

Cognitive Science: Essay 1

Oliver Lack a1809024

Provide an overview of connectionism. What are its main strengths and weaknesses? Is it a plausible approach to explaining cognition?

The purpose of this essay is to provide an overview of connectionism and show it as a plausible approach to explaining human cognition. Firstly, I will address the popular alternative inspired by Alan Turing's work, the classical computational theory of mind; giving an overview of the computational and representational differences between connectionist and classical architecture. Secondly, I will focus on the biological plausibility of connectionist models with regard to parallel distributed processing and representation. Furthermore, I will contrast the similarities between properties of human cognition and connectionist models. To conclude, I will address the argument for systematicity of thought by Fodor and Pylyshyn.

Connectionism is an approach to explaining human cognition that provides a promising alternative to the classical computational theory of mind. Alan Turing's description of an abstract automatic machine was the major influence on classical computational theory. His machine arranged physical properties of symbols in a semantically coherent form. These symbols are manipulated in regard to their syntactic structure, but with respect to their semantic properties. In other words, "if you take care of the syntax, then the semantics will take care of itself" (Haugeland, 1985, p. 106) Turing's answer for mechanising computation was to make a physical machine mimic the behaviour of semantic one. Based on the Turing machine, the conventional digital computer partitions a continuous variable into discrete syntactical binary states (high and low voltage states). These syntactical states are given semantic properties to create *symbols* (discrete syntactical units used as vehicles for representations). Well-designed rules (programmed software) sensitive only to the syntactic properties (physical structure) of these symbols manipulate them to produce results with respect to their semantic properties. Causal work is only computationally performed on the syntactic properties of these symbols; the rules are designed to keep syntactic processing consistent with intended semantic interpretation (O'Brien, 1999, pg. 5). The classical theory applies this method to the mind, suggesting cognitive processors are a form of symbol manipulation. Connectionist architecture differs computationally and representationally from the once popular classical computational theory of mind.

Connectionist models use networks comprised of a large number of connected nodes capable of being excited or inhibited by input signals. These nodes are arranged into layers, usually consisting of input and output layers mediated by multiple hidden layers. Each node has a weighted connection to each node in the following layer (Fig. 1). The weight of each individual connection can increase or decrease the strength of the signal and is given by a real number, adjustable by the system. These connections are used for transferring and altering the signal as it is passed through the network. Additionally, each node has a minimum threshold at which it will activate when given sufficient input from the sum of all attached connections. Once activated, the node will output a signal to each node in the following layer (Crane, 2016, pg. 106). The input signals in the first layer of the network cause a pattern of activation to flow through the subsequent layers of the network. The activation of each layer is passed through the network until the output signal activation is stabilised at the output layer. (Sterelny, 1991, pg. 171).

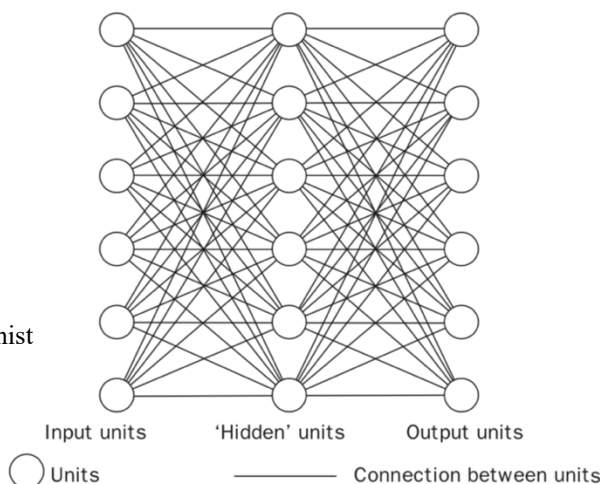


Fig. 1 Diagram of a connectionist model (Crane, 2016, pg. 107)

Connectionist models use a different form of processing, distinct from the serial processing in a conventional computer. These models use parallel distributed processing, which spreads the input activation through all the nodes in each layer; processing information simultaneously rather than serially from one node to the next. Parallel distributed processing allows a large number of individual computations simultaneously; making it much faster than a conventional computer at tasks confronted by living organisms (Churchland, 1995, pg. 12). Compared to serial processing, models that use parallel distributed processing excel at quickly accessing relevant information from large information stores and recognising patterns (for example in words, objects, faces and scenes) from degraded information (Sterelny, 1991, pg. 173). Human cognition seems to have a strong similarity to the strengths and strategies used by parallel distributed processing. Sterelny argues that humans are good at what serial computers are bad at, and vice versa (Sterelny, 1991, pg. 173). The lack of similarity to serial computation gives good grounds to suspect parallel distributed processing as a more plausible approach to explaining human cognition.

