Published in final edited form as:

Neurobiol Learn Mem. 2014 February ; 108: 1–4. doi:10.1016/j.nlm.2013.12.006.

The study of associative learning: Mapping from psychological to neural levels of analysis

Andrew R. Delamater and

Department of Psychology, Brooklyn College – City University of New York, United States

K. Matthew Lattal

Department of Behavioral Neuroscience, Oregon Health & Science University, United States,

Fax: +1 503 494 6877

Andrew R. Delamater: andrewd@brooklyn.cuny.edu; K. Matthew Lattal: lattalm@ohsu.edu

Abstract

One of the major achievements of the last century of research in experimental psychology is the identification of a coherent set of theories and principles to characterize the nature of simple forms of associative learning. Major advances are also currently being made at a rapid pace in the neurobiology of associative learning, and, interestingly, we are beginning to see how a mapping from a psychological level of analysis to underlying neurobiological mechanisms is possible. This collection of papers honors the illustrative careers of four major learning theorists from the experimental psychology tradition (Robert Rescorla, Allan Wagner, Nicholas Mackintosh, Anthony Dickinson) who have helped shape our understanding of behavioral principles. The collection of works in this special issue reflects common interests among researchers working at both psychological and neurobiological levels of analysis towards a more comprehensive understanding of basic associative learning processes as they relate to several key issues identified and intensively studied by these influential learning theorists. These consist of the questions regarding (1) the critical conditions enabling learning, (2) the contents of learning, and (3) the rules that translate learning into performance. In one way or another, the separate contributions in this issue address these fundamental questions as they relate to a wide variety of currently exciting topics in the study of the neurobiology of learning and memory.

The study of basic learning processes has a rich and venerable history. Early philosophers, Russian physiologists, and Darwin's evolutionary theory provided the backdrop from which modern day learning theory emerged (e.g., see Boakes, 1984). Several key issues included (a) the importance of experience in shaping learning and behavior (i.e., nature vs nurture), (b) understanding what constituted an explanatory mechanism (e.g., reflex arc conceptions vs functionalist accounts), and (c) delineating how complex behavioral systems evolved. When Pavlov (1927) and Thorndike (1898) first made public their systematic methods for studying the development of conditioned behaviors, the community was extremely excited by the prospects. These two paradigms – what have now become known as classic examples of simple forms of associative learning – allowed the scientist to measure fairly directly the

importance of experience in producing so-called “intelligent” behavior. Further, through its analysis in terms of newly established associative linkages between environmental inputs and behavior, the promise of developing a purely mechanistic understanding of behavior was nearly within grasp. And finally, as one understood more fully the mechanistic capacities and their functional significance across species, one could see how a truly effective comparative analysis of behavior might develop (as opposed to earlier approaches based on the collection of anecdotal evidence), at least from the perspective of an emerging experimental psychology (see Bitterman, 1975).

In retrospect, it should come as no great surprise that the better part of the 20th century was devoted to fully examining at a behavioral level of analysis some of the key psychological principles that underlie simple Pavlovian and Thorndike's instrumental conditioning. This was done with different motivations, however. On the one extreme, scientists were of the view that the study of behavior represented its own scientific level of analysis (e.g., Skinner, 1938, 1950; Tolman, 1932, 1936). It was thought that understanding the functional relationships between environmental constraints and behavior was, at this level, all one needed to do to fully predict, gain control over, or understand behavior. Whether or not one attempted to reduce from this level to the level of underlying brain mechanisms would not change the fundamental importance of discerning the more molar principles that could be used to deduce in any circumstance what behaviors might arise (e.g., see also, Hull, 1943; Spence, 1956; Thorndike, 1911). Although this approach sometimes acknowledged the contribution of brain physiology to behavior, neurobiological processes were not always viewed as the antecedent cause of behavior.

In contrast, others were of the view that the study of learning at the behavioral level was a tool that could be used to obtain the ultimate goal, which was to understand the nature of underlying brain mechanisms. Pavlov (1932) in his famous “Reply of a physiologist to psychologists” was quite insistent that his analysis of the conditioned reflex – which by today's standards would be regarded largely as a behavior-level description – was useful only to the extent that one could infer the nature of the underlying neural mechanisms involved. Indeed, Konorski (1948; 1967) later developed this theme more fully and, like Pavlov before, imagined behavior to be a kind of microscope into underlying brain mechanisms (see also Hall, 2002).

A major challenge in understanding the brain mechanisms of learning during the time of Pavlov and Konorski was the absence of tools to manipulate neurobiological circuits that were thought to underlie learning. Gross manipulations, such as regional ablation, could implicate brain regions in certain aspects of behavior, but the absence of reversibility made it difficult to draw firm conclusions about a region's involvement in specific learning processes. The 1960s saw a rapid increase in the development of pharmacological approaches that allowed for the regional, temporal, and cellular specificity that was needed to begin to understand how brain changes relate to changes in learning (see Davis & Squire, 1984). These approaches have continued to be refined and today we have an array of tools available to understand how specific genes and proteins contribute to learning. In recent years, these neurobiological approaches have allowed scientists to explain at the neuronal

level some of the concepts that have been so historically important in developing associative learning theories since the time of Pavlov and Thorndike.

