The Self Modeling Problem

Toward a Neurobiologically Plausible & Mechanistic Explanation of Subjectivity

Nick M. Heinz[1] & Aramis D. M. Valverde[2]

Affiliation.
1. San Francisco State University
2. New York University: Bioethics

Contact.
aramis.d.m.valverde@nyu.edu

More Info.
admvalverde.com

Abstract

The ineffability and subjective nature of conscious experience—how it feels to be something—pose central challenges for scientific accounts of consciousness. We employ a representationalist perspective to analyze the cognitive and neural systems that could plausibly contribute to ineffability and subjectivity.

This poster introduces the "self-modeling problem" as key to understanding the ineffable aspects of consciousness. The self-modeling problem arises when the system that evaluates perceptual representations attempts to evaluate itself and fails to do so due to the structure of the system. This limitation could partially underlie the difficulty in explaining the nature of consciousness and in introspecting subjectivity.

This poster also introduces a basic cognitive system of awareness based on the functions of the cortex and basal ganglia, as well as their interactions with the thalamus and hippocampus. This cognitive system is then used to inform a philosophical argument for the emergence and structure of subjectivity.

Subjectivity, as generated by the system of awareness, is categorized into three types. Semantic Subjectivity, the 'what it's like' evaluative quality of experience; Affective Subjectivity, the 'what it's like' emotional aspect of experience; and Perceived Subjectivity, the 'I am a thing experiencing the world' quality of experience.

The overall objective is to contribute to the task of understanding why consciousness is perceived as difficult to comprehend and to explain how Representational Subjectivity, Affective Subjectivity, and Perceived Subjectivity arise from a mechanistic and neurologically informed perspective.

Key Concepts

Consciousness is a system which cannot evaluate itself, as it has no means by which to do so, and as such, an evaluation of that sort is not possible in principle.

Subjectivity, broken down into Semantic, Affective, and Perceptual Subjectivity, can be explained mechanistically through the processes of a cognitive system of awareness.

A rough minimal system of awareness can be derived from the functions and interactions between three basic cognitive systems:
1. A system representing the world, including the self, and evaluating those representations, corresponding to the cortex.
2. A system processing and quantifying those evaluations using metrics like reward potential, corresponding to structures in the basal ganglia.
3. A system using those quantifications to modify and contextualize the representations and propagational dynamics of the system representing the world and self, corresponding loosely to the thalamus and hippocampus.

The dynamics, structure, and training of this system of awareness can potentially describe and explain subjectivity and ineffability through:
1. The self-modeling problem, which describes the inability of the system to evaluate its own processes, could be used to explain ineffability.
2. Cortical-Cortical processing within the system of awareness could be used to explain Semantic subjectivity.
3. Cortical-Basal Ganglia-Thalamus X Hippocampus mediated processing within the system of awareness could be used to explain Affective Subjectivity
4. A process of representational prioritization, using dynamics and training over the development of the entire affective system, could be used to explain Perceptual Subjectivity.

The Self Modeling Problem

The self-modeling problem is proposed here to be a significant contributing factor to the ineffability of consciousness. This problem arises from the cognitive system's limitations when it attempts to evaluate the structure of its own processing. Similar to Gödel's incompleteness theorem, which states that a complex system cannot prove its own consistency[1], the self-modeling problem posits that the cognitive system cannot introspectively evaluate itself as it uses introspection to evaluate the introspective process. When the system that evaluates representations turns inward to evaluate itself, it encounters a blind spot, leading to the perception of consciousness as ineffable and subjective.

Consider the analogy of a fire: a fire cannot burn itself because the chemical process of oxidation, which constitutes fire, cannot be applied to itself. Fire is a process of oxidation, which requires oxidizable materials (wood, paper, etc.) and oxygen as inputs, and as such, it cannot be applied to itself. Similarly, the cognitive system that processes and evaluates mental representations cannot apply this process to itself. This self-referential limitation means that when we try to introspect and understand our own conscious experience, we encounter an insurmountable problem and subsequent failure. This is why consciousness often seems mysterious and beyond the reach of explanation.

