Does Connectionism undermine Fodor's Language of Thought Hypothesis?

Jonathan Fryer*

University of Bristol

In 1975, Fodor hypothesised that thought is structured in much the same way as language.[1] Thoughts have semantics, a combinatorial syntax, and store information symbolically. In the 1980s, Connectionism looked to undermine his view. It suggested that mental information is stored non-symbolically in neural nets; it was considered a "paradigm shift" for cognitive theories.[2] In the 1990s, further work by Chalmers and Rowlands undermined Fodor's Language of Thought Hypothesis.[3 4 5] Modern cognitive research into Deep Learning uses an inherently Connectionist framework.

This paper separates Fodor's hypothesis from his arguments in its support. It argues that Fodor's Language of Thought Hypothesis is still a legitimate theory of cognition. However, it accepts that Fodor's arguments in favour of his hypothesis are fallacious. The paper examines three of Fodor's arguments for a language of thought: the only game in town argument, the argument from systematicity and productivity, and the argument from isomorphism.[6 7 8 9] It shows each to be flawed.

Further, this paper dismisses the dilemma Fodor and Pylyshyn present the Connectionist: that they must either merely implement his Language of Thought Hypothesis or concede that it is an inadequate theory of cognition.[10] he paper uses Chalmers' Backpropagation Model, a system that encodes grammatical information without using symbols, to escape the dilemma.[11]

Throughout, I argue that despite successfully undermining his arguments, Connectionism does not undermine Fodor's Language of Thought Hypothesis. I provide two positive reasons to upholding the Language of Thought Hypothesis. This paper concludes that – at present – neither Connectionism nor Fodor's Language of Thought Hypothesis has undermined the other.

1 Introduction

Connectionism undermines the arguments Fodor provides for a language of thought, but it does not undermine the Language of Thought Hypothesis (LOTH) itself. I distinguish between the LOTH – the thesis that thought is syntactically

*Jonathan Fryer completed his undergraduate studies at Bristol University with a First-Class Honours BA in Philosophy. His philosophical interests include the philosophy of psychology, ethics of all kinds (meta, normative, applied), and philosophy in the digital age. Having taken a year's break to recover from a COVID-filled undergraduate experience, he will begin an MPhil in Philosophical Studies in September at University College London. Jonathan considers himself "rubbish at science", and credits his interest in this area of philosophy to his lecturer at Bristol University, Max Jones.

1. Jerry A. Fodor, *The Language of Thought* (Harvard University Press, 1975).
2. Jerry A. Fodor, "Why There Still has to Be a Language of Thought," in *Psychosemantics* (MIT Press, 1987), 82.
3. D. Chalmers, "Why Fodor and Pylyshyn Were Wrong: The Simplest Refutation" (1990).
4. David J. Chalmers, "Syntactic transformations on distributed representations," in *Connectionist natural language processing* (Springer, 1992), 46–55.
5. Mark Rowlands, "Connectionism and the Language of Thought," *British Journal for the Philosophy of Science* 45, no. 2 (1994): 485–503, https://doi.org/10.1093/bjps/45.2.485.
6. Fodor 1975.
7. Fodor 1987.
8. J. A. Fodor, "The Language of Thought," *Critica* 10, no. 28 (1978): 140–143.
9. Jerry A. Fodor and Zenon W. Pylyshyn, "Connectionism and Cognitive Architecture: A Critical Analysis," *Cognition* 28, nos. 1-2 (1988): 3–71, https://doi.org/10.1016/0010-0277(88)90031-5.
10. Fodor and Pylyshyn 1988.
11. Chalmers 1990.

structured – and the arguments Fodor provides in favour of it. My separation of hypothesis and supporting argument is vital: the arguments support the LOTH, but they are not the LOTH.

I focus on three of Fodor's arguments:

1. The 'only game in town' argument[12]
2. The argument from systematicity and productivity;[13] and
3. The argument from isomorphism.[14]

Section 1 sets out the LOTH and the three arguments. I show that the LOTH can still be true, even if all three arguments are flawed. I then set out the Connectionist challenge in Section 2,[15] using Sanderson's model program that recognises hand-written digits as an example.[16] The existence of coherent Connectionist models undermines the argument that a language of thought is the 'only game in town'. It does not undermine the LOTH.

Fodor responds to the Connectionist challenge in a paper with Pylyshyn, in which he employs (2).[17] I discuss this in Section 3. They argue that symbol manipulation — a property of classical cognitive architecture — is required to explain the nomic necessity of systematicity and productivity in thought. An adequate theory of cognition must be able to explain this. Connectionism, therefore, either merely implements classical architecture or is an inadequate theory of cognition.

