The Boundaries of Cognitive Closure: Argument for Mysterianism in the Philosophy of Consciousness

Danil Kutnyy
danil.kutny@gmail.com

Abstract

The "hard problem" of consciousness has long been debated in philosophy, with mysterianism suggesting that it may be inherently unsolvable due to cognitive or epistemic limitations. This paper introduces a new argument for mysterianism, drawing on insights from the complexity of artificial neural networks. Using a simple multilayer neural network trained to classify images as an example, it is shown that even understanding a single artificial neuron's role in information processing can be beyond our cognitive capabilities. When considering the complexity of biological neurons, which far exceed artificial neurons in intricacy, the challenges become even more pronounced. This raises questions about the feasibility of understanding consciousness, a vastly more complex phenomenon, by suggesting that our cognitive limitations extend to fundamental principles of interpreting complex systems. The paper emphasizes the challenges posed by layered abstractions, drawing parallels with other multi-level systems like microprocessors to argue that certain problems may be insurmountable.

1. Introduction

The problem of consciousness has resisted scientific and philosophical explanation for centuries. The "hard problem" of consciousness, as articulated by David Chalmers, highlights the challenge of explaining why and how physical processes in the brain give rise to subjective experiences (Chalmers, 1995)[1]. Mysterianism, a position associated with philosophers such as Colin McGinn, argues that consciousness may be inherently beyond human understanding due to cognitive limitations (McGinn, 1989)[2].

This paper seeks to extend mysterianism by drawing on the opacity encountered in artificial neural networks. By using a trained multilayer neural network as an example, it will be argued that the difficulties in understanding how artificial neurons process information may indicate a broader limitation in human cognition when it comes to interpreting complex, layered systems such as the brain.

2. Traditional Mysterianism: Cognitive Closure and the Explanatory Gap

Mysterianism suggests that humans are cognitively closed to the problem of consciousness, meaning that we may lack the mental faculties required to solve it. This position is often grounded in evolutionary considerations, proposing that human brains evolved for survival and reproduction, not for solving abstract metaphysical problems (Nagel, 1974)[3]. The "explanatory gap," as described by Joseph Levine, refers to the difficulty in explaining how subjective experiences (qualia) arise from neural processes, despite progress in neuroscience (Levine, 1983)[4].

Common criticisms of mysterianism argue that it prematurely concludes the problem is unsolvable, potentially stalling scientific progress (Dennett, 1991)[5]. However, traditional mysterian arguments typically do not draw upon analogies with complex artificial systems, leaving room for novel arguments that might bolster the position.

3. The Challenge of Interpreting Neurons

Artificial neural networks (ANNs), particularly deep networks with multiple layers, provide an informative analogy for understanding the challenges faced in interpreting complex systems. Consider a simple multilayer neural network trained on the CIFAR-100 dataset, which consists of images from 100 categories (Krizhevsky & Hinton, 2009)[6]. The network has fully connected layers, with each neuron receiving inputs from numerous connections, each connection weighted differently. As information passes through the network, the weights transform the input data into increasingly abstract representations, ultimately leading to a classification output.

Examining a single neuron in the network poses significant interpretative challenges. This neuron receives signals from potentially hundreds or thousands of connections, with each weight modifying the signal individual. Understanding what this neuron "does" involves deciphering how these weights interact to transform the input into some output feature. The feature itself may not correspond to any single recognizable pattern or visual component; instead, it could represent an abstract aspect of the image data, such as a combination of edges, colors, or textures (Bengio, Courville, & Vincent, 2013)[7].

For humans, comprehending what exactly this neuron is "looking for" or how it processes the diverse signals simultaneously is an immensely complex task, potentially on the verge of unsolvability. The difficulty is not just in tracking each weight's role, but in understanding how the complex, non-linear transformations produced by these weights give rise to a particular output.

Here, César Hidalgo's concept of "personbytes" becomes relevant, as it illustrates the limits of individual cognitive capacity when handling vast quantities of interrelated information (Hidalgo, 2015)[12]. According to Hidalgo, each person has a finite capacity for managing knowledge, which constrains the complexity they can engage with alone. Understanding a neuron's function in high-dimensional spaces, then, surpasses the limits of any one mind's *personbytes*, explaining why we struggle to grasp how a neuron might encode or respond to specific features.

This cognitive limitation may offer support for Mysterianism, which argues that some aspects of consciousness or understanding lie beyond human capability. If each neuron operates in a space of thousands of interconnections, then piecing together the roles of these connections into a coherent "understanding" may not just be difficult but cognitively impossible, restricted by our own mental bandwidth.

4. Introducing Layers of Abstraction

To further illustrate the challenge of understanding complex systems, consider the analogy of a microprocessor, which operates through multiple layers of abstraction. At the foundational level are transistors, the basic building blocks that control electrical signals. These transistors form logic gates, which perform simple logical operations (AND, OR, NOT) and combine inputs to produce outputs.

