From Connectome to Cognition:

Proposing a Connectome-Based Hypernetwork Model to Interpret the Conjunction Fallacy & Possible Representational Structure of Qualia

Aramis D. M. Valverde

Affiliation.
New York University: Bioethics
Contact.
aramis.d.m.valverde@nyu.edu
More Info.
admvalverde.com

Abstract

This poster introduces a connectome-based hypernetwork model and framework designed to offer a novel explanation for human performance on the conjunction fallacy, and to explore potential representational structures related to qualia in the brain. The conjunction fallacy, a well-documented cognitive error where combined events are erroneously perceived as more probable than individual ones, provides a compelling context to examine the underlying mechanisms of semantic and representational knowledge processing through the lens of theoretical neuroscience.

Leveraging recent advancements in connectomics and hypernetwork science, this model utilizes the formalization of the intricate high-dimensional architecture of the neocortical connectome as a directed weighted hypergraph. We then employ that hypergraph formalization to investigate the possible functions of the corresponding neocortical circuits. This theoretical approach allows for the hypothetical simulation and analysis of the processes by which relational & semantic representations might be neurally instantiated by a single mechanism, and contribute to cognitive biases like the conjunction fallacy. By integrating mathematical models from hypernetwork science with existing empirical data, we propose to trace the possible neural systems that underlie these cognitive errors, potentially offering deeper insights into how the brain organizes and processes representational information.

This poster also explores the theoretical possibility of representing qualia, or personal experiences of consciousness, within neural network ensembles functionally bound into percepts, aiming to sketch a tentative link between phenomenological experiences and their neural underpinnings. This approach cautiously explores the potential neural foundations of representational processing and cognitive biases, setting the stage for further investigation guided by an empirical representationalist framework.

The proposed model of representation and theoretical framework, along with empirical explorations into its validity, may help delineate the intricate neural dynamics behind cognitive functions and percepts. In doing so, they may also contribute to a foundation for bridging abstract cognitive concepts with their physical neural bases.

Key Concepts & Background

The Conjunction Fallacy

The conjunction fallacy is a well-documented and replicated[1] logical error where individuals mistakenly rate the likelihood of two events occurring together as higher than the likelihood of a single event occurring alone. Despite extensive research, there is not yet a clear and conclusive explanation for its occurrence. Recent research has indicated that the use of mental simulation to ascertain likelihood may underlie a physics based version of the fallacy[2], but there is no work extending that idea experimentally to the fallacy at large.

Relational & Semantic Representation

This study proposes an abstract neuro-computational model utilizing directed hypergraphs to represent relational and semantic structures in the brain. The conjunction fallacy serves as an ideal subject for studying representational processing owing to its clear reliance on one's internal conceptual and relational structures. If relational processes and mental simulation do underlie the fallacy, there is reason to believe that this formalization and the accompanying model of event likelihood determination will be able to simulate human error.

Directed Hypergraphs

The research leverages directed hypergraphs to model the complexity of neural connections, offering a basic understanding of cognitive processes. This mathematical structure was chosen due to compelling evidence from cortical connectomics studies[3] which determined propagations in mammalian cortex take graph like paths through a directed hypergraph network.

Furthermore, directed hypergraphs are data structures that can effectively model complex relational networks. Their properties, such as nestedness, compositionality, and robustness to node loss[4], make them suitable for representing the intricate representational architecture of the neocortex.

A directed hypergraph is a generalized version of a graph, where edges can join an arbitrary number of vertices. A rough but more or less adequate way of conceptualizing directed hypergraphs and propagations through them is by thinking of them as directed and generalized Venn Diagrams of concepts, where sub-sets are able to take part in an arbitrary number of sets and point to other sets. The label of the Venn Diagram is the name of the concept. Lets say we are thinking about edible things, we may think about fruits and proteins, those would be two sub-categories. We can also think about red edible things, which would encompass strawberries and red meat. The directedness is akin to situations where a set is a part of another set, so all meats are proteins, but not all proteins are meats.

