Connectionism coming of age: legacy and future challenges

Julien Mayor[1]*, Pablo Gomez[2], Franklin Chang[3] and Gary Lupyan[4]

[1] Department of Psychology and Educational Sciences, University of Geneva, Genève, Switzerland
[2] Department of Psychology, De Paul University, Chicago, IL, USA
[3] Department of Psychological Sciences, University of Liverpool, Liverpool, UK
[4] Department of Psychology, University of Wisconsin, Madison, WI, USA
*Correspondence: julien.mayor@unige.ch

ABOUT 50 YEARS AFTER THE INTRODUCTION OF THE PERCEPTRON AND SOME 25 YEARS AFTER THE INTRODUCTION OF PDP MODELS, WHERE ARE WE NOW?

In 1986, Rumelhart and McClelland took the cognitive science community by storm with the Parallel Distributed Processing (PDP) framework. Rather than abstracting from the biological substrate as was sought by the "information processing" paradigms of the 1970s, connectionism, as it has come to be called, embraced it. An immediate appeal of the connectionist agenda was its aim: to construct at the algorithmic level models of cognition that were compatible with their implementation in the biological substrate.

The PDP group argued that this could be achieved by turning to networks of artificial neurons, originally introduced by McCulloch and Pitts (1943) which the group showed were able to provide insights into a wide range of psychological domains, from categorization, to perception, to memory, to language. This work built on an earlier formulation by Rosenblatt (1958) who introduced a simple type of feed-forward neural network called the perceptron. Perceptrons were limited to solving simple linearly-separable problems and although networks composed of perceptrons were known to be able to compute any Boolean function (including XOR, Minsky and Papert, 1969), there was no effective way of training such networks. In 1986, Rumelhart, Hinton and Williams introduced the back-propagation algorithm, providing an effective way of training multi-layered neural networks, which could easily learn non linearly-separable functions. In addition to providing the field with an effective learning algorithm, the PDP group published a series of demonstrations of how long standing questions in cognitive psychology could be elegantly solved using simple learning rules, distributed representations, and interactive processing.

To take a classic example, consider the word-superiority effect, in which people can detect letters within a word faster than individual letters or letters within a non-word (Reicher, 1969). This result is difficult to square with serial "information-processing" theories of cognition that were dominant at the time (how could someone recognize "R" before "FRIEND" if recognizing the word required recognizing the letters?). Accounting for such findings demanded a framework which could naturally accommodate interactive processes within a bidirectional flow of information.

The so-called "Interactive-activation model" (McClelland and Rumelhart, 1981) provided just such a framework.

The connectionist paradigm was not without its critics. The principal critiques can be divided into three classes. First, some neuroscientists (Crick, 1989) questioned the biological plausibility of backpropagation, when they failed to observe experimentally complex and differentiated back-propagating signals that are required to learn in multi-layered neural networks. A second critique concerned stability-plasticity of the learned representations in these models. Some phenomena require the ability to rapidly learn new information, but sometimes newly learned knowledge overwrites previously learned information (catastrophic interference; McCloskey and Cohen, 1989). Third, representing spatial and temporal invariance—something that apparently came easily to people—was difficult for models, e.g., recognizing that the letter "T" in "TOM" was the "same" as the "T" in "POT." This invariance problem was typically solved by multiplying a large number of hard-wired units that were space- or time-locked (see e.g., McClelland and Elman, 1986). Finally, critics pointed out that the networks were incapable of learning true rules on which a number of human behavioral, namely language-learning was thought to depend (e.g., Marcus, 2003; cf. Fodor and Pylyshyn, 1988; Seidenberg, 1999).

The connectionist approach has embraced these challenges: Although some connectionist models continue to rely on back-propagation, others have moved to more biologically realistic learning rules (Giese and Poggio, 2003; Masquelier and Thorpe, 2007). Far from being a critical flaw of connectionism, the phenomenon of catastrophic interference (Mermillod et al., 2013) proved to be a feature that led to the development of complementary learning systems (McClelland et al., 1995).