It is within this spirit that we today think of an “associative analysis” as helping us make sense of the intricate complexities of the variety of changes that can be observed to take place at the levels of individual neurons, synapses, or neural circuits. To a large extent, associative analyses and neurobiological analyses have proceeded in parallel, with major discoveries in one field not having influence on the other for many years. This, in part, is due to the tradition in which learning theorists and neurobiologists have been trained. Associative learning theory is strongly rooted in the field of psychology, with theories being influenced by behavioral, cognitive, social, developmental, and evolutionary psychological approaches. Neurobiological approaches to learning often come from very different perspectives, with leading researchers coming from the fields of molecular biology and genetics. We believe that these two approaches need each other. In the case of learning theory, a purely top-down approach that avoids cellular and molecular mechanisms would mean that one may miss important distinctions between seemingly similar psychological processes that are mediated by different brain systems or molecular processes. In the case of neurobiology, a purely bottom-up approach that avoids theoretical concepts at the level of psychological processes places one at a severe disadvantage in reaching a cogent understanding of the psychological functions performed by various neural circuits within which neural plasticity at various levels would be observed to occur. Ultimately, we are interested in psychological functions – sensations, thoughts, memories, feelings, etc. – and characterizing those neural changes in such terms requires a firm understanding of the psychological functions being tapped by the behavior in question. This is the most powerful argument to favor the kind of theorizing that Pavlov and Konorski initiated.

The latter 30 years of the 20th century saw major advances in our understanding at the behavioral level of some of the key psychological mechanisms at work in simple Pavlovian and instrumental learning. Some of the major contributors, in our view, to this development were Robert Rescorla, Allan Wagner, Nicholas Mackintosh, and Anthony Dickinson. Each of these scientists brought to the discipline a level of sophistication in psychological theorizing and experimental methodology that had not often been seen in the preceding century of studies on animal learning and behavior. As a result we now know a considerable amount more about the basic psychology at work in various learning tasks, and the tradition that these authors helped promote will, we suspect, continue to spawn new insights. Some of these theorists' contributions have now been supported through neural systems level analyses (witness the success of the “prediction error” concept as a driving force between associative learning (e.g., Rescorla & Wagner, 1972) and neural plasticity (e.g., Schultz, 1998)), and, as a result, it is becoming possible to see how a mapping from a psychology-level understanding to underlying neural mechanisms can occur. At the same time, as our discipline advances we also suspect that we will observe important constraints on psychological theorizing with discoveries about the nature of the underlying neural systems at work. In the early and middle part of the 20th century, particularly during the heyday of Behaviorism, such a two-way across-level fertilization of information would not have even been conceivable, but such is the case today.

One purpose, then, of this special issue on the “Associative Perspectives on the Neurobiology of Learning and Memory” is to celebrate the careers of each of these four exceptional scientists. Each one of them has closed down their laboratories within the last several years and is now in various phases of retirement. That their approaches, as well as their specific ideas, continue to provide guidance and offer insights can easily be seen in the various papers contained within these pages.

Much of the last 40 years of research in the field of “associative learning” has been concerned chiefly with three basic questions. These include (1) ‘what are the critical conditions required to support associative learning?’, (2) ‘what is the content or structure of associative learning?’, and (3) ‘what are the mechanisms involved in translating associative learning into observable performance?’ Formulating the key problems of associative learning in this way (e.g., see Mackintosh, 1983; Rescorla & Holland, 1976) goes a long way towards specifying criteria for any complete understanding of the learning process. Moreover, it gives us a good way to organize research findings in this field. The present collection of authors was chosen because each one of these investigators has spent a considerable amount of time thinking about issues both at the associative level of analysis as well as at the level of underlying neural systems. It is our hope that the specific collection of papers included here will illustrate that an effective approach to understanding basic learning processes will include some integration of ideas at both the psychological and neural systems levels of analysis.

We start this special issue with the paper that expands on the integrative approach that we have attempted to delineate above. Fanselow, Zelikowsky, Perusini, Rodriguez Barrera, and Hersman review isomorphisms between psychological and neurobiological concepts, focusing on how the idea of processing of stimulus elements has driven theories at both levels of analysis and using this idea to describe the activity of amygdala neurons during learning. This paper nicely leads into the rest of the special issue, which we have organized into five clusters of related papers. The first papers focus on understanding the conditions that cause learning to occur, with an emphasis on prediction error and unconditioned stimulus processing mechanisms. The second series of papers also focuses on the conditions that cause learning, emphasizing the conditions that cause learned behavior to be extinguished. The third series focuses on understanding the content of learning – what is it that the animal learns when certain experimental contingencies are arranged? The fourth series focuses on understanding how learning is expressed in performance. Finally, we end this special issue with a series of papers that extend learning theories, principles, and methodologies to other domains. Of course, many of the papers outlined below fall into multiple categories and each makes a contribution beyond the simple heuristics that we have used in our groupings.