For connectionist models to be held as biologically plausible explanations of cognition we must assume all cognitive processors are performed in the brain; if the brain consists of neural networks, all cognitive processors must use neural networks. Neuroscience suggests that the brain is comprised of multiple parallel processing systems built from networks of neurons (Churchland, 1993, pg. 139-142). Connectionist networks reflect the properties of the brain through their resemblance of nodes to neurons and weighted connections to synapses (Sterelny, 1991, pg. 175). The abilities and limitations of neurons can be accounted for in connectionist networks. Although, Crane would argue that nodes in a connectionist network have none of the complexity of real neurons (Crane, 2016, pg. 107). Others would argue that connectionist architecture bares compelling similarity to the complexities of biological properties. For instance, Sterelny outlines the '*100 step*' argument by comparing the slow speed of neural processing to the exceptionally faster serial processing in a conventional digital computer. He concludes that due to the processing speed limitations of a neuron and the ability of parallel processing to simultaneously execute multiple computations, human cognition must use parallel processing; precisely the processing that connectionism provides (Sterelny, 1991, pg. 172). Fodor illustrates that classical architecture does not exclude parallel execution of multiple symbolic processors; suggesting that an argument for parallel processing is not an argument against classicism or for connectionism (Fodor, 1988, pg. 56). Therefore, connectionism only succeeds over classical models of cognition based on serial processing. Nonetheless, connectionist models are fundamentally based on parallel distributed processing, allowing them to reflect properties of neural network structures and neuronal speed limitations.

From Turing's perspective, computation may be defined as a process that produces a semantically coherent sequence of symbols. However, there seems no reason to restrict computation to symbol manipulation. O'Brien suggests that what is essential to computation is not symbol manipulation, but the use of representational vehicles (O'Brien, 1999, pg. 3). The classical theory rejects this, claiming that human thought is the result of syntactically driven computation of neural symbols; it assumes that the localised symbols used in cognitive processing are also the vehicles of mental representation. Connectionism does not assume that these localised symbols of computation are the individual vehicles of mental representation (Cummins, 1989, pg. 157). Rather, the connectionist claims that the vehicle of mental representation is the state of the network's activation as a whole; the collective activation pattern of all nodes in the network contribute to the mental representation. An item of information expressed across the activity of multiple nodes is known as distributed representation (Clark, 2001, pg. 66). Information storage and processing are not separated as a consequence of distributed representation, unlike digital computation which separates storage from processing. The storage of information is spread across the system, allowing multiple representations to be encoded over the same network connections and nodes. Clark calls this overlapping of information '*superpositional storage*' (Clark, 1993, pg. 192). When one extracts information from the network, many connection weights and nodes play a role in the retrieval and processing of the information (Ramsey, Stich, and Garon, 1991, p. 212).

Connectionism's main strengths are due to its distinct position on representation and superpositional storage. Distributed representation allows connectionist models to remain relatively functional when attaining a loss or damage to parts of the network. Additionally, they have the ability to accurately generalise novel information. Because information storage is superpositional and the processing of information is distributed, this allows the system to act as a pattern completer (Clark, 2001, pg. 67). Given degraded or novel information, a connectionist network has the ability to recognize patterns and produce semantically coherent outputs fairly accurately. This has compelling similarity to the strengths of human cognition and the weaknesses of machines that use symbol manipulation. Humans are exceedingly good at pattern recognition and generalisation of novel information (Sterelny, 1991, pg. 173). For example, one's ability to recognize distorted audio such as an unfamiliar accent or distorted visual stimulus such as an unknown face from across the room. Humans have an inferential ability to generalise different stimulus that conventional computers do not. When faced with damages or imperfections to the network or input, machines that use symbol manipulation tend to fail or completely crash (Sterelny, 1991, pg. 173). Additionally, humans can suffer damage to parts of the brain with only gradual degradation to cognitive ability. For instance, Fodor explains that damage to the temporal cortex does not result in a loss of specific memories (Fodor, 1988, pg. 52); Sterelny states that damage to language centres do not lead to a complete loss of linguistic abilities, rather a gradual degradation of ability dependent on the severity of damage (Sterelny, 1991, pg. 175). These examples suggest that memories and other cognitive abilities are distributed among the entire network and degrade conditional to the damage. This *graceful degradation* of ability is also seen in connectionist models when parts of the network are damaged (Clark, 2001, pg. 66). If processing and storage is distributed throughout the brain's neural network, connectionist models may be a plausible approach to explaining the flexibility and pattern recognition abilities of cognition.