The Subjective Probability Determination System

The Subjective Probability Determination System, or just likelihood determination system, is proposed here to be a mechanism which evaluates the likelihood of a proposed scenario against another scenario by simulating the first scenario and then evaluating the overall global activation it elicits. The system then retains that evaluation, simulates the second scenario, evaluates global activation, and then compares the global activation of the first level of activation against the second. More specifically, it is a system which activates the neuro-representational ensembles relevant to the scenario, uses a population code to evaluate the overall activation elicited by the scenario, retains that activation through either a serotonergic or reciprocal activation process, simulates the second scenario in the same way, and then determines which one elicited the greatest global activation.

Representationalism About Consciousness.

There are multiple versions of Representationalism about consciousness[5]. This study takes the position that the phenomenal character of an experience is wholly encapsulated by representational content and the neural ensembles which are functionally bound to them (i.e. visual cortex ensembles), and requires processing and evaluation by a system which performs consciousness. In this view, representations are about internal experiences and not external objects. They gain their "about-ness" through relations to other representational ensembles and the processes which modulate processing of representations and the relations to other representations.

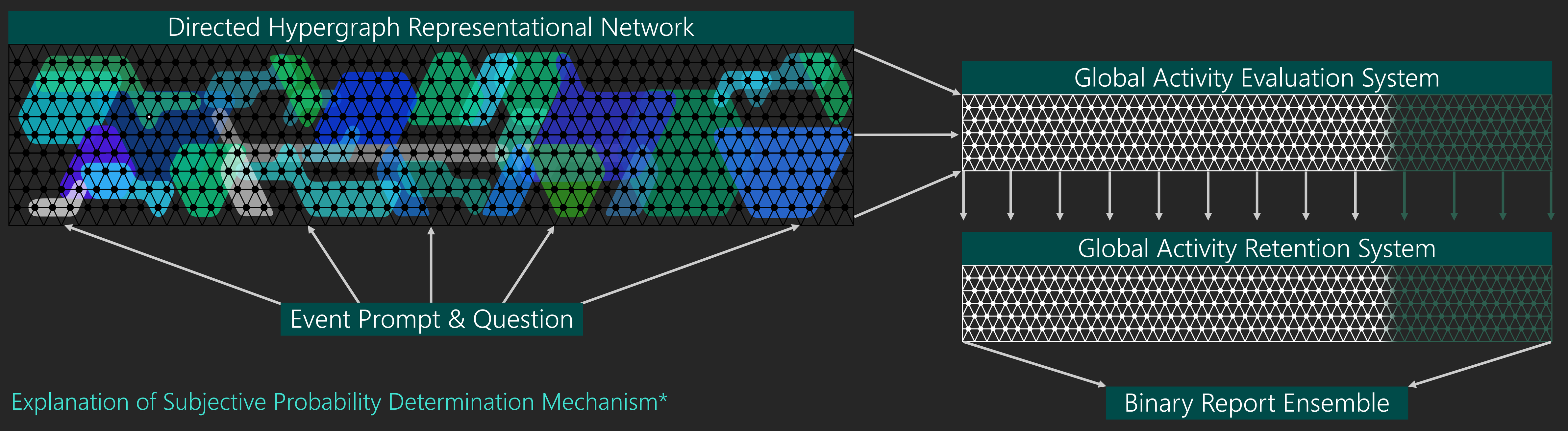
Conjunction Fallacy Example: Linda

```
Linda is 31 years old, single, outspoken, and very bright.
She majored in philosophy. As a student, she was deeply
concerned with issues of discrimination and social justice,
and also participated in anti-nuclear demonstrations.

Which is more probable?

(a) Linda is a bank teller
(b) Linda is a bank teller and is active in the
    feminist movement.
```

Figure 1: The Linda Problem, the original version of the conjunction fallacy.
Eighty-eight percent of participants chose B, the incorrect answer[6].