Progress has also been made on the invariance problem. For example, within the speech domain representing the similarity between similar speech sounds regardless of their location within a word has been addressed in the past by Grossberg and Myers (2000) and Norris (1994) and this issue presents a new more streamlined and computationally efficient model (Hannagan et al., 2013). An especially powerful approach to solving the location invariance problem in the visual domain is presented by Di Bono and Zorzi (2013), also in this issue.

A key challenge for connectionism is to explain the learning of abstract structural representations. The use of recurrent networks (Elman, 1990; Dominey, 2013) and self-organizing maps, has captured important aspects of language learning (e.g., Mayor and Plunkett, 2010; Li and Zhao, 2013), while work on deep learning (Hinton and Salakhutdinov, 2006) has made it possible to model the emergence of structured and abstract representations within multi-layered hierarchical networks (Zorzi et al., 2013). The work on verbal analogies by Kollias and McClelland (2013) continues to address the challenges of modeling more abstract representations, but truly understanding how neural architectures give rise to symbolic cognition is a gap that remains. Although learning and representing formal language rules may not be completely outside of the abilities of neural networks (e.g., Chang, 2009), it seems clear that understanding human cognition requires understanding how we solve these symbolic problems (Clark and Karmiloff-Smith, 1993; Lupyan, 2013). Future generations of connectionist modelers may wish to fill this gap and in so doing provide a fuller picture of how neural networks give rise to intelligence of the sort enables us to ponder the very workings of our cognition.

WHAT'S NEXT?

The articles assembled in this issue demonstrate the range of topics currently addressed by connectionist models: from word learning in atypical populations (Sims et al., 2013), to sentence processing (Hsiao and MacDonald, 2013), to multimodal processing (Bergmann et al., 2013), to interactions between language and vision (Smith et al., 2013). We expect this diversity to continue to increase. We also hope to see greater increasing integration between connectionism and a computationally similar but philosophically distinct models employing Bayesian inference. Although the computational similarities between these two approaches have been previously recognized (McClelland, 1998), detailed tutorials like the one contained in this volume (McClelland, 2013) provide new clarity on the relationship between these two approaches.

The influence of theoretical constructs introduced by the connectionist approach have become part and parcel of cognitive science (although they are now often not accompanied by the label "connectionism" or "PDP"). The distributed representations that challenge classical symbolic models and which emerge naturally in neural networks are now no longer theoretical constructs and can be directly observed in the brain (Kriegeskorte et al., 2008; Chang et al., 2010). Evidence for rapid warping of these representations by task demands (of the sort described by e.g., McClelland and Rogers, 2003) is also being confirmed through modern neuroimaging (e.g., Çukur et al., 2013) [1]. Many connectionist models have stressed prediction as a way of learning structure and statistical inputs (e.g., Dell and Chang, 2014). This too finds wide support in contemporary neuroscience (Friston, 2010) leading some to even argue that prediction is the unifying feature of all cognitive and perceptual processes (Clark, 2013, for review). Interactive processing—another core feature of the connectionist paradigm—has become similarly foundational. The interplay between bottom-up and top-down information is now recognized to be critical from everything as simple as simply detecting the presence of a visual stimulus, to consciousness itself (e.g., Dehaene et al., 2003; Gilbert and Sigman, 2007; Lupyan and Ward, 2013).

Finally, contemporary neural networks, most notably those utilizing so called deep-learning, have found success in solving practical problems such as image and speech recognition, and natural language processing. For example, algorithms based on the deep-learning approach are now used by Google to extract high-level features from images with, in some cases, above-human performance (Ciresan et al., 2011; Le et al., 2011).

REFERENCES

Bergmann, C., ten Bosch, L., Fikkert, P., and Boves, L. (2013). A computational model to investigate assumptions in the headturn preference procedure. *Front. Psychol.* 4:676. doi: 10.3389/fpsyg.2013.00676

Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., and Knight, R. T. (2010). Categorical speech representation in human superior temporal gyrus. *Nat. Neurosci.* 13, 1428–1432. doi: 10.1038/nn.2641

Chang, F. (2009). Learning to order words: a connectionist model of heavy NP shift and accessibility effects in Japanese and English. *J. Mem. Lang.* 61, 374–397. doi: 10.1016/j.jml.2009.07.006