Above the gates, we have arithmetic logic units (ALUs), responsible for executing arithmetic and logical operations on binary data. These are integrated into the central processing unit (CPU), which orchestrates the flow of data and executes instructions.

Next is assembly language, providing a more human-readable representation of machine code, enabling programmers to write instructions that the CPU can execute without delving into binary complexities.

At a higher level, high-level programming languages (like Python or Java) allow for the creation of complex applications with syntax closer to human language.

5. Layers of Abstraction in Neural Networks

Similar to microprocessors, artificial neural networks (ANNs) also operate through multiple layers of abstraction. At the foundational level are weights, which represent the strength of connections between neurons. Each neuron receives inputs from multiple connections, and these weights determine how much influence each input has on the neuron's output.

Above the weights are the neurons themselves, which perform non-linear transformations on the weighted inputs. Each neuron's activation function determines whether it should fire based on the sum of its inputs, contributing to the network's overall functionality.

These neurons are organized into layers, where each layer transforms the data into increasingly abstract representations. The output layer then generates the final classification or prediction based on the processed information.

Many ANNs consist of multiple hidden layers, allowing for complex interactions and hierarchical feature extraction. Each layer builds upon the previous one, similar to how the CPU integrates various components to execute complex tasks (Koch &Segev, 2000).[8]

Just as the mystery of a microprocessor's behavior persists if we fail to understand its logic gates, the same applies to consciousness. Without a clear understanding of how neurons function and interact at these foundational levels, we may be left with a ground-up incomplete understanding of consciousness. There remains a significant gap in our knowledge between the complex operations of neural networks and the fundamental nature of consciousness, echoing the mysteries inherent in the workings of microprocessors. Just as a computer's high-level programs, like Microsoft Word, require insight into their foundational components—such as logic gates performing simple operations—understanding consciousness necessitates comprehension of its fundamental building blocks, such as neurons.

Moreover, cognitive closure at the level of these foundational elements means that even if we grasp the complex interactions of neurons, we might still struggle to solve the hard problem of consciousness. The inability to fully understand these essential components could render certain aspects of consciousness perpetually beyond our reach, much like how a lack of understanding of basic logic gates limits our comprehension of advanced computing.

6. Cognitive closure on each level of abstraction

In addition to the mystery gap and cognitive closure surrounding the understanding of neurons, we must consider that the interconnections between multiple neurons—within layers of a neural network—and the multi-layered architecture itself may also exceed our cognitive capacities. Just as understanding a single neuron's function is challenging, grasping the intricate web of relationships and interactions among hundreds or thousands of interconnected neurons poses an even greater difficulty. Each layer in a neural network adds another dimension of complexity, leading to emergent properties that cannot be easily traced back to the individual components.

This multi-layered interconnectivity can create additional cognitive closures at various levels of abstraction. Just as we may struggle to comprehend the role of a single neuron, we may find it equally challenging to understand how clusters of neurons within a layer work together to process information, or how entire layers interact to produce higher-level outputs. The clumsiness and complexity at these abstraction levels can consist of too many interrelated elements for our minds to grasp fully.

Thus, these multiple cognitive closures—from individual neurons to interconnected layers—compound the challenge of explaining consciousness. As we attempt to unravel the workings of neural networks, we may face not just one but several layers of cognitive barriers that inhibit our understanding, making the task of elucidating consciousness even more daunting.

7. Biological Complexity of Neurons and Brains

Consequently, the organization of neural networks within the brain is shaped by evolutionary processes, leading to systems that are inherently even more complex and challenging to interpret than a microprocessor. Unlike the meticulously designed architecture of a computer, which is built for functionality and clarity, biological systems have evolved through a gradual and often haphazard process, resulting in layers of complexity that can be daunting to decipher.

For instance, despite significant intellectual efforts in synthetic biology, our understanding of even the simplest single-celled organisms remains incomplete. When researchers attempted to construct a minimal synthetic cell, they discovered that approximately 25% of its genes had unknown functions (Hutchison et al., 2016)[10]. This highlights a fundamental gap in our knowledge, suggesting that even basic life forms possess intricacies that elude our comprehension.

In this context, biological neurons present an even greater challenge. Each neuron is a highly intricate cell, featuring thousands of synapses and a variety of ion channels, all involved in complex biochemical processes (Markram, 2006)[9]. In contrast, artificial neurons are simplified models designed for specific functions, often falling short of capturing the full operational richness of their biological counterparts. As we consider the task of understanding consciousness, it becomes apparent that while artificial neurons may operate at the edge of interpretability, real biological neurons represent an even more intricate puzzle. This disparity highlights another point for cognitive closure to appear; the complexity inherent in the structure and function of biological neurons may place them beyond our grasp, rendering the understanding of consciousness increasingly elusive.

8. Limitations of Current Research Methods

Finally, even if none of the described arguments for potential cognitive closure were factors, the complexities of understanding consciousness would remain daunting. To illustrate this, we can refer to research conducted by neuroscientists on microprocessors, highlighting the limitations of our current methods in unraveling complex systems.