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

Subjective Probability Determination Mechanism



Explanation of Subjective Probability Determination Mechanism*

(It is important to note that the system outlined below is largely a theoretical framework based on a synthesis of known processes and their conjectured functions. The Directed Hypergraph Representational Network, the overall basis of the computational process, and the basic process of the global activity evaluation system are conjectured to be roughly equivalent to the corresponding neural processes used by humans. However, the Global Activity Retention System and Binary Report Ensemble are constructs used to create an actionable signal to approximate the specific binary response elicited by the conjunction fallacy as it is most well known. The principal object of investigation is the Directed Hypergraph Representational Network, the measurement of global activation corresponding to the question, and the role of priming and potentiation.)

The primary function of the outlined system is to evaluate the likelihood of a given scenario. We will use the Linda problem to illustrate the proposed process.

The system functions by using the beginning part of the event prompt as a network primer, which through activation potentiates in the short term (with paired pulse facilitation[7]) the representative ensembles corresponding to the characteristics attributed to Linda, namely, left-wing activism, outspokenness, intelligence, and independence. As such, the Directed Hypergraph Representational Network now has a bias towards the activation of the potentiated ensembles.

Subsequently, when the question is asked, "Which is more likely, that she is a bank teller, or a bank teller and an activist?", the neural system initiates a propagation corresponding to "bank teller," which then propagates through the corresponding ensembles in the Directed Hypergraph Representational Network.

The Global Activity Evaluation System then quantifies the global activity elicited by the prompt using a population code.

The Global Activity Retention System receives the population code, and retains the received population code measure of global activation through internal reciprocal activation.

Subsequently, the Binary Report Ensemble, which determines the direction of change in activation (more or less activation), receives a continuous measure from the Global Activity Retention System corresponding to the maximum degree of activation

At this point we have a retained measure of activation and a process which determines whether there is greater or lesser global activation and reports it in a binary manner, which is receiving a continuous signal.

Now, the scenario of Linda being a bank teller and a feminist is simulated. The neuro-representational ensembles corresponding to being a bank teller and a feminist are activated, and the signal propagates. Because the network was primed with concepts closely related to feminism (left-wing activism, intelligence, outspokenness, and independence), this propagation elicits greater activation across the network. The Global Activity Evaluation System's population code for activation reflects this, and the signal is passed on to the Global Activity Retention System, which updates and activates additional ensembles to reflect the new input, producing a stronger downstream signal.

The stronger signal reaches the Binary Report Ensemble, which, due to the change in signal strength, now fires. This provides a binary result by determining whether there is more or less activation compared to before, thereby answering the question, "which is more likely, that Linda is a) a bank teller, or b) a bank teller and a feminist?"

This system is based on known neural architectures and processes exist in mammalian brains, with the exception of the Global Activity Retention System and the Binary Report Ensemble. While the Directed Hypergraph Representational Network and Global Activity Evaluation System are conjectured to closely mirror the actual processes in the brain, the reporting mechanism was modified to accommodate the proposed experiment below.

Processing Flow Chart

Event Prompt
Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti -nuclear demonstrations.

↓

Representational Network
Activation elicited by prompt flows through, and potentiates the network towards the semantic representations mentioned by paired-pulse facilitation, meaning that if there is a second input to the same neural ensemble, it is much more likely to fire and to release more neurotransmitter.

↓

Question
The question is asked, "Which is more probable? (a) Linda is a bank teller (b) Linda is a bank teller and is active in the feminist movement." and the propagation for the evaluation begins for Linda being a bank teller.

↓

Representational Network
The propagation for Linda being a bank teller flows through. Given that there are no ensembles which were activated for bank teller (when is the last time you went to the bank and thought someone was outspoken?), the propagation is not potentiated.

↓

Global Activity Evaluation
Aggregates and communicates the amount of global activity through the representational network as it relates to Linda being a bank teller using a population code.

→

Global Activity Retention
Retains the global activity registered in the Global Activity Evaluation system.

↑

Question is Asked
The system is now asked to evaluate if Linda is likely a feminist and bank teller.

↑

Representational Network
Propagations elicited by question for "bank teller" and "feminist" flow through. Crucially, due to the proposed structure of representations, the ensembles which correspond to feminist, which include segments of those for the potentiated attributes, benefit from paired pulse facilitation. So the propagation is stronger than before.

↑

Global Activity Evaluation
Aggregates and communicates the amount of global activity in the representational network as it relates to Linda being both a bank teller and a feminist.

