Wordcount main text: 1977

Associative Learning, From Conceptualization to Implementation

Ángel Eugenio Tovar[1] & Gert Westermann[2]

[1]Facultad de Psicología, Universidad Nacional Autónoma de México, Av. Universidad 3004, Coyoacan, Mexico

[2]Department of Psychology, Lancaster University, Lancaster LA1 4YF, United Kingdom

Abstract

Two main commitments appear in Rey's Unified Radical Associationism: 1) using Hebbian learning as a unified computational framework in psychological science, 2) using associations as a unique construct to account for mental activities. Hebbian learning has been shown to account for complex behavioral repertoires in experimental paradigms coming from different psychological traditions. Here we review the processing mechanisms underlying such a complex repertoire, we suggest two additional points to extend the associationist hypothesis, and we point to the need to understand how implemented associative learning mechanisms are balanced and modulated. We believe these points are key for future research to support or question both commitments in Rey's Unified Radical Associationism.

We applaud Rey's call for a radical associationism. We believe that too often, high-level processes ('rich' interpretations, Haith, 1998) are put forward where low-level ('lean') explanation would suffice. At the very least, radical associationism can push us to explore the limits of lean interpretations by building explanations of cognitive processes from the bottom up instead of postulating high-level descriptions of behavior as the underlying mechanisms of cognition (see also McClelland et al., 2010).

Rey puts forward 10 initial points to circumscribe the *associationist hypothesis*. He also proposes the computational framework of Hebbian learning as the central axis to unify learning theories in psychological science. However, the classical conception of Hebbian learning may seem limited to only explain the associative strength between co-occurring events, and therefore not useful as an account for learning in more complex situations such as human language acquisition, where learning associations between nonadjacent elements or deriving associations from other preexisting ones happens all the time.

In this commentary we first review a set of mechanisms that provide implemented Hebbian neural networks with the power to account for complex cognitive processes such as those involved in language; we then suggest two additional points to Rey's 10-point list that we believe would significantly expand the reach of the *associationist hypothesis*, and we finally discuss a set of critical questions pointing towards future research in associative learning to test Rey's *Unified Radical Associationism.*

Key Features in Hebbian Neural Networks for Complex Processing

Hebbian neural networks simulate experimental paradigms through a series of simple steps: external objects/stimuli (e.g., *A* and *B*) are presented to the network through input vectors following experimental schedules, some units in the network (e.g., neurons *A* and *B*) selectively respond to these objects, and connection weights (e.g., *WAB*) between active units change in Hebbian fashion based on co-occurrence of stimuli. Network performance is usually evaluated through analyzing unit activations (e.g., responses) and connection weights (e.g., interpreted as associative strengths).

The above description makes it straightforward to understand how these networks account for the learning of co-occurring events. However, a number of reasonable questions also emerge, such as how to prevent overgrowing in connection weights; how to account for associative learning between objects that appear distant in time; and whether it is possible to derive new associations based on previous ones. We next describe some key components and mechanisms that enrich the computational framework of Hebbian learning to considerably expand the empirical phenomena covered by these networks.

Spreading Activation and Activation Decay are key components to capture complex associations, sequence learning and associations between distant objects. For example, after unit *A* has been repeatedly coactive with *B* and a strong connection between A and B (*WAB*) has formed, further presentations of *B* will trigger spreading activation to *A*, enabling the network to recall previous associations, accounting for phenomena such as priming and prediction/anticipation (Tovar et al., 2018). Moreover, through this mechanism it is possible to account for derived or "emergent" associations between objects that have never been presented together, because recalling an object *A*, for example during presentations of *B and C together*, allows the activation of *A* to co-occur with activation of *C*, triggering Hebbian learning between *A* and *C* and accounting for derived transitivity (i.e., if A=B and B=C, then A=C; Tovar & Westermann, 2017)

A second mechanism enabling more complex learning is the gradual decay of unit activations when a stimulus is removed. In this way, the unit's residual activation may co-occur with other newly appearing stimuli. The amount of activation decay modulates the associative strengths between previous and current stimuli, enabling the modeling of experimental paradigms where sequences are relevant such as the canonical tasks used in Statistical Learning: word-segmentation, artificial grammar learning and serial reaction time (Tovar & Westermann, 2023).

Connection Decay and Connection Limit: Hebbian learning is typically conceptualized as connection strengthening such as Long-Term Potentiation (LTP) in biological neural networks. But connections do not grow endlessly, they stabilize at a functional level. This can be computationally achieved by including two constraints. One is connection decay that captures an adaptive synaptic process of weakening of connection weights following low coactivation values. This process is sometimes called Anti-Hebbian learning and it is biologically inspired by reports of Long-Term Depression (LTD) of synaptic efficacy. The second constraint is a connection limit: in previous work (Tovar et al., 2018; Tovar & Westermann, 2017) we have proposed a self-regulated parameter that limits the growth of connections and captures metaplasticity as it tunes the amount and direction of weight changes depending on the difference between the current weight (e.g., *WAB*) and the current coactivation (*AB*). This limiting effect is mathematically equivalent to the lambda parameter in the Rescorla & Wagner model (1972), showing a convergent approach in both associative models.

