Combined Hebbian and Error-driven learning: Transfer of learning in infants

Francisco Calvo Garzón
Investigador contratado *Programa "Ramón y Cajal"*
Departamento de Filosofía
Universidad de Murcia
Murcia – 30100 SPAIN
E-mail: fjcalvo@um.es

Gary Marcus (2001) distinguishes two separate ontologies in the connectionist realm: implementational connectionism and eliminative connectionism. The former accounts for cognitive phenomena by positing sets of explicit rules that serve the purpose of symbolic manipulation. The latter, in terms of computational abilities which are the result of an associative memory. He argues that only those connectionist models that incorporate (classical) rules can account for the phenomenon of transfer of learning in infants (Marcus et al., 1999). Seidenberg and Elman (1999) have tried to counter to Marcus by means of a simple recurrent network (SRN) trained on a *categorization* task. In this paper I show how a biologically plausible *prediction*-SRN can preserve its computational equivalency with respect to classical counterparts while eschewing the need to posit rule-governed underlying mechanisms; a criticism that has been raised against Seidenberg and Elman's categorization-based reply (Calvo and Colunga, under review).

My reply to Marcus is three-fold: I argue, first, (i) that he does not provide a robust criterion to decide when a network implements a rule. Second, (ii) that his proposal to implement recursive combinations by means of encodings where tokens that belong to the same representational type are identical lacks biological plausibility. And third, (iii) that even if points (i) and (ii) were wrong, the neural substrate may impose limitations on the mechanisms that permit the system carry out the appropriate computations. These limitations may tip the balance in favour of a biologically plausible Hebbian-cum-error-driven network arquitecture (O'Reilly and Munakata, 2000).

In this context, I shall explore what I dub "*non*-classical connectionism" (NCC)—the class of models that have different combinations of pattern associator/autoassociative memory/competitive network topologies, with bi-directional connectivity and inhibitory competition, that employ combined Hebbian and activation-phase learning algorithms (Rolls and Treves, 1998; O'Reilly and Munakata; 2000)—as a coherent framework for evaluating the transfer-of-learning-in-infants debate. The model proposed has a simple recurrent network architecture that has been supplemented with the following non-classical features: **(i)** bidirectional propagation of activation; **(ii)** inhibitory competition (kWTA); **(iii)** an error-driven form of learning (McClelland's GenRec); and **(iv)** Hebbian model learning (O'Reilly, 1998).

NCC, I contend, provides the basic toolkit to address the worries raised by Marcus with regard to the level of realism of connectionist natural language parsers. Marcus correctly points out that "Seidenberg and Elman (1999) do not give an account of how the supervisor's rule could itself be implemented in the neural substrate" (2001, p. 65). The teaching signal of a fully-supervised categorization task is not ecologically grounded. In a self-supervised prediction task, however, that's not an issue. Activation-based signals in a prediction task are *not* to be interpreted in terms of rule-implementation. The teaching signal exploited *is* an activation state of experience.

Combined Hebbian and error-driven learning furnishes neuromodelers with a biologically plausible framework to simulate cognitive activity. Neurobiology tells us that both forms of learning take place in the cortex, and that their combined effect may be used to account for modifications at the synaptic level—O'Reilly & Munakata, 2000, pp. 168-170. Error-driven learning makes use of an activation-phase algorithm that, via bi-directional connectivity and symmetric weight matrices, permits the network to alter the knowledge acquired in the weights by computing the difference between an initial phase where the networks activations are interpreted as its "expectation" of what's to happen, and a later phase where the environment provides the output response to be taken as the teaching signal. Hebbian learning, on the other hand, makes its contribution by representing in hidden space the first-order correlational structure of the data pool. Inhibitory competition at the output layer, furthermore, helps by forcing the network, via kWTA, for example, to code for sparse distributed representations (O'Reilly, 1998). The network, thus, encodes the existing statistical regularities with no need to process algebra-like information. My working hypothesis is that infants make use of discrepancies based on expectations to make successful predictions.

The neurosimulation reported successfully delivers a correct syntactic interpretation of the infants' data, and backs up empirically the working hypothesis of eliminative connectionism. The data is accounted for in statistical terms—without the positing of devices that store particular values of variables to perform variable bindings, such as register sets, or other classical resources. Marcus's charge of implementation is therefore not applicable, since an ecologically grounded prediction tasks doesn't incorporate universally open-ended rules. I conclude that Marcus fails in his attempt to override eliminative connectionism.

References

Calvo, F and Colunga, E. (under review) "The Statistical Brain: Reply to Marcus' *The Algebraic Mind*."

Proceedings to the 25th Annual Conference of the Cognitive Science Society. Hillsdale, NJ: Lawrence Erlbaum Associates.

Marcus, G.F. (2001) *The Algebraic Mind*. Cambridge, Mass.: MIT Press.

Marcus, G.F., Vijayan, S., Rao, S. B., Vishton, P. M . (1999) Rule learning in seven-month-old infants. *Science* 283, 77-80.

O'Reilly, R. and Munakata, Y. (2000) *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain*. Cambridge, Mass.: MIT Press.

O'Reilly, R. (1998) Six principles for biologically based computational models of cortical cognition. *Trends Cognit. Sci.* 2, 455–462.

Rolls, E.T. and Treves, A. (1998) *Neural Networks and Brain Function*. Oxford: Oxford University Press.

Seidenberg, M.S. and Elman, J.L. (1999) Generalization, rules, and neural networks: A simulation of Marcus et al. *http://crl.ucsd.edu/~elman/Papers/MVRVsimulation.html*.