Ciresan, D. C., Meier, U., Masci, J., Gambardella, L. M., and Schmidhuber, J. (2011). "Flexible, high performance convolutional neural networks for image classification," in *International Joint Conference on Artificial Intelligence IJCAI-2011.* (Barcelona), 1237–1242.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204. doi: 10.1017/S0140525X12000477

Clark, A., and Karmiloff-Smith, A. (1993). The cognizer's innards: a psychological and philosophical perspective on the development of thought. *Mind Lang.* 8, 487–519. doi: 10.1111/j.1468-0017.1993.tb00299.x

Crick, F. (1989). The recent excitement about neural networks. *Nature* 337, 129–132. doi: 10.1038/337129a0

Çukur, T., Nishimoto, S., Huth, A. G., and Gallant, J. L. (2013). Attention during natural vision warps semantic representation across the human brain. *Nat. Neurosci.* 16, 763–770. doi: 10.1038/nn.3381

Dehaene, S., Sergent, C., and Changeux, J.-P. (2003). A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proc. Natl. Acad. Sci. U.S.A.* 100, 8520–8525. doi: 10.1073/pnas.1332574100

Dell, G. S., and Chang, F. (2014). The P-Chain: Relating sentence production and its disorders to comprehension and acquisition. *Philos. Trans. R. Soc. B Biol. Sci.* 369, 20120394. doi: 10.1098/rstb.2012.0394

Di Bono, M. G., and Zorzi, M. (2013). Deep generative learning of location-invariant visual word recognition. *Front. Psychol.* 4:635. doi: 10.3389/fpsyg.2013.00635

Dominey, P. F. (2013). Recurrent temporal networks and language acquisition—from corticostriatal neurophysiology to reservoir computing. *Front. Psychol.* 4:500. doi: 10.3389/fpsyg.2013.00500

Elman, J. (1990). Finding structure in time. *Cogn. Sci.* 14, 179–212. doi: 10.1207/s15516709cog1402_1

Fodor, J. A., and Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: a critical analysis. *Cognition* 28, 3–71. doi: 10.1016/0010-0277(88)90031-5

---

[1] It is useful to note that the methods that make these analyses possible, most notably multi-voxel pattern analyses (MVPA, e.g., Norman et al., 2006) and "representational dissimilarity matrices" (Kriegeskorte et al., 2008) are adaptations of methods developed for analyzing dynamics of artificial neural networks.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787

Giese, M., and Poggio, T. (2003). Neural mechanisms for the recognition of biological movements and action. *Nat. Rev. Neurosci.* 4, 179–192. doi: 10.1038/nrn1057

Gilbert, C. D., and Sigman, M. (2007). Brain states: top-down influences in sensory processing. *Neuron* 54, 677–696. doi: 10.1016/j.neuron.2007.05.019

Grossberg, S., and Myers, C. W. (2000). The resonant dynamics of speech perception: interword integration and duration-dependent backward effects. *Psychol. Rev.* 107, 735. doi: 10.1037/0033-295X.107.4.735

Hannagan, T., Magnuson, J. S., and Grainger, J. (2013). Spoken word recognition without a TRACE. *Front. Psychol.* 4:563. doi: 10.3389/fpsyg.2013.00563

Hinton, G. E., and Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science* 313, 504–507. doi: 10.1126/science.1127647

Hsiao, Y., and MacDonald, M. C. (2013). Experience and generalization in a connectionist model of Mandarin Chinese relative clause processing. *Front. Psychol.* 4:767. doi: 10.3389/fpsyg.2013.00767

Kollias, P., and McClelland, J. L. (2013). Context, cortex, and associations: a connectionist developmental approach to verbal analogies. *Front. Psychol.* 4:857. doi: 10.3389/fpsyg.2013.00857

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., et al. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* 60, 1126–1141. doi: 10.1016/j.neuron.2008.10.043

Le, Q. V., Ranzato, M.'A., Monga, R., Devin, M., Chen, K., Corrado, G. S., et al. (2011). Building high-level features using large scale unsupervised learning. *arXiv preprint arXiv*:1112.6209.

