1

The Cognitive Neuroscience of Learning

Introduction and Intent

Robert C. Honey and Robin A. Murphy

If an organism's behavior is to become better tuned to its environment, then there must be plasticity in those systems that interact with that environment. One consequence of such plasticity is that the organism's mental life is no longer bound to the here and now but reflects the interplay between the here and now and the there and then. Scientists from a variety of disciplines have studied the processes of learning that provide the basis for this interplay. While some have inferred the nature of the underlying conceptual or hypothetical processes through the detailed analysis of behavior in a range of experimental preparations, others have examined the neural processes and brain systems involved by making use of these and other preparations. To be sure, the preparations that have been employed often vary considerably in terms of their surface characteristics and the uses to which they are put. But this fact should not distract one from attempting to develop a parsimonious analysis, and it with this principle in mind that this handbook was conceived. Its focus is on the cognitive neuroscience of learning. Our frequent use of the qualifier Associative, as in Associative Learning, reflects either our bias or the acknowledgment of the fact that the formal analysis of all learning requires an associative perspective.

According to an associative analysis of learning, past experiences are embodied in the changes in the efficacy of links among the constituents of that experience. These associative links allow the presence of a subset of the constituents to affect the retrieval of a previous experience in its entirety: they provide a link, both theoretically and metaphorically, between the past and the present. We focus on this process because it has provided the basis for integration and rapprochement across different levels of analysis and different species, and it has long been argued that associative learning provides a potential shared basis for many aspects of behavior and cognition – for many forms of learning that might appear superficially distinct.

> Hence, the temporary nervous connexion is a universal physiological phenomenon both in the animal world and in our own. And at the same time it is likewise a psychic phenomenon, which psychologists call an association, no matter whether it is a combination of various

The Wiley Handbook on the Cognitive Neuroscience of Learning, First Edition.
Edited by Robin A. Murphy and Robert C. Honey.

> actions or impressions, or that of letters, words, and thoughts. What reason might there be for drawing any distinction between what is known to a physiologist as a temporary connexion and to a psychologist as an association? Here we have a perfect coalescence, a complete absorption of one by the other, a complete identification. Psychologists seem to have likewise acknowledged this, for they (or at any rate some of them) have made statements that experiments with conditioned reflexes have provided associative psychology … with a firm basis. (Pavlov, 1941, p. 171)

The breadth of application evident in Pavlov's treatise, and that of some of his contemporaries and successors, has often struck many as overly ambitious, provocative, or even plain misguided. The idea that what seems to be a rather simple process might play a role in such a broad range of phenomena is certainly bold; and some have argued that such an enterprise is flawed for a variety of reasons: where is the direct evidence of the operation of associative processes, how could such a simple process be sensitive to the inherent complexity and ambiguity in the real world, and so on. These and other criticisms have been acknowledged and have played an important role in shaping, for example, investigations of the brain bases of associative learning, and the development and assessment of more complex associative models that explicitly address a broad range of phenomena. This is not to say that the critics have been silenced or have even become any less vocal, and nor is it to imply that they have accepted the changes in the scientific landscape for which they have been partly responsible: they want the changes to be more radical, more enduring. Not to put too finer point on it, they want associationism to be like Monty Python's parrot: an ex-theory. We hope that the contents of this handbook will serve to illustrate that the associative analysis of learning is flourishing, with each chapter highlighting recent advances that have been made by cognitive and behavioral neuroscientists.

The research conducted by cognitive and behavioral neuroscientists uses complementary techniques: ranging from the use of sophisticated behavioral procedures, which isolate key theoretical processes within computational models, to new software tools, that allow vast quantities of imaging data to be rendered in a form that enables changes in neural structures, systems, and their connectivity to be inferred. Some behavioral and neuroscientific techniques are clearly better suited or better developed for some species than others. However, the prospect of understanding the associative process at a variety of levels of analysis and across different species, which was envisaged by previous generations, is now being realized. The chapters in this handbook are intended, both individually and collectively, to provide a synthesis of how cognitive and behavioral neuroscientists have contributed to our understanding of learning that can be said to have an associative origin. To do so, we move from considering relatively simple studies of associative processes in the rat, through to learning involving time and space, to social learning and the development of language. Clearly, the superficial characteristics of the experiences that shape these different forms of learning are quite different, as are the behavioral consequences that these experiences generate. However, there remains the possibility that they are based, at least in part, on the operation of shared associative principles. Where and how these principles are implemented in the brain is an important facet of this handbook. In pursuing answers to these basic questions, of where and of how, we might be forced to reconsider our theoretical analysis of the processes involved

in associative learning. This synergy is an exciting prospect that can only be exploited when a common issue is studied from differing vantage points.