For Connectionism to undermine (1) and (2), it must escape this dilemma. In Section 4, I argue that the nomic necessity requirement is unnecessarily stringent. Section 5 discusses how Chalmers undermines (2) by creating a structure-sensitive, non-implementational Connectionist model.[18] This re-establishes Eliminative Connectionism as a legitimate theory of cognition. Connectionism being a legitimate theory of cognition further undermines (1). However, it does not undermine the LOTH.

Moreover, Rowlands shows (3) to be fallacious.[19] Section 6 follows their argument that logically structured representations do not follow from an isomorphism of the causal relations between representations and the logical relations between propositions. Although Chalmers and Rowlands succeed in undermining Fodor's arguments, I argue that they do not undermine the LOTH itself.

In Section 7, I provide two positive reasons for a language of thought, before concluding that the LOTH remains a legitimate cognitive theory.

First, some clarifications. For Connectionism to undermine Fodor's LOTH, the criticism must come from Connectionists; it must be about the nature of mental states and mental processes.[20] Both parties believe that representations exist and are physicalist about brain states — they believe states and processes of the mind to be identical to states and processes of the brain.[21][22] Whilst there are some Connectionists who deny representational states — such as Churchland[23] — the majority of debate assumes their existence. Thus, I will not discuss Eliminativism about representations or anti-realism about mental states. Further, Connectionist models match what we know about the neurological structure

12. Fodor 1975.
13. Fodor, 1987; Fodor and Pylyshyn, 1988.
14. Fodor 1987.
15. Michael Rescorla, "The Language of Thought Hypothesis," in *The Stanford Encyclopedia of Philosophy*, Summer 2019, ed. Edward N. Zalta (Metaphysics Research Lab, Stanford University, 2019).
16. Grant Sanderson, "But what is neural network? Chapter 1, Deep learning," Youtube, 2017, https://youtu.be/aircAruvnKk.
17. Fodor and Pylyshyn 1988.
18. Chalmers 1990, Chalmers 1992.
19. Rowlands 1994.
20. Fodor and Pylyshyn 1988, 3.
21. Fodor 1987, 282.
22. J. J. C. Smart, "The Mind/Brain Identity Theory," in *The Stanford Encyclopedia of Philosophy*, Spring 2017, ed. Edward N. Zalta (Metaphysics Research Lab, Stanford University, 2017).
23. Churchland 1990 as found in Rescorla.

of the brain, but Connectionism on a nonrepresentational level is not relevant.[24] It is possible for a brain to be neurologically Connectionist but implement classical representational architecture.[25] The only relevant Connectionism is at the representational level. Moreover, non-Fodorian LOT theories (such as Schneider's) are not discussed in detail.[26] I include Deep Learning[27] in my definition of Connectionism, but it is not relevant to my argument, so will not be examined in detail.

2 The LOTH and Fodor's Supporting Arguments

Fodor's hypothesis is that mental states are syntactically structured, and that mental processes are syntactical operations on mental states.[28] A state's structure determines its causal role in mental processes. Thinking occurs in a mental language. To have a belief that p is to bear an appropriate relation to a mental representation whose meaning is that p.[29] This mental representation takes the form of a sentence with combinatorial syntax and semantics. For example, to have the thought "I believe that X and Y" is to bear an appropriate relation to a complex mental representation whose meaning is that "X and Y". The complex representation gets its meaning from its constituents and how they are combined: from the meaning of its atomic constituents (X, Y) and from its syntactic parts (the conjunctive, and). As such, thought is combinatorial and structure sensitive.[30]

Distinct from the LOTH are the arguments that Fodor presents to support it. In The Language of Thought (1975), Fodor provides (1), which has widely become known as his 'only game in town' argument. He notes that our only remotely plausible cognitive theories of decision-making, concept learning and perception require a representational system to be coherent.[31] Representation presupposes a medium of representation, and a medium of representation requires symbolisation. Symbolisation requires symbols and thus a LOT.[32]

He supports this conclusion in *Why There Still Has to be a Language of Thought (1987)* with (2): our linguistic capacities are productive and systematic. Language is productive: we can conceivably say infinitely many unique and new thoughts, despite our finite physical resources.[33] This can be explained if thought is combinatorial — we can combine the constituents of sentences in as many ways as we would like. For example, *"I believe that it is very warm"* is distinct from *"I believe that it is very, very warm"* and *"I believe that it is very, very, very…ad infinitum…warm".* Even with the seven words used above, an infinite number of different mental sentences might be constructed. and our ability to understand some sentences means we understand others.

Productivity might be denied, since it requires idealisation – we never actually use any more than a finite part of any mental capacity, so our mental capacities might not necessarily be infinite. Fodor acknowledges this, but thinks idealisation is justified if it leads to independently well confirmed theories.[34] Systematicity, though, does not require idealisation, so this objection is not the focus of this essay.