Li, P., and Zhao, X. (2013). Self-organizing map models of language acquisition. *Front. Psychol.* 4:828. doi: 10.3389/fpsyg.2013.00828

Lupyan, G. (2013). The difficulties of executing simple algorithms: why brains make mistakes computers don't. *Cognition* 129, 615–636. doi: 10.1016/j.cognition.2013.08.015

Lupyan, G., and Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proc. Natl. Acad. Sci. U.S.A.* 110, 14196–14201. doi: 10.1073/pnas.1303312110

Marcus, G. F. (2003). *The Algebraic Mind: Integrating Connectionism and Cognitive Science*. Cambridge: The MIT Press.

Masquelier, T., and Thorpe, S. J. (2007). Unsupervised learning of visual features through spike timing dependent plasticity. *PLoS Comput. Biol.* 3:e31. doi: 10.1371/journal.pcbi.0030031

Mayor, J., and Plunkett, K. (2010). A neurocomputational model of taxonomic responding and fast mapping in early word learning. *Psychol. Rev.* 117, 1–31. doi: 10.1037/a0018130

McClelland, J. L. (1998). "Connectionist models and Bayesian inference," in *Rational Models of Cognition*, eds M. Oaksford and N. Chater (Oxford: Oxford University Press), 21–53.

McClelland, J. L. (2013). Integrating probabilistic models of perception and interactive neural networks: a historical and tutorial review. *Front. Psychol.* 4:503. doi: 10.3389/fpsyg.2013.00503

McClelland, J. L., and Elman, J. L. (1986). The TRACE model of speech perception. *Cogn. Psychol.* 18, 1–86. doi: 10.1016/0010-0285(86)90015-0

McClelland, J. L., McNaughton, B. L., and O'Reilly, R. C. (1995). Why there are complementary learning-systems in the hippocampus and neocortex - insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* 102, 419–457.

McClelland, J. L., and Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nat. Rev. Neurosci.* 4, 310–322. doi: 10.1038/nrn1076

McClelland, J. L., and Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychol. Rev.* 88, 375–407. doi: 10.1037/0033-295X.88.5.375

McCloskey, M., and Cohen, N. J. (1989). Catastrophic interference in connectionist networks: the sequential learning problem. *Psychol. Learn. Motiv.* 24, 109–164. doi: 10.1016/S0079-7421(08)60536-8

McCulloch, W. S., and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* 5, 115–133. doi: 10.1007/BF02478259

Mermillod, M., Bugaiska, A., and Bonin, P. (2013). The stability-plasticity dilemma: investigating the continuum from catastrophic forgetting to age-limited learning effects. *Front. Psychol.* 4:504. doi: 10.3389/fpsyg.2013.00504

Minsky, M. L., and Papert, S. A. (1969). *Perceptrons*. Cambridge, MA: MIT Press.

Norman, K. A., Polyn, S. M., Detre, G. J., and Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn. Sci.* 10, 424–430. doi: 10.1016/j.tics.2006.07.005

Norris, D. (1994). Shortlist: a connectionist model of continuous speech recognition. *Cognition* 52, 189–234. doi: 10.1016/0010-0277(94)90043-4

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *J. Exp. Psychol.* 81, 275. doi: 10.1037/h0027768

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol. Rev.* 65, 386. doi: 10.1037/h0042519

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors *Nature* 323, 9. doi: 10.1038/323533a0

Rumelhart, D. E., and McClelland, J. L. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Vol. 1. Cambridge, MA: Foundations MIT Press.

Seidenberg, M. S. (1999). Do infants learn grammar with algebra or statistics? *Science* 284, 434–435. author reply: 436–437.

Sims, C. E., Schilling, S. M., and Colunga, E. (2013). Beyond modeling abstractions: learning nouns over developmental time in atypical populations and individuals. *Front. Psychol.* 4:871. doi: 10.3389/fpsyg.2013.00871

Smith, A. C., Monaghan, P., and Huettig, F. (2013). An amodal shared resource model of language-mediated visual attention. *Front. Psychol.* 4:528. doi: 10.3389/fpsyg.2013.00528

Zorzi, M., Testolin, A., and Stoianov, I. P. (2013). Modeling language and cognition with deep unsupervised learning: a tutorial overview. *Front. Psychol.* 4:515. doi: 10.3389/fpsyg.2013.00515