↑

Global Activity Retention
Updates the its measure of global activity as registered by the evaluation system, reflecting the now higher degree of activation

↑

Binary Report Ensemble
The ensemble is responsible for measuring the change and direction of change of the Global Activity Retention system. It checks if population code goes up or down, corresponding to if global activation went up or down. Since it went up, the second activation is determined to be the correct one, and the system has determined that it is more likely that Linda is a feminist and a bank teller, and as such the system as a whole has committed the conjunction fallacy.

Experimental Validation Proposal

The objective of this experimental validation proposal is to examine the underlying cognitive processes associated with the conjunction fallacy through the development and testing of a neuro-computational model of relational and semantic representation. This proposal is structured into three main sections: model development, validation of basic processes, and validation through personalized conjunction fallacy tasks.

1. Develop the computational model of relational representation and the tools to personalize it. This involves creating a program based on the hypergraph formalism and validating the expected dynamics. Initially, the model will be benchmarked against human performance on similarity tasks.

2. Evaluate the hypothesis that the conjunction fallacy arises from the proposed system, where the profile primes the semantic relational representational network, thereby increasing the global activation of that simulation. This is accomplished through the multi-trial use of the model and comparisons with participant or open-source data. If the model is accurate, we expect not only a consistent replication of participant responses, but also a similar multi-trial rate of error with human participants, as some questions elicit the fallacy more than others.

3. Evaluate the hypothesis that the representational structure corresponds with the directed hypergraph formalism by probing participant relational structures. For example, assess whether participants have a negative or positive opinion of police. Then provide a version of the conjunction fallacy that accounts for these relationships and probes for the effects of loosely related concepts being primed. The model may be able to accommodate individual variations on highly subjective or valanced questions.

The proposed experimental validation aims to rigorously test a neuro-computational model of relational and semantic representation in the context of the conjunction fallacy. If successful, this research will provide significant insights into the neural and computational foundations of relational representation and enhance our understanding of human cognitive processes. Future work on this project will require collaborative efforts to properly validate and expand the model. This includes efforts to expand the framework to artificial systems and incorporate neuroimaging methods to confirm the suspected areas responsible for these computations, namely the anterior temporal lobe.

Conclusion

This study has proposed that the conjunction fallacy is more than a cognitive bias, that it is a deeply rooted consequence of the way our brains represent and process relational representative information.

The Directed Hypergraph Representational Network has been put forward as a robust framework for the investigation of representational processes. Furthermore, in examining a proposed model of probability determinations which leveraged conjectured directed hypergraph mediated processes, the poster outlined how the structure of representations and neural ensemble potentiation could jointly influence thought towards committing the conjunction fallacy

The proposed experimental validation aims to show that this model can replicate human responses and error rates across various scenarios, demonstrating its potential to mirror the complexities of human thought. Additionally, personalized conjunction fallacy tasks are expected to highlight the model's adaptability to individual differences, further underscoring its relevance and applicability.

This model does not account for all instances of the conjunction fallacy, or even a large subset of them. As such there remains a great deal of work to be done to definitively demonstrate its relevance to and coherence with human probability processing and in instantiating and understanding the fallacy. Nonetheless, we believe that this is a utile and computationally robust account, which coheres with connectomic research on mammalian cortical processing.

Turning towards consciousness and qualia, if you take a strongly representationalist approach to consciousness, you need to subject your account of representation not just to tests of representational capacity and coherence with human processing, but also conscious processing. What that means is yet to be defined, but it seems that if you are going to state some thing is the neural basis of representations, and you are a representationalist about consciousness, then you ought to be able to state how representations are processed into conscious percepts.

Overall, this proposal aims to advance our understanding of cognitive biases through the development of a neuro-computational model. If strong representationalist theories of consciousness are correct, this model could also provide fundamental insights into the neural basis of consciousness, paving the way for future research in artificial intelligence, education, and cognitive therapy.

Future directions include a more robust model which is completely attributable to known neural processes and informed by neuroimaging experiments, the proposed experiment above, and the full integration of this model into our larger model of awareness.