Language is systematic: the ability to produce and comprehend some thoughts is intrinsically connected to the ability to produce and comprehend many other thoughts. If you understand the sentence *"Mary loves John"*, then you understand the sentence *"John loves Mary"*.[35] Our linguistic capacities are productive and systematic because they have combinatorial structure. Thought is also productive and systematic; this must be because it too has combinatorial structure. I will discuss how he uses (2) to try and undermine Connectionism later.

24. Cameron Buckner and James Garson, "Connectionism," in *The Stanford Encyclopedia of Philosophy*, Fall 2019, ed. Edward N. Zalta (Metaphysics Research Lab, Stanford University, 2019).
25. Fodor and Pylyshyn 1988, 6.
26. Susan Schneider, *The Language of Thought: A New Philosophical Direction* (MIT Press, 2011), Section 7.
27. Buckner and Garson.
28. Jerry A. Fodor, "Connectionism and Cognitive Architecture," Youtube, 2018, 10:49-11:02, https://youtu.be/vyrn1JWgqFA.
29. Rescorla.
30. Fodor and Pylyshyn 1988, 8.
31. Fodor 1987, 31 (decision-making); 36 (concept-learning); 51 (perception).
32. Fodor 1975, 55.
33. Fodor 1987, 292.
34. Fodor 1987, 293.
35. Fodor 1987, 294.

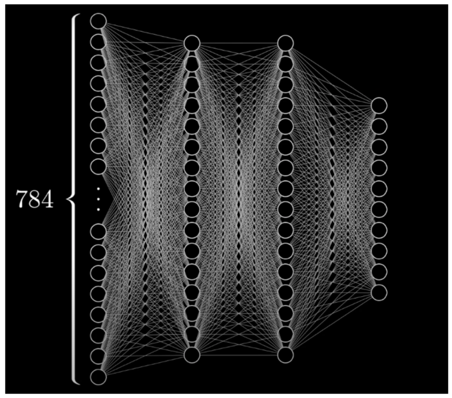

Figure 1: A Neural Network

Fodor presents (3) in Propositional Attitudes (1978) by pointing to the isomorphism of causal relations between representations and logical relations between propositions.[36] One can map the causal relations between representations onto a set of logical relations between propositions without losing the meaning of the representations. The propositions index the representations. From this isomorphism, he concludes that the representations must have logical form.

Fodor uses these three arguments to establish his LOTH. However, they are distinct from it – even if the arguments turn out to be invalid and/or unsound (as they do), this does not affect the truth of the LOTH. Analogously, I might present a fallacious argument that 'proves' that grass is green; the issues with my argument do not alter the fact that grass is green. I will show that Connectionism undermines (1)-(3), but it has not undermined the LOTH.

3 The Connectionist Challenge

Connectionism is not one idea or hypothesis, but a vast range of ideas. There is, though, a general form of Connectionism that threatens to undermine Fodor's. It offers an alternate theory of cognitive processing, using a different account of representation, mental states, and mental processes. Thus, it threatens Fodor's hypothesis that mental states are syntactically structured.

Connectionism claims that cognitive functioning can be explained by collections of units in a neural network. *Figure 1* provides a simple example of a neural network designed to learn to recognise hand-written digits.[37]

There are three types of units (or neurons): input units, hidden units, and output units. The units are organised into layers.[38] *Figure 1* has an input layer of 784 neurons, two hidden layers of 16 neurons and an output layer of 10 neurons.

The 784 neurons in the input layer (1st left to right in *Figure 2*) correspond to the 784 pixels on a 28x28 computer screen. When an image of a hand-written digit is on the screen, each pixel lights up in a certain way. Each neuron represents the greyscale value of its corresponding pixel as a number between 0 and 1. This is the neuron's activation value. The output layer has 10 neurons, representing the digits 0-9. The activation value of these neurons represents how much the system 'thinks' that a given image corresponds with a given digit.[39] Every neuron in the first hidden layer is connected to all 784 neurons from the input layer. Every neuron in the second hidden layer is connected to all 16 in the first, and each in the output layer connected to all 16 in the second.

36. As found in Rowlands 2019, 492.
37. Sanderson 2017a 03:47, based upon Nielsen 2015, 13.
38. Buckner and Garson.
39. Michael A. Nielsen, *Neural networks and deep learning*, vol. 25 (Determination press San Francisco, CA, USA, 2015), 14.