Fodor and Pylyshyn developed one of the main arguments against connectionism, suggesting connectionist models lack the systematicity essential to thought. Examples of the systematic nature of thought put forward in the argument use language comprehension and production. For instance, the ability to have the thought 'Keanu is jealous of Camu' is intrinsically connected to one's ability to have the thought 'Camu is jealous of Keanu'. The syntactic properties of the sentences are different, yet they are semantically related. Fodor and Pylyshyn would argue that symbol manipulation is required to explain the ability to have semantically related thoughts. Symbol manipulation can explain systematicity by assuming the parts of a thought are split into separate symbols (e.g. 'Keanu', 'Camu' and 'jealous'). The parts comprising the first thought are the same parts used in the second thought. The rearrangement of these parts allows one to think many other semantically related thoughts (Fodor, 1988, pg. 37-41). Connectionism uses distributed representation and lacks compositional syntax and semantics. Therefore, it may be unable to explain the systematicity of thought. However, connectionist models have been offered that may allow for systematicity (Van Gelder, 1990; Smolensky, 1987). Furthermore, Sterelny doubts other animals have systematicity. For example, a chimpanzee's inability to think of a dangerous banana that may eat him, despite possessing each individual part of the thought (Sterelny, 1991, pg. 183). Dennett and Clark share this doubt, supporting the exclusivity of systematicity in human thought. Additionally, they propose that language has allowed humans to separate form from content; giving the ability to have open ended, semantically related thoughts (Clark, 2001, pg. 78; Dennett, 1991, pg. 27). This counterargument suggests that the systematicity of thought is a consequence of language skills and not cognitive architecture (shared with non-linguistic animals). The fitting responses to Fodor's problem of systematicity encourages many to retain connectionism as a plausible theory of cognition. Although, few may find the weakness enough to abandon connectionist architecture altogether.

Connectionism displays strengths that resemble cognitive abilities such as pattern recognition, flexibility and graceful degradation. This resemblance is a result of the network's distributed representation and superpositional storage. Connectionism seems to account for some of the weaknesses held by the classical computational theory and arguably possesses a closer similarity to

cognition, despite its weaknesses. Systematicity of thought remains a weakness to some. Although, strong responses seem to prevent abandonment of the theory and motivate explanation for the role of language in the systematicity of thought. Connectionism is still a plausible approach to human cognition with significant resemblance to abilities and properties of neural networks.

References

- Churchland, P. 1995, *The Engine of Reason, The Seat of the Soul*
- Clark, A. 1993, *Associative Engines*
- Clark, A. 2001, *Mindware*
- Crane, T. 2016, *The Mechanical Mind*, 3rd Edition
- Cummins, R. 1989, *Meaning and Mental Representation*
- Cummins, R. 2010, *Systematicity and the Cognition of Structured Domains*
- Dennett, D. 1991, *Consciousness Explained*
- Fodor J, Pylyshyn Z. 1988, *Connectionism and cognitive architecture Cognition*
- Haugeland, J. 1985, *Artificial Intelligence: The Very Idea*. Cambridge
- O'Brien, G. 1999, 'Connectionism, Analogicity and Mental Content', *Acta Analytica*, Vol. 22, pg. 111-131
- Smolensky, P 1987, 'The Constituent Structure of Connectionist Mental States: A Reply to Fodor and Pylyshyn', *The Southern Journal of Philosophy*, Vol. 26, pg. 137-161
- Ramsey W, Stich S, Garon J. 1991, *Connectionism, eliminativism, and the future of folk psychology*
- Sterelny, K. 1991, *The Representational Theory of Mind*
- Van Gelder, T. 1990, 'Compositionality: A Connectionist Variation on a Classical Theme', *Cognitive Science*, Vol. 14, pg. 355-384