LTP/LTD threshold: With both connection strengthening and decay triggered by unit coactivations, determining which coactivation levels should result in strengthening and which ones in decay emerges as an interesting question. We have suggested a computational implementation following two complementary notions: one is the BCM model (Bienenstock et al., 1982) which proposed an LTP/LTD threshold; here, over-threshold coactivations (e.g., strong *A* and strong *B*) result in strengthening the connection weight, and below-threshold coactivations (e.g., weak *A* and strong *B*) result in connection decay. Notably, for the BCM theory the threshold is not a fixed value but slides as a function of previous postsynaptic activity (experience). The second notion comes from reports of animal models of intellectual disability showing an unbalanced LTP/LTD with increased LTD and weak LTP (Rueda et al., 2012), which can be captured through a relatively higher threshold. This is interesting because when we have modeled such imbalance with a higher LTP/LTD threshold, the resulting performance in the neural network mimics and predicts the behavior of participants with intellectual disabilities in Statistical Learning and Equivalence tasks (Tovar et al., 2018; Tovar & Westermann, 2017). But we have also found that slight changes in the threshold are needed to account for typical performance across different experimental paradigms. Thus setting the LTP/LTD balance is both an important problem for future research (see below) and a promising approach to model human processing and disorders, directly linking neural processing to cognitive behaviors.

By enriching the traditional Hebbian algorithm with the above mechanisms, Hebbian neural networks evolve from learning one-to-one associations to displaying a complex performance that models learning during challenging tasks. For example, during Statistical Learning paradigms such models learn associations between nonadjacent dependencies, because activation decay and spreading activation allow the establishment of new links between distant objects. Moreover, weight decay, connection limits and the LTP/LTD threshold keep these distant associations with a medium strength accounting for empirical reports of medium accuracy, discriminability and changes in response speed in human participants when acquiring such repertoires (Tovar & Westermann, 2017a, 2023). Notably, these Hebbian networks account for numerous paradigms coming from different traditions such as Behavior Analysis, Implicit Learning and Statistical Learning, providing evidence in support of Rey's proposal of Hebbian learning as a central framework for unifying learning theories.

Two additional points to enrich the associationist hypothesis

The ten points put forward by Rey provide a comprehensive characterization of the associationist framework. Eventually, this framework's challenge is to explain higher cognitive functions as the outcome of associative processes. We propose two further points to add to this framework that we believe have great potential to enhance learning in associative networks to provide mechanistic accounts of the emergence of higher-level processes:

11. Networks grow and shrink, adapting their functionality to the task. Over the first years of life, biological networks develop in an experience-dependent, constructive way (Quartz & Sejnowski, 1997). Such guided growth presents substantial advantages over learning in maximal, fixed architectures (Quartz, 1993) in terms of speed and robustness, and can lead to functional specialization in developing sub-components (Westermann & Ruh, 2012), integrating the statistical structure of the environment into the functional structure of the resulting network and showing rule-like and abstract generalization behavior (Westermann, 2016).

12. Learners choose their input, arguably to optimize their in-the-moment learning. This point embeds the associative learning mechanism in a behaving agent. We have shown that the order in which information is encountered (e.g., blocked vs. interleaved presentation of stimuli from two classes) shapes the learning outcome (Tovar et al., 2018). These results marry with recent research on infants' learning showing that infants take an active role in their information selection to maximize their learning progress (Poli et al., 2020; Twomey & Westermann, 2018). With a fixed learning mechanism, therefore, active learners can generate an optimized sequence of information selection leading to enhanced learning over a random (or non-optimal) sequence.

Integrating these features into the associationist framework will enhance this framework to account for developmental processes and, we believe, enable systems that more naturally support higher level cognitive processes.

Future Directions in Associative Hebbian Learning

One relevant research agenda to advance the associationist hypothesis is focused on understanding the interactive effects of processing mechanisms/constraints in associative systems. Taking the case of the Hebbian constraints reviewed above, a close inspection of their functioning rapidly reveals their interactive nature because changing the value of a given parameter has effects on the functioning of the others. For example, a relatively higher LTP/LTD threshold restricts the number of connections undergoing positive changes in the network, but increasing the rate of activation decay would rescue a proportion of such connections. Now consider that these network parameters also interact with the processing restrictions from our two additional suggested points: structural adaptation of the network (growing or shrinking) and active information selection. This highlights the complex role of balancing between the interactive learning restrictions; moreover, the optimal balance (i.e., optimal parameter space) in associative systems differs depending on task demands (Tovar & Westermann, 2023); for example, a lower LTP/LTD threshold is useful for learning distant associations, but not all tasks require this.