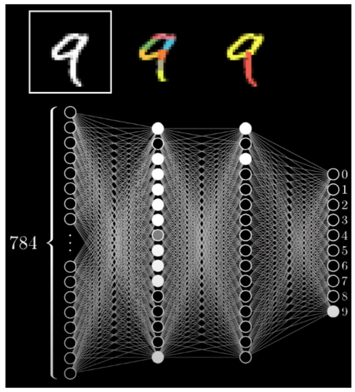

Figure 2: The Layers

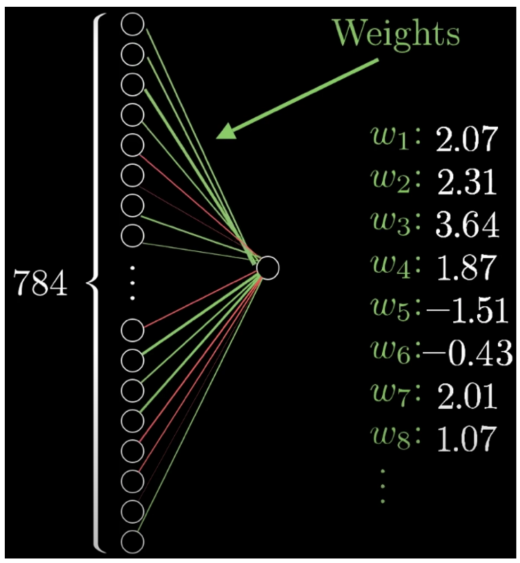

Figure 3: Weighted Connections

The 16 neurons in each of the hidden layers represent a subcomponent of a hand-written digit.[40] For example, the number 9 is a loop on top of a line; the number 8 is one loop on top of another. The rightmost hidden layer (3rd) represents these subcomponents. The leftmost hidden layer (2nd) represents subcomponents of these subcomponents. For example, a loop is composed of various small lines. *Figure 2* demonstrates what each layer might represent for an image of the number 9.[41]

Each neuron can be thought of as a function, taking the outputs of all the previous neurons, and producing a number between 0 and 1. In the first hidden layer, the activation value of each neuron is determined by the activation values of all 784 input units. As well as the activation value of each input unit, we also assign a value to the connection between x and each of the 784 neurons from the first layer, as shown in *Figure 3*.[42] This value is called the weight of a neuron's connection.

The network learns to recognise hand-written digits by finding the right weights and biases to produce the greatest activation value at the output unit for the correct digit. This process is loosely analogous to biological networks,

40. Nielsen, 11.
41. Grant Sanderson, "What is backpropagation really doing? Chapter 3, Deep learning," Youtube, 2017, 07:40, https://youtu.be/Ilg3gGewQ5U.
42. Nielsen, 14.

where neurons firing causes other neurons to fire.[43] Network learning methods generally fall under two categories: supervised and unsupervised.[44] The difference between the two is that supervised learning has an element of human oversight. Chalmers' model uses backpropagation, a supervised learning method. I will explain this process in Section 5, when discussing Chalmers' model.

If Connectionism is a legitimate theory of cognition, (1) seems undermined. Upon the relevant Connectionist models, representations are not operations over symbols. Information is stored non-symbolically in the weights of connections between the units in a neural net. There is another game in town. It designs systems that exhibit intelligent behaviour without retrieving or operating upon structured symbolic expressions.[45]

If Eliminative Connectionism is shown to be the correct theory of cognition, then the LOTH would be undermined. Eliminative Connectionism is Connectionism whose models do not implement a LOT. There are also Connectionist models that operate upon symbols.[46] These models implement the LOTH upon Connectionist hardware. Thus, even though Eliminative Connectionism undermines (1), the existence of Implementational Connectionism means the LOTH is not undermined. The existence of another possible explanation does not disprove the LOTH.

Nevertheless, in the 1980s, Connectionism looked to be a "paradigm shift" for cognitive theories – the LOTH seemed under threat.[47] Fodor and Pylyshyn responded to this threat.[48]

4 Fodor and Pylyshyn's Response: The Connectionist's Dilemma

F and P adapt the argument from systematicity and productivity (2) into an argument against Connectionism.[49] They argue that a theory of mental computation is explanatorily adequate only if it explains the nomic necessity of systematicity and productivity in thought. Symbol manipulation is the only way to explain the nomic necessity of systematicity and productivity. Whilst there might be Connectionist models that are systematic and productive (for example, ones that implement classical/LOT architectures), Connectionism does not require these qualities. This is an issue because systematicity is a necessary quality of thought. Thus, the Connectionist is presented with a dilemma: to endorse symbol manipulation, making Connectionism nothing but a way to implement a LOT, or to reject symbol manipulation. In rejecting it, they would be unable to explain the nomic necessity of systematicity and productivity; Connectionism would be an inadequate theory of cognition.

Accepting an implementational role is a real option for Connectionists in terms of accurately accounting for the nature of mental states and processes. However, Connectionism cannot undermine Fodor's LOTH if it implements it. Implementational theories – for example, Marcus, who argues for neural networks that implement symbol manipulation[50] – are not useful for undermining the LOTH. If the Connectionist cannot escape this dilemma, (1) is no longer undermined – the LOTH would be the only coherent option. Eliminative Connectionism must escape the charge of inadequacy by accounting for productivity and systematicity.

