phenomenon itself, allowing for the hypothesis that an analogous form is possible of artificial design.

Keywords: Computation; circuitry; neural network; general reasoning; deep learning

Definitions

The base term reason derives from a statement offered in explanation or justification [1], while reasoning is the drawing of inference or conclusion through its use [2], and finally, the attribute of generalized refers to applicability of kind to a well defined group [3].

The properties and processes of general reasoning are undefined in the literature, leading to a perspective relegated at best at the boundaries of pure philosophy [4]. It is therefore of interest to confine the term as a process of information flow instead [5]. This is a dependence of the flow of the elements of matter and energy, a conceptual link between the human brain and the phenomenon itself, allowing for the hypothesis that an analogous form is possible by artificial design. However, it is also possible that the term is not grounded in the physical world and instead merely resembles a concept with sole existence in the Mind, and therefore its applicability is bounded by the constraints of metaphysics and its methods.

Reasoning as an informational process

Since advanced information processing is a property of the human brain, its mechanistic basis is in the neurons and their interconnections, a network of kinds of biological cells [6]. This is distinguished from the lesser forms of information processing in biological organisms, such as the regulatory networks of gene and protein expression and their dynamics for a biological code used by evolution and in the development of animals and their body plan [7].

The above perspective resembles a form of biological computation. Therefore, general reasoning as a phenomenon may be restricted in both the theory of physics and the informational sciences. It follows that the generalized scenario requires a "program" or "algorithm" that represents a distinct category of kinds [8], a putative lineage of related kinds of reasoning which are both unique and derived in relationship to any other forms of mental processing. An example of this in a skill of mathematical reasoning [9]. The summation of two particular integer values is a form of abstract reasoning with application to the physical world. Further, its generalized form applies to other cases of summing integer values, leading to its applicability across the other branches of mathematics and the sciences, as in the interpretation of discrete objects across a visual scene. This extends a putative process of a specific arithmetic operation to other instances not confined to the that of the Mind, regardless of its final interpretation as a kind of interpolation, or the putative, but orthogonal process, of sampling for the extrapolation of knowledge [10].

There is also the question on whether general reasoning is partly or wholly defined *a priori* in the human brain as a substance or is likewise dependent on an experiential existence, as originates from sensation and the interactions in the physical world. However, it is not biologically plausible, nor is there a known mechanical basis for, programming the developing brain with this attribute *a priori* [6]. Therefore, it may be taken as an assumption that general reasoning emerges from experience and programming instead of from an uninitialized neural network. Furthermore, it may be hypothesized that the informational elements of this activity forms patterns [11] that are expressed by a notation grounded in past knowledge of mathematics and computation, such as a subnetwork, a circuit, or an algorithm [5]. It follows that this form of biological computation is interconvertible with a computational framework, whether the translational process is a tractable and bounded computation or not. Another open question is whether, or to what degree, does this process depend upon a modular design [12], since any category of generalized reasoning must have a physical basis for its occurrence, a mechanism based on the conventional elements of matter and its composition from a set of smaller discrete set of elements (atomic perspective of reality).

Reasoning as circuitry in computation

If this knowledge of biological computation is robustly and relatively equivalent to that of an artificial form, as can be represented by a human engineering and artificial design [11], then the artificial neural network serves as the analogous form of the biological kind [5].

Recent work in machine interpretability of transformer circuitry [13] shows examples of information processing in this artificial setting, and includes a learning process that can lead to a general form of an algorithm and its putative computation, as in the skill of generalized reasoning, as observed in a large language model [14]. An example is shown where two circuits are formed in parallel, a process that is a type of "grokking" [15], that leads to the formation of circuitry dedicated to this higher form of information processing and search of the algorithmic space. The study further contends that the generalized circuit spanned particular but local layers across the neural network, but it is undetermined to whether this circuitry is confined by locality, and therefore a gradual building upon of the increasingly larger features of the network, or that the circuit can span without interruption across the network layers, and therefore bridge and "bind" the lower order features with that of the higher order features [16].