Our hope is that this handbook will also help to bridge some gaps between research that has originated from different philosophical orientations and involved different levels of analysis. Briefly, there is a longstanding division between those who use purely behavioral studies to infer the nature of associative processes and those whose principal interests are in the neural bases of learning and memory. Researchers from both traditions make use of a variety of behavioral measures to draw inference about hypothetical processes, on the one hand, and about the role of various systems, structures, or neuronal processes, on the other. At its heart, the dialog does not concern the legitimacy or rigor of the research that is conducted within either tradition, but rather concerns whether or not the research conducted at one level of analysis or in one tradition provides any information that has utility to the other. Of course, it need not; and it is certainly true that historically there has been surprisingly little crosstalk between researchers from the two traditions – a fact that is likely to constrain the opportunity for productive synergy. We believe that this is a pity and hope that the chapters in this handbook will illustrate, in different ways, how such crosstalk can be mutually beneficial.

The study of associative learning is the application of an analytic technique for describing the relation between the here and now and the there and then, and for how the brain deals with this relation and its contents. It is ultimately a description of how the brain works. A theme throughout the chapters in this volume is the conclusion that where we want to understand the brain's workings, we will need to consider how the brain performs the functions described by associative analysis. To this end, we need both the analytic tools for describing the functions and a description of how these functions are implemented at the level of tissue. We are completely aware that the two levels might look very different but also that a complete description will require both.

The counterargument – that we might understand the brain without the associative framework – can be allied to a similar challenge faced by experts in neurophysiology. Here, the question posed is whether brain imaging (which includes any one of a number of techniques for representing the internal workings of the brain in a visual or mathematical manner) goes beyond simple functional mapping of processes and can be used to uncover how the brain codes experience and communicates this experience. Passingham, Rowe, and Sakai (2013) present a convincing defense of the position that at least one technique, fMRI (a technique for using blood flow to track changes in brain activity) has uncovered a new principle of how the brain works. What is perhaps more interesting for this volume is that the principle in question looks very much like the types of associative processes described herein.

As suggested in Passingham *et al.* (2013), it is quite common, and relatively uncontroversial, to use the technique of fMRI to make claims about the localization of cognitive processes. However, it is more difficult to argue that this or similar techniques have informed our understanding of the principles by which the brain processes information. In the case that Passingham *et al.* identify, fMRI was used to show how processing in area A and processing in area B are related to one another with some types of stimuli or context, but activity in area A is related to area C in another context. They then speculate about how this might be achieved through different subpopulations of neurons being active in area A depending on the context. Students of associative learning

will recognize the issue of how context-dependent stimulus processing is achieved as one that has dominated the recent associative landscape. It has led to the development of various formal models, some bearing more than a passing resemblance to the implementation described immediately above (e.g., Pearce, 1994; Wagner, 2003), that have been subject to experimental testing through behavioral and neuroscientific analyses. This form of integrated analysis is one to which the associative approach lends itself, as the contents of this volume will, we hope, illustrate.

References

Passingham, R. E., Rowe, J. B., & Sakai, K. (2013). Has brain imaging discovered anything new about how the brain works? *Neuroimage*, *66*, 142–150.

Pavlov, I. P. (1941). The conditioned reflex. In *Lectures on conditioned reflexes: conditioned reflexes and psychiatry* (Vol. 2, p. 171). London, UK: Lawrence & Wishart.

Pearce, J. M. (1994). Similarity and discrimination: A selective review and a connectionist model. *Psychological Review*, *101*, 587–607.

Wagner, A. R. (2003). Context-sensitive elemental theory. *Quarterly Journal of Experimental Psychology*, *56B*, 7–29.