43. Sanderson 2017a, 04:32-05:00
44. Buckner and Garson.
45. Fodor and Pylyshyn 1988, 2.
46. Chalmers 1990, 341.
47. Fodor 1987, 82.
48. Fodor and Pylyshyn, 1988.
49. Reconstructed in Rescorla.
50. Gary F. Marcus, *The Algebraic Mind: Integrating Connectionism and Cognitive Science* (MIT Press, 2001).

5 Escaping the Dilemma

The dilemma presented by F and P is unnecessarily stringent by requiring systematicity and productivity as a nomic necessity. There are classical architectures that lack systematicity and productivity.[51] Therefore systematicity and productivity cannot be a nomic necessity for classical architectures. By their own requirement, then, the LOT would be an inadequate theory of cognition. Thus, it seems fair to drop the requirement that productivity and systematicity be necessary. Connectionist models still need to be productive and systematic, though.

F and P's paper roused numerous responses (Clark 1989; Smolensky 1987, 1990; van Gelder 1990; Elman 1990; Pollack 1990).[52] For this question, the only useful response to F and P is to produce an Eliminativist Connectionist model that explains systematicity and productivity. Therefore, I focus on Chalmers, who produces a non-classical model of cognition that operates structure-sensitive processes.[53] Others have also attempted this – notably, Pollack (1988, 1990) and Smolensky (1987, 1990).[54] However, they provide productive, systematic models by extracting the original constituents of a representation. This effectively renders their models implementational and, as such, not useful for undermining the LOTH.[55]

The success of Chalmers' model would undermine (1). It would also undermine (2) in its employment as a counterargument against Connectionism.[56]

6 Chalmers' Eliminativist Connectionist Model

Chalmers argues against F and P's claim that Connectionist models cannot support systematic operations in a non-classical way. He claims that F and P misrepresent the Connectionist endeavour and underestimate the difference between localist and distributed representations.[57]

When describing Connectionism, F and P provide an example of a localist Connectionist model where atomic symbols are represented by single nodes, connected by associative links.[58] They briefly assert that it would change nothing if these nodes were replaced by a distributed pattern of activation.[59] Chalmers, though, demonstrates that there is a clear and significant difference between localist and distributed representation. In a localist model, nodes represent atomic symbols. They are connected with associative links. F and P representing Connectionism in this way is problematic – many Connectionists define themselves on attempting to do away with the atomic symbol in theories of meaning.[60] Distributed models do not use atomic symbols. They have groups of separately functioning nodes that have functional properties far beyond that of an isolated unit. Representation does not occur at the level of the node, but at a much higher level. Information is stored in the activation values of nodes and weights of connections. At that higher level, patterns of activation between nodes combine compositionally and autonomously to produce distributed, malleable representations. Thus, a small difference in the activity of a subset of nodes can cause substantial differences in later processing.[61]

F and P claim that Eliminativist Connectionist models cannot account for productivity or systematicity. In other words, they cannot carry out structure-sensitive operations. Responding to this criticism is paramount for Connectionism to be able to undermine Fodor's LOTH. Chalmers responds by undergoing a series of experiments, in which

51. Buckner and Garson.
52. Found in Chalmers 1990, 340, and Terence E. Horgan and John L. Tienson, *Connectionism and the Philosophy of Mind* (Kluwer Academic Publishers, 1991)
53. Chalmers 1990, 1992.
54. Found in Chalmers 1990, 344.
55. Chalmers 1990, 340.
56. (2) will be fully undermined in section 5
57. Chalmers 1990, 340.
58. Fodor and Pylyshyn 1998, 10.
59. Fodor and Pylyshyn 1998, 15.
60. Chalmers 1990, 343.
61. Chalmers 1990, 343.

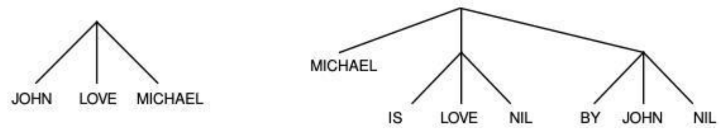

Figure 4: Sentence Representations

he demonstrated that structure-sensitive operations are possible upon distributed representations.[62]

The experiments looked to use distributed representations to transform sentences from their active to their passive forms. Chalmers combined five different names and verbs to produce 250 different sentences, all of similar syntactic form to the active "John loves Michael", or the passive "Michael is loved by John". He used Pollack's Recursive Auto-Associative Memory (RAAM),[63] a system that recursively encodes symbolic tree structured representations of sentences in distributed form.[64] *Figure 4* shows how the sentences were given syntactic structure.

