Introduction

Very often we naturally assume that our brains think the same way we speak – that is, that our thoughts and ideas can each be broken down and mapped to a nice set of internal symbols, similar to words in spoken language. However, in some cases – especially when it comes to internal representation of what we perceive – this explanation falls short. It turns out that the way we store something such as a visual image is much better understood by comparing it to the way information is stored in an intricately layered road map, in which no piece of information can be removed without affecting nearby pieces, than to a set of words and phrases in a spoken or written language, in which each element is for the most part independent of the others. This concept of a map-like representation of perception is known as connectionism. Connectionism offers what seems to be a much more accurate model for perceptual representation than that which is put forth by symbolic architecture.

Background

In general, the connectionist model is more compatible with the biological reality of the brain than is symbolic architecture. The connectionist model incorporates what is known as a neural network, which consists, basically, of multiple interconnected nodes (points at which data is manipulated). The existence and strengths of the connections between these nodes, as well as the functions contained within each node, determine what the output will be for a given set of input. When given a set of inputs and expected associated outputs, the network can be set up to "train" itself – that is, nodes and their connections will change to fit the data they are given. Neural networks do not involve symbol manipulation.

Symbolic architecture holds that there exists within the brain a sort-of neural language, which represents everything in a manner similar to the way we use spoken or written language.

In symbolic architecture, information, thoughts, etc. are stored as individual units (symbols) which are then manipulated / transformed using stored rules.

It turns out, based on our knowledge of the inner-workings of the brain, that connectionism produces what seems to be a much more accurate model than that afforded by symbolic architecture. Brains consist of interconnected neurons that rest or fire depending on the varying levels of activation or inhibition that they experience as a result of the actions of the neurons that they are connected to. The properties of the neuron, the transmission medium between it and nearby neurons, and the status of those nearby neurons are all contributing factors to the neuron's tendency to fire or rest.

It is only reasonable that network models would more closely resemble the brain since the models were, in fact, derived from a simplified interpretation of the way the brain works. The functionality of nodes was inspired by the functionality of the neuron, and the connections between nodes came from a modeling of the neuron's axons and dendrites. The connectionist model is, however, still much simpler than the true architecture of the brain in that it does not simulate such things as the differences between neurotransmitters. (Bechtel & Abrahamsen, 2002, p. 45)

In the brain, simple processes (e.g. word recognition) are executed very rapidly, usually in only a few tenths of a second. Because it takes several milliseconds for a neuron to fire, it can be argued that simple tasks would need to perform, at most, somewhere around a hundred sequential operations. Unfortunately, symbolic architectures have a hard time being compatible with this, as then tend to require thousands of operations for the same simple tasks. However, since network processing allows for the execution of numerous operations in parallel, they tend to satisfy the hundred-sequential-operations limit much more often. (Bechtel & Abrahamsen, 2002, p. 45) While structural similarity does not inherently guarantee functional similarity, it can be a factor, as seems to be the case for connectionist models.

Graceful Degradation

The human brain is extremely reliable, but it can easily become overloaded with excessive tasks or information. When this happens, instead of giving up altogether or completely eliminating a large piece of information, it experiences what is known as graceful degradation. Basically, this means that it gradually begins to do a less-than optimal job – if given too many requirements or too much data it will begin to ignore a few small pieces of information (e.g. minor details in a remembered image) throughout the brain. This is also what usually occurs when a portion of the brain is damaged, destroyed, or removed, although occasionally a noticeably distinct / encapsulated deficiency will appear. That is to say, small (and sometimes even moderately large) amounts of damage do not seem to have any noticeable affect on the brain's performance. (Bechtel & Abrahamsen, 2002, p. 48)

Symbolic architectures do not experience graceful degradation. If a rule is lost, the system cannot respond at all to any situation which would have employed that rule. Of course, redundant data and rule storage could help to alleviate this problem, but would likely still fall short of displaying graceful degradation. Connectionist architecture, on the other hand, because of its lack of rules and symbol manipulation, does not experience the same difficulties. When a node or connection is removed the network does indeed become less accurate, but as long as the number of nodes and connections is large, this drop in accuracy is not likely to dramatically alter any given output, but will have its effect spread out over the whole network to a very slight degree. In this way, the network resembles graceful degradation much more closely than symbolic architecture. (Bechtel & Abrahamsen, 2002, p. 48)

Conceptually, the difference is because connectionist networks have many small factors that contribute to any piece of information. Altering only a few of those factors (destroying a few connections) would cause the information to become less accurate, but still somewhat useful. In symbolic architecture, each piece of information and each rule exists as a single unit. If that unit is destroyed, all the information it contained is lost. The symbolic architecture is, in a sense, all-or-nothing, while the connectionist network can experience graduated decline in information accuracy / performance – graceful degradation.

In the case of visual perception, we often retain an image of what we have seen, though over time the details that we can recall from that image become fewer and fewer – the memory becomes “fuzzier”. What is happening is that this visual image is encoded globally throughout a large region of the brain – that is, the nodes and connections that make up a representation are physically spread out over a large portion of the brain (Ishai, Ungerleider, et. al., 1999); when neurons are removed from this distribution (e.g. when they die off), the system still functions normally for the most part. (Merkle, 2001, p. 18) This is consistent with an understanding of the brain’s visual representation as a network that experiences graceful degradation.