When we try to consciously examine a conscious experience, the qualia of a thing, without referencing that same qualia, what we are attempting to do is isolate conscious experience itself, sans representations. This cannot be done, as experiences are representational in nature. Trying to consciously examine your perception of a sensation, the representational byproduct of consciousness, is akin to using the process without an input and expecting to get an output, in a situation where the input and output are the same thing. You cannot isolate experience by itself, because when you attempt to do so you are simultaneously trying to not focus on representative content and also trying to ascertain the qualities of that lack of representational content through representational content.

When we attempt this, say either in trying to understand consciousness itself or the qualia of a given thing, say the taste of chocolate, nothing happens, and the network defaults to the experience of having failed to isolate experience. This failure is then perceived as demonstrating the ineffability of consciousness because the introspective tools we used were indeed inadequate for the task. The process that evaluates and updates representational content is not equipped to ascertain its own functioning. It evolved to process representations, creating, manipulating, and updating representations of the body, the perceived self, and objects of the mind, both external and internal, but not to introspect on its own workings.

This self-modeling problem could help explain why conscious report is characterized by its ineffability. The cognitive system, designed to process and respond to external stimuli through the creation and evaluation of internal representations, lacks the mechanisms to fully analyze and articulate its own processes. Therefore, the act of introspection does not reveal the structure of our experiences but results in the representational perception of an ineffable quality. This limitation underlies the hard problem of consciousness, making it challenging to explain why our experiences feel the way they do, why chocolate tastes like chocolate.

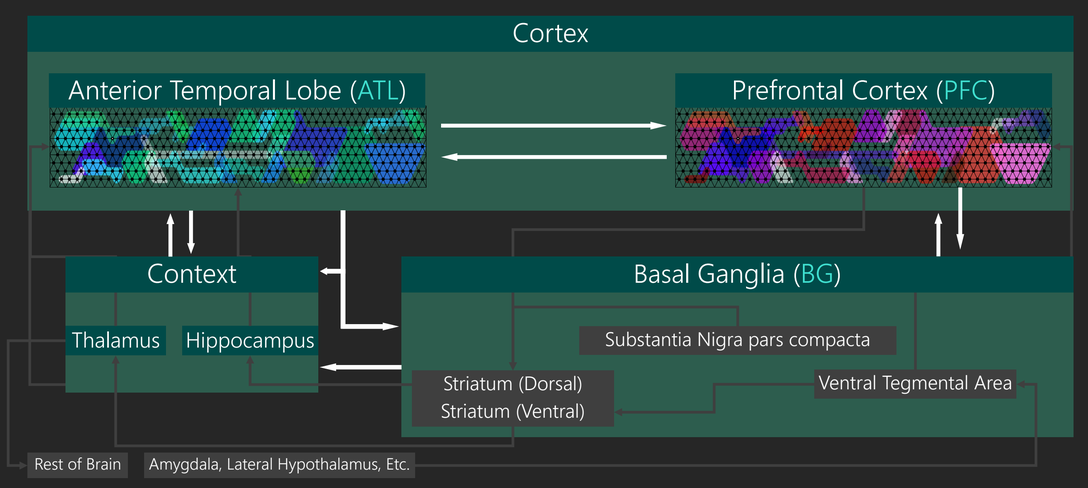
Simplified Cognitive System of Awareness

Figure 1:

Diagram of Simplified Cognitive System of Awareness

This diagram illustrates a simplified version of the structure of the proposed basic minimal system of awareness.*

(Most of the proposed full basic minimal system of awareness is in grey. These structures and the nuances of their interactions are important and will be outlined elsewhere, however for the scope of this poster discussion of the full system is not feasible, and consequently we have simplified the system to its most basic form. The text below also simplifies the dynamics and structures involved. The version of the system we will discuss below corresponds to the white directional interactions and green boxes. Work on the interactions in grey are forthcoming.