RAAM uses backpropagation to create patterns for each of the internal nodes of the trees. Backpropagation is a form of supervised learning. Whilst untrained, a neural net might produce activation values at the output nodes that are wildly inaccurate. Backpropagation is an algorithm that computes a list of changes required to the weights and biases to produce the correct results.[65] It compares the net's outputs to the 'correct' outputs provided by a training data set and works backwards, seeing how the weights and biases of the connections from hidden and input layers have led to the 'incorrect' values of the output nodes. Over many training cycles, backpropagation fine-tunes the weights and biases of the connections between nodes, until the network produces the 'correct' outputs. As the theory goes, during this process the network generalises the syntactic rules of the operations it learns.

Chalmers' RAAM assigned each word a primitive localist representation and learned to represent all 250 sentences. Once encoded in the RAAM, the representation is considered distributed. 150 of these encodings were randomly selected to train the Transformation Network, another backpropagation network. The Network was to take an encoded distributed representation of an active sentence as input and transform it into the appropriate encoded distributed representation of a passive sentence as output. The RAAM then decoded the represented output sentences. To truly see whether the network was structure-sensitive, it needed to successfully operate on sentences outside of its training data set – this would test whether the network had generalised the syntactical rules it was being fed.

Thus, after the Transformation Network was trained, the RAAM encoded the other 50 active sentences, fed them through the Transformation Network, and then decoded the Transformation Network's output pattern. In all 50 cases, the output pattern decoded to the correct passivized sentence; generalisation rate was 100 per cent. As noted by Chalmers (1990), this shows that distributed representations formed by RAAM can effectively facilitate structure-sensitive operations in a non-classical way.[66] If Chalmers' model can account for structure-sensitive operations, then it can explain why understanding the sentence "Mary loves John" entails understanding "John loves Mary". It can account for systematicity and, it follows, productivity. Thus, if Chalmers' model is satisfactory, then F and P's dilemma is escaped: eliminativist Connectionism is explanatory of systematicity and productivity. (1) and (2) are undermined.

The effectiveness of RAAM and Backpropagation Models, though, is questionable. Buckner and Garson note that such models fall short when they are applied to truly novel sentences.[67] Marcus argues that multilayer perceptron approaches that backpropagation cannot capture the flexibility and power of everyday reasoning.[68]

Debate regarding the fine-tunings and legitimacy of RAAM and Backpropagation Models is beyond the scope of this essay. What is important, though, is not that Chalmers' model is perfect, but that it is another game in town.

62. Chalmers 1990, 1992.
63. Chalmers 1990, 5.
64. Seth Rait, "DRAAM: Deep (Recursive Auto-Associative Memory) And Applied eMbeddings," Undergraduate Honors Thesis (Undergraduate Honors Thesis, Brandeis University, 2018), 12.
65. Sanderson, "But what is neural network? Chapter 1, Deep learning."
66. Chalmers 1992.
67. Buckner and Garson.
68. Marcus, 169.

There are many other Eliminativist Connectionist models that have also been purported to explain systematicity and productivity of thought. Moreover, new research into Deep Learning has opened up the opportunity for further discovery down the line. Rait (2018) argues that a new Deep Learning RAAM (DRAAM) could provide novel insights into cognitive processing, since the main issue with RAAM at Pollack's time of writing was its technical limitations.[69] Loula, Barni and Lake (2018) report that their nets qualified as demonstrating strong semantic systematicity.[70] Whether any of these models are without criticism is less important for this essay – the nature of mental representations is still an undetermined issue, but Eliminativist Connectionism remains a game in town.

(1) is certainly undermined. (2) might still stand. Whilst Eliminative Connectionism is a legitimate prospect, systematicity and productivity might be evidence of the syntactic structure of thought. This still does not undermine Fodor's LOTH. Connectionism provides an accurate neurological account of the brain which should be adopted, but this does not preclude the possibility of Connectionist models implementing the LOTH. The LOTH is one of the two current theories of high-level cognition that might be true. I now turn to Rowlands, who attempts to directly undermine Fodor's reasoning behind the LOTH.[71] I argue that Rowlands' successfully shows (2) and (3) to be fallacious. However, he also fails to undermine the LOTH itself.

7 Rowlands' Critique of the LOTH

The LOTH makes two distinct claims:

(C1) Mental representations are structured entities

(C2) Mental representations have the structure of propositions or sentences.[72]

Rowlands argues that all the arguments that Fodor provides to support and motivate the LOTH are based upon a fallacious conflation of (C1) and (C2). Fodor assumes that the arguments prove (C2) when they only prove (C1). This fallacy is clear in the argument from productivity and systematicity (2), and the argument from isomorphism (3). Rowlands focusses on (3).

