

## Computation and the Brain

Patricia Smith Churchland  
University of California, San Diego

Rick Grush  
Center for Semiotic Research, Aarhus, Denmark

To Appear in *The MIT Encyclopedia of Cognitive Sciences*.  
Rob Wilson and Frank Keil, General Editors. MIT/Bradford.

Two very different insights motivate characterizing the brain as a computer. One depends on mathematical theory that defines computability in a highly abstract sense. Here the foundational idea is that of a Turing machine. Not an actual machine, the Turing machine is really a conceptual way of making the point that any well-defined function could be executed, step by step, according to simple “if-you-are-in-state-P-and-have-input-Q-then-do-R” rules, given enough time (maybe infinite time) [see COMPUTATION]. Insofar as the brain is a device whose input and output can be characterized in terms of some mathematical function -- however complicated -- then in that very abstract sense, it can be mimicked by a Turing machine. Given what is known so far brains do seem to depend on cause-effect operations, and hence brains appear to be, in some formal sense, equivalent to a Turing machine [see CHURCH-TURING THESIS]. On its own, however, this reveals nothing at all of how the mind-brain actually works. The second insight depends on looking at the brain as a biological device that processes information from the environment to build complex representations that enable the brain to make predictions and select advantageous behaviors. Where necessary to avoid ambiguity, we will refer to the first notion of computation as ‘algorithmic computation’, and the second as ‘information processing computation’.

What kind of computer, if any, the brain is, and the nature of its representations, remains controversial. The *algorithmic* computation school (traditional Artificial Intelligence research and “functionalism” in philosophy) favors the an analogy with serial digital computers that run various programs. Using this analogy, the idea is that mental processes such as reasoning are likened to the brain running a software program. Representations, according this approach, are like sentences or sentence constituents, and have a syntactic structure and a semantic content in much the same way that logical formulae can. This approach is most plausible for fairly high-level but narrowly constrained cognitive capacities such as doing logic, thinking in language [but see COGNITIVE LINGUISTICS]

or playing chess. Transformations that constitute the computations are modeled by sentence manipulation via rules of inference. One important assumption is that learning can be characterized on the model of hypothesis testing, and that the “learning device” itself does not change from birth to death. Another assumption is that the nature of the “mental program” is largely independent of the brain hardware, and a very different physical systems could run our “mental program”. A consequence of this assumption is the claim that neuroscience will purchase no insight into the mind, and that psychology is an “autonomous” science. (Fodor 1975) [See COMPUTATIONAL THEORY OF MIND]

The *information processing* computational approach, typified in connectionism, stresses the architecture of the brain itself. Connectionist (neural network) computing covers a very broad range of model types, but the common theme is the idea that the brain is a highly parallel machine consisting of many interconnected (typically analog) elements, and that its connectivity and wiring can change dramatically as a result of development and learning. Representation is characterized as a pattern of activation across the units in the neural net, where this can be described as a vector, and computations are therefore vector-vector transformations [See COGNITIVE MODELING, CONNECTIONIST; and NEURAL NETWORKS]. When such networks have internal feedback, they are more complex [See RECURRENT NETWORKS]. The hardware/software analogy favored in the algorithmic approach is challenged on the additional grounds that biological nervous systems are not general-purpose computers, but rather have evolved to perform highly specialized tasks. For two main reasons, much of the learning in brains is believed to be distinct from symbol/sentence manipulation. (a) Data show that changes in the wiring, including both growth and pruning, appear as a result of experience, and (b) behavioral evidence points to changes in the “learning device” itself during infancy, (see Quartz and Sejnowski 1997) Successful connectionist models include non-sentential learning, memory, perception and perceptual-motor control -- domains where algorithmic models have had mixed success. In addition, there are connectionist approaches to language (Elman et al. eds 1996) and many hybrid models that include both signal processing and symbol processing.

An important difference between the two approaches is that connectionism readily generalizes to brain function at all levels of organization, ranging from the level of molecules to whole systems and finally the entire nervous system in its body. This means that, unlike algorithmic computational models, connectionist models are applicable to sub-personal, sub-cognitive levels of brain function. Different functions take place at different organizational levels, and the character of the information processing computations the

nervous system uses to carry out these functions at various levels can be investigated using connectionist models. Thus to discover what mathematical function is executed by a biological neural network, an artificial neural network (ANN) can be trained to perform a similar information processing task. The ANN can then be investigated in great detail, and this can shed light on the operation of the biological neural system after which the ANN was designed.