Explanation of Simplified Cognitive System of Awareness

The outlined system's primary function is to evaluate representational information, characterize its valence, and feed that information back into representational processing to fine tune the dynamics for the representational content and context.

Here the ATL holds semantic representations[2] and relational data between said representations, and communicates the activation of those representations to the hippocampus and the PFC[3].

The PFC functionally binds those representations and evaluates the pattern of representations in the ATL through further representational processing. The PFC then modulates the representations in the ATL accordingly, and communicates the determination of the overall valence of the active pattern of representations in the ATL to the BG.

The BG integrates this information, along with inputs from the ATL and many other systems,[4] to create a measure of how to respond to the current state of events. The BG then communicate that measure to the thalamus and hippocampus[5].

The Thalamus modulates the cortex at large[6], including the ATL and PFC and further modulates a great number of other neural systems, changing the dynamics to reflect the BG's evaluation of the current state of events.

The Hippocampus integrates the signals from the BG, PFC, and ATL[7], and changes the potentiation of neural ensembles in the cortex and/or activates neural ensembles in the cortex, according to the received input from the BG, PFC and ATL.

Semantic Subjectivity

Semantic Subjectivity is the evaluation of the system of awareness concerning what is being represented in the mental space. Note that 'represented' is used synonymously with 'perceived,' reflecting our strongly representationalist[8] approach to consciousness.

Semantic subjectivity is the phenomenon that occurs when two people disagree on the quality of a film or their feelings towards red velvet cake. Let's examine how the reciprocal processing between the ATL and PFC can explain the emergence of semantic subjectivity through an example.

Example: Let's say Kenny and Avery are walking by a dog pound, and both see a husky scratching at the door of its cage. Kenny gets a warm feeling and smiles, and goes about his day. Avery walks a bit faster and similarly goes about his day. Both saw a dog, but Kenny evaluated it as a friendly and warm thing, and Avery felt threatened.

Possible Mechanistic Explanation:

The ATL holds object representations[9]. When a person perceives a car or a stapler, a neural ensemble in the ATL corresponds to it and its relation to other representational neural ensembles. The PFC maintains representations related to goals and the evaluation of the patterns of activation of the ATL, mainly to create a measure of reward value of the activity and also to guide behavior in a context appropriate way[10]. To do so, it has to evaluate the representational pattern and bind dynamically to it, so that other mechanisms can integrate the signals of both accordingly.

For Kenny and Avery, the ATL signal was the same: the percept of a dog. However, prior experience biased Kenny's PFC to evaluate dogs as rewarding, while Avery's experiences taught him that dogs were not rewarding. Thus, Kenny sees the dog scratching at the cage door and views it positively, with positive evaluations bound to the dog's representation. Avery, expecting no reward, keeps his distance.

Affective Subjectivity

Affective Subjectivity is the emotional or "felt" aspect of subjective experience. It includes your feelings and reactions to experiences, such as smelling a flower or the prospect of getting tenure.

Example: Kenny and Avery are walking in a park when a husky puppy runs up and barks. Kenny is happy to see the puppy, while Avery reacts negatively, feeling stressed and wanting to get the dog away.

Possible Mechanistic Explanation:

As before, the ATL signal for both Kenny and Avery is the same—the percept of a dog. And as before, the PFC's evaluation of the expected reward value is positive for Kenny and negative for Avery. Avery's PFC makes the determination that the experience is negatively valenced, while Kenny's PFC finds a positive valence.

The downstream processes in the basal ganglia integrate the determination of valence from the PFC[10] into a measure of how to respond to the current state of events. The BG then communicates this information to the thalamus and hippocampus. The hippocampus takes this input, alongside the representational information from the ATL, and potentiates the relevant representational ensembles in ATL according to prior experiences.