As mentioned earlier, (3) points to the isomorphism of causal relations between representations and logical relations between propositions.[73] One can map the causal relations between representations onto a set of logical relations between propositions without losing the structure of the representations. The propositions index the representations. Fodor argues that this structure-preserving mapping must be specifiable in terms of the logical form of propositions and concludes that the objects of these attitude must have logical form. Rowlands shows this inference to be a fallacy using the analogy of a painting[74]:

A painting has many features. Conceivably, all of its features can be put into a structure-preserving representation theorem that maps the features of the picture onto a set of propositions. This set of propositions would preserve the structure of the painting. We could therefore say that the propositions index the features of the picture, in just the same way that the propositions index the causal relations between representations. Nobody, though, would imply that a painting has logical, syntactical structure, merely because we can map its features on a set of propositions. It is possible to make any system isomorphic with another if you find the right mathematical function. Isomorphism does suggest (C1) – the representations do have structure. However, it does not suggest (C2) – isomorphism does not entail logic structure.

(2) commits the same fallacy.[75] It points to certain features of natural language that are mirrored in thought (productivity and systematicity) and argues that this mirroring must be because mental representations have the same

69. Rait, 2.
70. As found in Buckner and Garson.
71. Rowlands.
72. Rowlands, 489.
73. Rowlands, 493-494.
74. Rowlands, 491.
75. Rowlands, 494.

structure as propositions. Fodor is aware of this: after re-stating (2) in *Why There Still Has to Be A Language of Thought*, he notes that one might accuse him of affirming of the consequent; he waves it off as inference to the best explanation.[76] I argue that this dismissal of fallacy would be reasonable if (1) still stood. If the LOTH was the only game in town, then noting that thought has a shared property with language and concluding that this is because they have the same structure would be a reasonable inference to make. However, it has been shown that the LOTH is not the only game in town. Thus, arguments (2) and (3) unjustly make this inference. Fodor's arguments are demonstrably undermined.

8 Arguments in Favour of a LOTH

The purpose of this section is not to present a conclusive argument in favour of Fodor's LOTH, nor one against Connectionism. Instead, it looks to present two positive reasons for maintaining the language of thought as a legitimate cognitive theory.[77]

8.1 Brains and Neural Networks learn differently

The way neural networks learn is only loosely analogous to biological networks in the brain; there are significant differences between the two. The example in Section 2 used the MNIST data set, a data set of scanned images of handwritten digits created by the United States' National Institute of Standards and Technology.[78] The MNIST data set contains 60,000 images as its training data, and 10,000 images as test data.[79] Other networks use billions of training examples.[80] Human minds do not learn in this way – we do not consult billions of pieces of training data before being able to recognise hand-written digits. It would not physically be possible for brains to do so.

Chomsky's Poverty of the Stimulus argument states that children learn with far less data than something like a neural net requires.[81] Auxiliary verbs are a classic example of this: they are highly syntactically complex, and their employment in language often defies generalisation.[82] There are 1x1022 combinations of English auxiliary verbs, yet only 99 grammatically possible combinations.[83] A human brain could not run through a data set that large to learn the complex rules of auxiliary verbs, as 1x1022 is roughly one hundred billion times the number of neurons in the human brain.[84] Nevertheless, children consistently differentiate between auxiliary and lexical verbs without issue.[85] Despite achieving the same competence as a human brain in specific tasks, neural networks compute in a fundamentally different way.

It is worth noting that Chomsky's Poverty of the Stimulus argument is philosophically controversial, and the debate surrounding it is well beyond the scope of this essay.[86] That debate is further muddled by Fodor's belief that the language of thought is innate.[87]Nevertheless, an important point has been raised. Much of the "paradigm shift"[88] away from the LOT occurred because Connectionism was able to produce models that mirrored biological processes; there are significant differences between human learning and machine learning that Connectionism is yet to reconcile.

76. Fodor 1987, 293.
77. For a series of criticisms of Connectionism, see Marcus.
78. Nielsen, 15.
79. Nielsen, 16.
80. Nielsen, 2.
81. Noam Chomsky, *Poverty of Stimulus: Unfinished Business*, March, 5.
82. Stephen Laurence and Eric Margolis, "The Poverty of the Stimulus Argument," *British Journal for the Philosophy of Science* 52, no. 2 (2001): 226, https://doi.org/10.1093/bjps/52.2.217.
83. Stromswold as found in Laurence and Margolis, 224.
84. Laurence and Margolis, 224.
85. Laurence and Margolis.
86. Laurence and Margolis.
87. Laurence and Margolis, 240.
88. Schneider, *The Language of Thought: A New Philosophical Direction*, 82.

8.2 The LOTH is still pursued

Whilst this essay is not an examination of non-Fodorian language of thought theories, it is worth mentioning that modern studies continue to build upon the belief that thought is logically structured. A language of thought accounts for the productivity and systematicity of thought in simple terms. Symbol-manipulation still provides the best explanation for high-level cognitive phenomena.[89] Further, modern LOT theories, such as Schneider's, offer the LOT a "philosophical overhaul" to separate Fodor's hypothesis from his fallacious arguments, keeping symbol manipulation at the forefront of cognitive theories.[90]

Implementational Connectionism – Connectionism that endorses symbol manipulation and implements classical architectures – remains a genuine field of philosophical inquiry.[91] Schneider also notes that information-processing psychology predominantly operates with symbol-processing models.[92] The LOT is still considered a relevant theory of cognition and is being used to further our understanding of the mind and brain.