Content-Addressable Memory / Pattern Reconstruction

When recalling a specific situation, we can often use any of many cues to reconstruct the whole thing in our mind. For example, we can call to mind a specific portion of a movie by multiple things associated with it, such as the music, the actions of the characters, the conversation, etc. The whole experience can be called to mind from any one of those cues. Often, a cue can help us to reconstruct a piece of information that we were unable to recall before – we are able to complete the pattern. This concept is known as “content-addressable memory.”

In a symbolic architecture, these would be very difficult tasks to accomplish. To recall information from a given cue, it would be necessary to know in advance that the information would need to be indexed by that cue. It would then be necessary to specify ahead of time all the ways that the information could be accessed. (Bechtel & Abrahamsen, 2002, p. 49) On the other hand, because of the nature of a connectionist network, such information-retrieval ability is built-in. All the connections already exist. That is, it is just as easy to start at node A (a set of names) and see what associations lead to node B (foods liked) (index by name) as it is to start at node B and see what associations lead to node A (index by food liked). In a symbolic architecture, the information would have to be indexed separately for every situation in which information would need to be retrieved. In fact, to create all possible indexes would be to create a connectionist network of sorts.

In fact, it is sometimes possible in a network architecture to reconstruct a whole pattern based on a part of it. In this case, the network is determining which possible missing piece best fits with the information in the portion of the pattern that is already known – that is, whichever one is most strongly associated with the already-active nodes. (Bechtel & Abrahamsen, 2002, p. 50) In the case of perception, people are often better able to remember what they saw or heard of an event when they are returned to the area in which the event took place. The brain has an easier time rebuilding the event's internal pattern when the part of the pattern consisting of the surroundings is externally supplied.

Objection

Fodor & Pylyshyn disagree with the view that graceful degradation / damage tolerance is a property unique to connectionist architecture. In *Connectionism and Cognitive Architecture: A Critical Analysis*, they claim that the distribution of a representation over units in a connectionist network achieves damage resistance only if the representations are neurally (physically) distributed as well, and they claim that such a neural distribution (which implies damage resistance) can be accomplished using symbolic architecture as well. Since symbolic expressions are frequently made up of connected parts, the symbols that make up the expression must be in some way adjacent (directly connected). Fodor and Pylyshyn argue that this adjacency can be functional, but need not necessarily be spatial. That is, not all the symbolic elements need to be physically adjacent for symbolic architecture to exist. Additionally, they claim that the symbolic elements themselves can be implemented in a way consistent with the connectionist network without eliminating the higher-level symbol. That is, the symbol can consist of a distributed set of interconnected, spatially distributed neurons, but that "mini-network" as a unit is still able to be seen as a distinct symbol. Therefore they claim that it is plausible that a symbolic mental architecture is indeed consistent with graceful degradation. (Fodor and Pylyshyn, 1988, 339-340)

Response

However, this argument seems inconsistent. These symbols, even though they are implemented by small neural networks, must in some way connect to the other symbols to form a representation. Supposing that there are very few connections between these symbols, they would still be incompatible with graceful degradation, as a destruction of those few connections would bring the whole representation down, just as in the case of low-level symbols. While the symbol itself could experience graceful degradation internally, the representation as a whole would not. Alternatively, if a symbol has many connections to adjacent symbols, it would indeed experience graceful degradation, but would seem to be conspicuously similar to a full connectionist network in which arbitrary sections are selected to act as symbols. In this case, by maintaining that the architecture is symbolic, Fodor and Pylyshyn would be missing the fact that connectionism is not merely an implementation of symbolic architecture, but a wholly distinct concept that can cope with phenomena that symbolic architecture handles well while also dealing with phenomena that connectionism handles better, such as learning, generalization, variation, and perception. (Bechtel & Abrahamsen, 2002, p. 162)

Conclusion

Overall then, it seems that the functioning of the brain, especially in the area of perception, is much more consistent with a connectionist network than with a symbolic architecture. Not only in the area of perception, but in general as well, is it evident that the brain experiences graceful degradation / damage tolerance, pattern recognition, and content-addressable memory. These phenomena are easily explained by a connectionist network model of the brain, while a symbolic architecture model struggles to achieve the necessary functionality. Although not conclusive, these points do seriously question the long-held belief that perceptual representation is symbolic, while providing a strong foundation for a connectionist-model understanding of perception.

Reference

Bechtel, W., & Abrahamsen, A. (2002). *Connectionism and the mind: Parallel processing, dynamics, and evolution in networks* (2nd ed.). Malden, MA: Blackwell. (Original work published 1991)

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. In J. Haugeland (Ed.), *Mind design II: Philosophy, psychology, artificial intelligence* (2nd ed., rev. and enlarged ed., pp. 309-350). Cambridge, Massachusetts: The MIT Press. (1997)

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999, August). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences of the United States of America, 96*, 9379-9384. Retrieved March 10, 2006, from National Academy of Sciences of the United States of America Web site: http://www.pnas.org/cgi/content/full/96/16/9379

Merkle, F. (2001, April). Consciousness, meet neuroscience [Special section]. *Caltech Undergraduate Research Journal, 1*(1), 12-24. Retrieved March 7, 2006, from Caltech Web site: http://www.curj.caltech.edu/pdf/vol1iss1full.pdf