For Avery, negative experiences with dogs biased the ATL and PFC towards a further negatively valenced evaluation. Avery's thalamus similarly integrates the signal from the BG and biases a wide variety of neural sub-systems towards a dynamic corresponding to the negative stressful evaluation. For Kenny, the BG communicates the puppy is a rewarding stimulus, and so the same system integrates the signal and biases the processing of the cortex and multiple other neural sub-systems towards a generally positive evaluation.

In this scenario, the affective aspect arises from the widespread biasing of neural systems to accommodate the PFC's valence evaluation of the experience, as it relates to the perceived pattern of representations active in the ATL. The subsequent feeling was a system-wide reaction to the quantified valence, which, through the thalamus and hippocampus, biased the representational and evaluative ensembles in the PFC towards the corresponding state.

Perceptual Subjectivity

Perceptual Subjectivity is the feeling that you are a thing experiencing things, the evaluation that you are a thing distinct from the rest of the world, which is currently having experience wash over you. Here we argue that this feeling is mediated by two factors, one cognitive and one developmental.

Because people believe and experience that there is a self which is separate from everything else, and which takes in everything else as an input, we have been forced to examine consciousness from a broken foundation. If conscious contents are representational in nature, the self-representation ought to be as well. Here we posit that the self representation is a representation like all the others in structure and in how it is processed by the conscious system at large, but is prioritized by the cognitive system at large due to the attribution of sensory input and concurrent evaluation of that sensory input in conjunction to the self-representation. This is the reification of the self-representation.

Since the self-representation is affected by sensory inputs like pain and pleasure through functional binding, unlike other representations, it becomes functionally isolated. Because of this functional isolation and prioritization, we come to believe that the self is a discrete and separate thing which has experience wash over it as a separate thing.

We propose that the object of perception, the perceiver, and the process of perception are all the same and are best described by attributing the self to the process of perception. In this view you are the process of perception and all of your percepts. Under normal conditions, the reification of the self ensures that perception is attributed to the self-representation and somatic inputs, excluding other representations that are essentially the same in form and function.

This implies that if functional isolation is diminished, other representations would be attributed to the self, and the system would identify with those representations. We conjecture in another presentation that this is the case when ego death occurs through the use of serotonergic drugs like LSD. During ego death, the self-representation becomes bound to different representations due to the functional coactivation of those ensembles and others, such as those of your parents or the universe. We argue that filopodia and increased potentiation, which allow propagations to coactivate many representative ensembles, are the neurological factors underlying the phenomenology of ego death.

Conclusion

Consciousness is a complex concept, made all the more complex by our preconceived notions of what it ought and ought not be. Here, we have proposed a potential framework for explaining subjectivity—a critical aspect of consciousness—through a mechanistic and neurologically-based model of awareness. This system of awareness consists of a cortical system for representational and evaluative processing and a basal ganglia system for modulating the cortical system, mediated by the thalamus and hippocampus.

In doing so we have outlined a potential line of inquiry, which could possibly come to elucidate important aspects of subjectivity and consciousness at large. We believe firmly that the tracing of neural processes to phenomenal experience is the best path forward to understanding consciousness, and believe that this model and set of explanations sets out a framework by which to do so.

Moreover, we believe that the self-modeling problem and the development of subjectivity offer valuable perspectives for the study of consciousness. We find the self-modeling problem and the reification of the self to be a compelling account of the evaluation of a separate self experiencing the world and of the emergence of the ineffable and subjective aspects of consciousness. That the model we propose provides a medium which could instantiate those accounts is also encouraging.

As we move forward, we will work to complete and finalize an account of the function of the basic system of a awareness without simplifications, and continue the process of making this theoretical work concrete through empirical evaluation.

Note: This updated presentation prioritizes the self-modeling problem and the neural basis of subjectivity, diverging from the initial emphasis on the PRISM model as outlined in the originally submitted abstract. The PRISM model, which was more general, lacked the more direct ties to neurological systems that is seen here. Due to spacing limitations the experimental predictions and practical applications are to be found in the link below.

For Full Citations, Digital Poster, & Paper,
Please Visit admvalverde.com/assc27.html