The above example shows the blurring of the boundary between the neural network approaches in engineering and the alternative practices of neurosymbolic ones [16]. However, the underlying process can be stated as originating and emerging from an unstructured neural network as opposed to any *a priori* design as exemplified by a neurosymbolic approach. This suggests that a neural network is a basis for forming and emerging of higher order concepts, and that they are not defined *a priori*. A hypothesis to fully test this concept is difficult because it depends on reachability, that any experiment is a robust measure of circuit design. This concept, and that of the others in the above sections, refer to the elements of higher cognition and refer to the recognition of information for the processes of advanced computation. Therefore, it can be said that the priors of cognition are a neural network that is "programmed" for construction of the informational pathways and patterns that serve as the intermediate basis for advanced computation [11].

Conclusion

These definitions and the definition of processes with a mechanical basis are suitable for hypothesizing and constraining the space of possibilities of experiments in the quest of to validation of any theory of generalized reasoning. It also suggests that the problem of advancing the skill of it is not an unreasonable proposition in itself, given the assumption it exists *a priori* in human cognition. Moreover, there is supporting evidence for it in the mechanisms of the binding of the lower to the higher order concepts, and a reminder of the categorical nature of generalized reasoning and its kinds, as it must depend upon the physics of information flow, as any of these advanced informational processes involve physical motion, and not depend on an abstract and undefined set of traits, but instead tractable to a formalization and definability as a physical process.

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Merriam-Webster Dictionary (an Encyclopedia Britannica Company: Chicago, IL, USA). Available online: <https://www.merriam-webster.com/dictionary/reason> (accessed on 28 November 2024).
2. Merriam-Webster Dictionary (an Encyclopedia Britannica Company: Chicago, IL, USA). Available online: <https://www.merriam-webster.com/dictionary/reasoning> (accessed on 28 November 2024).
3. Merriam-Webster Dictionary (an Encyclopedia Britannica Company: Chicago, IL, USA). Available online: <https://www.merriam-webster.com/dictionary/general> (accessed on 28 November 2024).
4. Friedman, R. Cognition as a Mechanical Process. *NeuroSci* 2021, 2, 141-150.
5. Friedman, R. A Perspective on Information Optimality in a Neural Circuit and Other Biological Systems. *Signals* 2022, 3, 410-427.
6. Friedman R. Themes of advanced information processing in the primate brain. *AIMS Neuroscience*, 2020, 7, 373-388.
7. Davidson EH, Erwin DH. Gene regulatory networks and the evolution of animal body plans. *Science* 2006, 311, 796-800.
8. Hennig, W. *Grundzüge einer Theorie der Phylogenetischen Systematik*; Deutscher Zentralverlag: Berlin, Germany, 1950.
9. Marcus GF. *The algebraic mind: Integrating connectionism and cognitive science*. MIT press, 2003.
10. Balestriero, R.; Pesenti, J.; LeCun, Y. Learning in High Dimension Always Amounts to Extrapolation. *arXiv* 2021, arXiv: 2110.09485.
11. Friedman, R. Tokenization in the Theory of Knowledge. *Encyclopedia* 2023, 3, 380-386.
12. Friedman, R. Higher Cognition: A Mechanical Perspective. *Encyclopedia* 2022, 2, 1503-1516.
13. Nanda, N., Chan, L., Lieberum, T., Smith, J., Steinhardt, J. (2023) Progress measures for grokking via mechanistic interpretability. *arXiv*, arXiv: 2301.05217.
14. Yang, S., Gribovskaya, E., Kassner, N., Geva, M. and Riedel, S. Do Large Language Models Latently Perform Multi-Hop Reasoning? *arXiv* 2024, arXiv: 2402.16837.
15. Power, A., Burda, Y., Edwards, H., Babuschkin, I. and Misra, V. Grokking: Generalization beyond overfitting on small algorithmic datasets. *arXiv* 2022, arXiv: 2201.02177.
16. Chughtai, B., Chan, L. and Nanda, N. 2023, July. A toy model of universality: Reverse engineering how networks learn group operations. In *International Conference on Machine Learning* (pp. 6243-6267). PMLR.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.