For example, consider certain neurons in the parietal cortex of the brain whose response properties are correlated with the position of the visual stimulus relative to head-centered co-ordinates. Since the receptor sheets (retina, eye muscles) cannot provide that information directly, it has to be computed from various input signals. Two sets of neurons project to these cells: some represent position of the stimulus on the retina, some represent the position of the eyeball in the head. Modeling these relationships via an artificial neural net shows how the eyeball/retinal position can be used to compute the position of the stimulus relative to the head [see OCULOMOTOR CONTROL]. Once trained, the network's structure can be analyzed to determine how the computation was achieved. The results of such investigations often helps to guide research into real biological brains. For example, when ANN neurons exhibit certain responses while solving a problem, it leads to hypotheses concerning the sorts of response properties researchers might seek during single-cell recordings in real brains (Andersen (1995)). [See COMPUTATIONAL NEUROSCIENCE]

How biologically realistic to make an ANN depends on the purposes at hand, and different models are useful for different purposes. At certain levels and for certain purposes, abstract, simplifying models are precisely what is needed. Such a model will be more useful than a model slavishly realistic with respect to every level, even the biochemical. For example, if you want to know whether neuronal structures can compute sequences as output given sequences as input, as for example in language, a fairly abstract model will suffice. Excessive realism may mean that the model is too complicated to analyze or understand or run on the available computers. For other projects, such as investigating dendritic spine dynamics, the more realism at the biochemical level, the better [See AXONAL MODELING].

But why think of brains or neural systems or dendritic spines as in the computing/representing business at all, even in the connectionist sense? Like the liver, there are of course causal interactions between cells, but what makes it appropriate to say livers

filter but brains compute? Indeed, there is growing sympathy for approaches which see the brain, together with the body and environment, as dynamical systems best characterized by systems of differential equations describing the temporal evolution of states of the brain [See DYNAMIC APPROACHES TO COGNITION, and Port and van Gelder (1995)]. On this view both the brain and the liver can have their conduct adequately described by systems of differential equations.

The main reason for preferring a framework with computational resources derives from the observation that neurons represent various non-neural parameters, such as head velocity or muscle tension or visual motion, and that complex neuronal representations have to be constructed from simpler ones. Recall the example of neurons in area 7a. Their response profiles indicate that they represent the position of the visual stimulus in head-centered coordinates. Describing causal interactions between these cells and their input signals without specifying anything about representational role masks their function in the animal's visual capacity. It omits explaining how these cells come to represent what they do. Note that connectionist models can be dynamical when they include back projections, time constants for signal propagation, channel open times, as well as mechanisms for adding units and connections, and so forth.

In principle, dynamical models could be supplemented with representational resources in order to achieve more revealing explanations. For instance, it is possible to treat certain parameter settings as inputs, and the resultant attractor as an output, each carrying some representational content. Furthermore, dynamical systems theory easily handles cases where the 'output' is not a single static state (the result of a computation), but is rather a trajectory or limit cycle. Another approach is to specify dynamical subsystems within the larger cognitive system that function as models for external domains, perhaps the environment (see Grush, 1997). This approach allows both representational description of the model (the model represents the external domain), as well as an understanding of the overall system's function (to 'think about' the represented domain).

## References:

- Andersen, Richard A. (1995) 'Coordinate transformations and Motor Planning in Posterior Parietal Cortex'. In Gazzaniga, Michael (ed.) *The Cognitive Neurosciences*. Cambridge, MA: MIT Press.
- Elman, Jeffrey L., Elizabeth A. Bates, Mark H. Johnson, Annette Karmiloff-Smith, Domenico Parisi, and Kim Plunkett (eds, 1996) *Rethinking Innateness: A connectionist Perspective on Development*. Cambridge, MA: MIT Press.
- Fodor, Jerry (1975) *The language of thought*. Cambridge MA: Harvard University Press.
- Grush, Rick (1997) 'The Architecture of representation'. *Philosophical Psychology* 10(1):5-25.
- Port, Robert and Timothy van Gelder (1995) *Mind as motion: explorations in the dynamics of cognition*. Cambridge, MA: MIT Press.
- Quartz, Steven R., and Terrence J. Sejnowski (to appear) 'The neural basis of development: a constructivist manifesto'. *The Behavioral and Brain Sciences*.

## Further reading:

- Abeles, M. (1991) *Corticonics: Neural circuits of the cerebral cortex*. Cambridge, England: Cambridge University Press.
- Arbib, A.M. (1995) *The handbook of brain theory and neural networks*. Cambridge, MA: MIT Press.
- Boden, Margaret (1988) *Computer models of the mind*. Cambridge, England: Cambridge University Press.
- Churchland, Patricia, and Terrence Sejnowski (1992) *The Computational Brain*. Cambridge, MA: MIT Press.
- Churchland, Paul (1989) *A neurocomputational perspective*. Cambridge, MA: MIT Press.
- Churchland, Paul (1995) *The engine of reason, the seat of the soul*. Cambridge, MA: MIT Press.
- Koch, C., and I. Segev (1997) *Methods in neuronal modeling: From synapses to networks*, 2nd edn. Cambridge MA: MIT Press.
- Sejnowski, Terrence (1997) 'Computational neuroscience'. *Encyclopedia of Neuroscience*. Amsterdam: Elsevier Science Publishers.