9 Conclusion: Fodor's LOTH is not undermined

I have discussed three of Fodor's arguments in favour of the LOTH. Chalmers and Rowlands have convincingly demonstrated each to be unconvincing. However, none of this undermines Fodor's LOTH. Chalmers and Rowlands have showed that Fodor's arguments do not prove that thought is logically structured. This is not the same as showing that thought cannot be logically structured. There are positive reasons to uphold the LOTH.

This question is unfairly stacked against Fodor – his work is singular and complete, whereas Connectionism is a dynamic field of advancing scientific understanding. Thus, there will inevitably come a point where Fodor's writing seems dated and out of touch with modern science. Already, Connectionism acts as a strong non-representational framework within which to build theories of high-level cognition. However, Fodor's hypothesis is yet to be disproved.

It is entirely possible that some future scientific discovery – perhaps in *Deep Learning* models – proves Eliminativist Connectionism to be the only satisfactory theory of cognition and representation. Equally, some breakthrough might champion Implementational Connectionism. Neither has happened yet. Despite the flaws in his arguments, Fodor's LOTH is still a game in town; the LOTH is still being pursued. Thus, Connectionism does not undermine Fodor's Language of Thought Hypothesis. Equally, Fodor's *Language of Thought Hypothesis* does not undermine Eliminative Connectionism. At least not yet.

References

Buckner, Cameron, and James Garson. "Connectionism." In *The Stanford Encyclopedia of Philosophy*, Fall 2019, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University, 2019.

Chalmers, D. "Why Fodor and Pylyshyn Were Wrong: The Simplest Refutation." 1990.

Chalmers, David J. "Syntactic transformations on distributed representations." In *Connectionist natural language processing*, 46–55. Springer, 1992.

Chomsky, Noam. *Poverty of Stimulus: Unfinished Business*, March.

Fodor, J. A. "The Language of Thought." *Critica* 10, no. 28 (1978): 140–143.

89. Marcus, 170.
90. Schneider, 5.
91. Chalmers 344 gives Pollack (1988, 1990) and Smolensky (1987, 1990) as examples
92. Schneider, 4.

Fodor, Jerry A. "Connectionism and Cognitive Architecture." Youtube, 2018. https://youtu.be/vyrn1JWgqFA.

———. *The Language of Thought*. Harvard University Press, 1975.

———. "Why There Still has to Be a Language of Thought." In *Psychosemantics*. MIT Press, 1987.

Fodor, Jerry A., and Zenon W. Pylyshyn. "Connectionism and Cognitive Architecture: A Critical Analysis." *Cognition* 28, nos. 1-2 (1988): 3–71. https://doi.org/10.1016/0010-0277(88)90031-5.

Horgan, Terence E., and John L. Tienson. *Connectionism and the Philosophy of Mind*. Kluwer Academic Publishers, 1991.

Laurence, Stephen, and Eric Margolis. "The Poverty of the Stimulus Argument." *British Journal for the Philosophy of Science* 52, no. 2 (2001): 217–276. https://doi.org/10.1093/bjps/52.2.217.

Marcus, Gary F. *The Algebraic Mind: Integrating Connectionism and Cognitive Science*. MIT Press, 2001.

Nielsen, Michael A. *Neural networks and deep learning*. Vol. 25. Determination press San Francisco, CA, USA, 2015.

Rait, Seth. "DRAAM: Deep (Recursive Auto-Associative Memory) And Applied eMbeddings." Undergraduate Honors Thesis. Undergraduate Honors Thesis, Brandeis University, 2018.

Rescorla, Michael. "The Language of Thought Hypothesis." In *The Stanford Encyclopedia of Philosophy*, Summer 2019, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University, 2019.

Rowlands, Mark. "Connectionism and the Language of Thought." *British Journal for the Philosophy of Science* 45, no. 2 (1994): 485–503. https://doi.org/10.1093/bjps/45.2.485.

Sanderson, Grant. "But what is neural network? Chapter 1, Deep learning." Youtube, 2017. https://youtu.be/aircAruvnKk.

———. "What is backpropagation really doing? Chapter 3, Deep learning." Youtube, 2017. https://youtu.be/Ilg3gGewQ5U.

Schneider, Susan. *The Language of Thought: A New Philosophical Direction*. MIT Press, 2011.

Smart, J. J. C. "The Mind/Brain Identity Theory." In *The Stanford Encyclopedia of Philosophy*, Spring 2017, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University, 2017.