In a notable study, researchers successfully applied advanced imaging techniques and computational models to analyze the workings of a microprocessor (Jonas & Kording, 2017)[11]. Using methodologies of neuroscience such as functional magnetic resonance imaging (fMRI), they visualized activity within the processor and correlated these patterns with specific operational functions. Remarkably, they identified neuron-like structures within the microprocessor that corresponded to certain computational tasks, akin to how neuroscientists have identified neurons associated with specific concepts, such as the “Jennifer Aniston neuron” that activates when subjects recognize her image (Quiroga et al., 2005)[13].

Despite these achievements, the research highlighted significant limitations. While the researchers successfully pinpointed specific components and their functions, this knowledge did not lead to a comprehensive understanding of the microprocessor as a whole. The intricate interplay between hardware and software created layers of abstraction that proved challenging to dissect, leaving the overarching operational principles of the microprocessor elusive.

This investigation reinforces the argument that if we cannot comprehend the fundamental workings of a microprocessor—a construct specifically designed and understood by humans—then extrapolating these methods to the complexities of biological systems, particularly the brain, is inherently problematic. As a result, the insights drawn from this study support the Mysterian perspective, suggesting that, irrespective of cognitive closure, our current methodologies may be insufficient to resolve the enigma of consciousness.

Summary

Covering all the arguments, the Mysterian view on consciousness finds robust support in light of these challenges.

1. **Interpreting Artificial Neurons**: The challenge begins at the lowest level with artificial neurons. Despite their relative simplicity, these units already test our interpretive capacity, suggesting cognitive limitations even when working with simplified, human-made models.
2. **Layers of Abstraction**: Consciousness is not built upon isolated neurons but through layered networks, each adding new degrees of abstraction. As a result, understanding the layered structure of the brain—from single neurons to complex networks—likely requires insights beyond those needed to grasp the functioning of individual neurons. Each layer of abstraction contributes to consciousness in ways that likely to present challenges at least as great as that of understanding a single neuron—if not greater.
3. **Biological Complexity**: Unlike artificial neurons, biological neurons are vastly more intricate, with complex synaptic connections, diverse neurotransmitters, genomical activity, life support and defensive systems and many more . This additional complexity further escalates the challenge, as these features go well beyond just individual neurons, but spread in all the levels of complexity above in the abstraction ladder a suggests a level of organizational detail that might place understanding consciousness out of reach.
4. **Limitations of Current Research Methods**: Even if cognitive closure were not an absolute barrier, our current research tools and methods seem inadequate for fully deciphering consciousness which only reinforces a Mysterian view: consciousness may remain elusive due to the limitations in both our cognitive and methodological approaches.

Together, these arguments suggest that consciousness may be inherently difficult , if not impossible, for us to fully comprehend, highlighting Mysterianism as a compelling perspective.

8. Conclusion

The paper proposes a novel argument for mysterianism by drawing on the interpretive difficulties encountered in artificial neural networks. The layered abstraction argument suggests that if we cannot fully understand even simple neural networks, solving the problem of consciousness may be fundamentally out of reach. This view underscores the limitations of human cognition in dealing with complex, multi-layered systems, adding a new dimension to the debate over the hard problem of consciousness.

References

1. Chalmers, D. J. (1995). "Facing Up to the Problem of Consciousness." *Journal of Consciousness Studies*, 2(3), 200-219.
2. McGinn, C. (1989). "Can We Solve the Mind-Body Problem?" *Mind*, 98(391), 349-366.
3. Nagel, T. (1974). "What Is It Like to Be a Bat?" *The Philosophical Review*, 83(4), 435-450.
4. Levine, J. (1983). "Materialism and Qualia: The Explanatory Gap." *Pacific Philosophical Quarterly*, 64, 354-361.
5. Dennett, D. C. (1991). *Consciousness Explained*. Little, Brown and Company.
6. Krizhevsky, A., & Hinton, G. (2009). "Learning multiple layers of features from tiny images." *Technical Report, University of Toronto*.
7. Bengio, Y., Courville, A., & Vincent, P. (2013). "Representation Learning: A Review and New Perspectives." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8), 1798-1828.
8. Koch, C., & Segev, I. (2000). "The role of single neurons in information processing." *Nature Neuroscience*, 3, 1171-1177.
9. Markram, H. (2006). "The Blue Brain Project." *Nature Reviews Neuroscience*, 7(2), 153-160.
10. Hutchison, C. A., et al. (2016). "Design and synthesis of a minimal bacterial genome." *Science*, 351(6280), 1414-1423.
11. Jonas, E., & Kording, K. P. (2017). "Could a neuroscientist understand a microprocessor?" *PLoS Computational Biology*, 13(1), e1005268.
12. Hidalgo, C. A. (2015). *Why information grows: The evolution of order, from atoms to economies*. Basic Books.
13. Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, 435(7045), 1102-1107. doi:10.1038/nature03687