These considerations guide us to propose several key questions for future research towards a deeper understanding of associative learning:

How is the parameter balance in associative networks naturally achieved for each task?

How does a system (or organism) autonomously (implicitly or explicitly) find the optimal parameter combination in this multi-dimensional space? For example, how does explicitly discovering the efficacy of learning certain associations (adjacent or nonadjacent) modulate the network constraints?

Is the balance best explained by other processes, such as top-down modulation from attention, working memory, and other executive processes, or are Hebbian constraints and environmental contingencies enough?

What is the relationship between the achieved balance and development?

The above are not trivial questions; their answers are at the heart of confirming or questioning Arnaud Rey's proposal that "associations are all we need". While associations are necessary to account for complex behavior, we have highlighted the need to understand how these associations are governed and finely tuned, either self-regulated or modulated by other processes. This turns the question into *how Hebbian learning interacts with/or depends on other cognitive processes,* and whether a single process is sufficient as argued by Rey, or whether multiple processes (e.g., one processing system for low level associations, and a second high-level system for selective, attention-guided learning; Conway, 2020) are necessary. Answers to our list of questions may provide evidence to discern between both perspectives and would ultimately support, undermine or reshape radical associationism.

Conclusion

There are two major commitments in Arnaud Rey's proposal. The first one, for which there is already some evidence, is using Hebbian learning as a *unified* computational framework for learning theories. The second is using associations as a *unique* construct to account for mental activities. Rey's theoretical positioning pushes research to explore the explanatory limits of associative learning. Here we have motivated a research agenda to understand how associative learning mechanisms are regulated. Discovering what is involved in bringing associative mechanisms into balance is essential to test and advance the exciting associative project proposed by Arnaud Rey.

References

Bienenstock, E. L., Cooper, L. N., & Munro, P. W. (1982). Theory for the development of neuron selectivity: Orientation specificity and binocular interaction in visual cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *2*(1), 32–48. https://doi.org/10.1523/JNEUROSCI.02-01-00032.1982

Haith, M. M. (1998). Who put the cog in infant cognition? Is rich interpretation too costly? *Infant Behavior and Development*, *21*(2), 167–179. https://doi.org/10.1016/S0163-6383(98)90001-7

McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., & Smith, L. B. (2010). Letting structure emerge: Connectionist and dynamical systems approaches to cognition. *Trends in Cognitive Sciences*, *14*(8), 348–356. https://doi.org/10.1016/j.tics.2010.06.002

Poli, F., Serino, G., Mars, R. B., & Hunnius, S. (2020). Infants tailor their attention to maximize learning. *Science Advances*, *6*(39), eabb5053. https://doi.org/10.1126/sciadv.abb5053

Quartz, S. R. (1993). Neural networks, nativism, and the plausibility of constructivism. *Cognition*, *48*(3), 223–242. https://doi.org/10.1016/0010-0277(93)90041-S

Quartz, S. R., & Sejnowski, T. J. (1997). The neural basis of cognitive development: A constructivist manifesto. *Behavioral and Brain Sciences*, *20*(4), 537–556. https://doi.org/10.1017/S0140525X97001581

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. En A. H. Black & Prokasy (Eds.), *Classical Conditioning II: Current Research and Theory: Vol. Vol. 2* (pp. 64–99). Appleton-Century-Crofts.

Rueda, N., Flórez, J., & Martínez-Cué, C. (2012). Mouse models of Down syndrome as a tool to unravel the causes of mental disabilities. *Neural Plasticity*, *2012*, 584071. https://doi.org/10.1155/2012/584071

Tovar, Á. E., & Westermann, G. (2017). A Neurocomputational Approach to Trained and Transitive Relations in Equivalence Classes. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01848

Tovar, Á. E., & Westermann, G. (2023). No need to forget, just keep the balance: Hebbian neural networks for statistical learning. *Cognition*, *230*, 105176. https://doi.org/10.1016/j.cognition.2022.105176

Tovar, Á. E., Westermann, G., & Torres, A. (2018). From altered synaptic plasticity to atypical learning: A computational model of Down syndrome. *Cognition*, *171*, 15–24. https://doi.org/10.1016/j.cognition.2017.10.021

Twomey, K. E., & Westermann, G. (2018). Curiosity-based learning in infants: A neurocomputational approach. *Developmental Science*, *21*(4), e12629. https://doi.org/10.1111/desc.12629

Westermann, G. (2016). Experience-Dependent Brain Development as a Key to Understanding the Language System. *Topics in Cognitive Science*, *8*(2), 446–458. https://doi.org/10.1111/tops.12194

Westermann, G., & Ruh, N. (2012). A neuroconstructivist model of past tense development and processing. *Psychological Review*, *119*(3), 649–667. https://doi.org/10.1037/a0028258