The first series of papers focus on understanding the conditions that cause learning. Perhaps the most important discovery in the last 50 years in learning theory is that learning proceeds as a function of the discrepancy between the predicted and obtained outcome. The amount of learning on a given trial follows the size of this prediction error, with large errors resulting in large amounts of learning and small errors resulting in little learning. Li and McNally review the importance of prediction error in fear conditioning and how this key

psychological concept has changed the way the field views the function of certain neurobiological systems. McDannald, Jones, Takahashi, and Schoenbaum describe results from studies that manipulate prediction error in various ways to determine how the orbitofrontal cortex is involved in reward expectancy processes. In an empirical report, Young, Dreumont, and Cunningham show how decreasing prediction error by pharmacologically reducing processing of the unconditioned stimulus can impair learning, consistent with a neurobiological mechanism for prediction error coding.

The second series of papers also focuses on the question of circumstances that produce learning. Prediction error is important not just for acquiring new learning, but also for inhibiting that learning. When the prediction error is negative (generally speaking, whenever the expected outcome is more than the obtained outcome), associative strength is thought to decrease. The most common way to study negative prediction error is extinction – when the CS is presented in the absence of the predicted US, the response is weakened. Delamater and Westbrook review findings that validate the importance of prediction error in driving the changes in learning that occur during extinction, while at the same time review studies that suggest there are at least two distinct types of negative prediction errors that drive extinction learning. They consider both neurobiological and theoretical mechanisms. Todd, Vurbic, and Bouton review some key findings showing that extinction in Pavlovian and instrumental preparations shares many common properties. Abraham, Neve, and Lattal describe how dopamine, which is most commonly associated with reward processes, may mediate extinction of fear through appetitive-aversive interactions of the sort envisioned by Konorski and Dickinson. In an empirical report, Panayi and Killcross dissociate the role of the orbitofrontal cortex in within- and between-session extinction in appetitive Pavlovian procedure. They place their findings within the context of several influential associative and nonassociative accounts of extinction. Finally, Maren describes results from fear conditioning experiments in which extinction may be appropriately thought of in terms of habituation and recovery from extinction in terms of dishabituation (or disinhibition) processes.

The third series of papers focuses on the question of the content of learning – what is it that an organism learns after experiencing certain environmental contingencies? An important theoretical issue for behavioral and neurobiological analyses of Pavlovian conditioning is determining whether organisms encode stimulus situations in terms of configurations or elements. This issue has been extremely important for both associative learning theories (e.g., Pearce, 1987) and neurobiological theories (e.g., Fanselow, 1999; Rudy, 2009). In this issue, Honey, Iordanova, and Good describe how to disentangle these two processes at the level of brain and behavior. Their focus is on behavioral approaches using sensory pre-conditioning and complex discriminations to understand the associative content and neurobiological representation of configural and elemental learning, and the particular role of the hippocampus in episodic-like memory encoding. The question of content of learning also has been critical for the development of theories of instrumental learning. Hart, Leung, and Balleine review evidence that instrumental performance is mediated by different associative processes and that these processes are represented by distinct striatal pathways that are coordinated and modulated by the basolateral amygdala.

The fourth series of papers focuses on the question of how learning is expressed in performance. Of the three questions posed by Rescorla and Holland (1976), the question of how learning maps into performance is perhaps the most difficult to answer. In general, neurobiological approaches to learning and memory are often quick to assume that the absence of behavior reflects the absence of learning, but we know that this is often not always the case. In a series of papers, Miller and colleagues have developed a performance model of associative learning phenomena, the latest instantiation of which is known as the “sometimes competing retrieval” theory (SOCR), and have claimed that what we often consider to be a deficit in learning may actually reflect a deficit in expression of learning that can be overcome with the right conditions for performance. Witnauer, Urcelay, and Miller provide simulation results of a number of published empirical findings based on SOCR and contrast these with simulation results from other associative models in reaching the conclusion that their performance model provides a better fit to existing behavioral data. The results bring into question (1) how to properly incorporate performance mechanisms into our learning theories and neurobiological approaches, and (2) how the traditional prediction error concept can be reconciled with performance-based processes that make different assumptions concerning prediction error. Gallistel and Balsam also suggest that much of the way that modern neuroscience conceptualizes how learning occurs is incomplete, at best. They suggest that some of the key concepts in the neurobiology of learning, such as synaptic or associative strength, fail to capture adequately the temporal dynamics of associative learning. Further, these authors offer an information-theoretic approach to learning, where, to a large extent, organisms are assumed to store events within a temporal memory system and where behavior is assumed to reflect a set of decision processes based on computations performed on the raw data within this temporal memory system.

We end this special issue with a series of papers that extend learning theories, principles, and methodologies into different domains of research and application. The analysis of associative learning is known for its rigorous experimental designs. One of the key concepts that has emerged from the associative learning tradition is the importance of not confounding the conditions for learning with the conditions for performance. This approach is nicely illustrated in an empirical report by O’Neale, Ellis, Creton, and Colwill who apply the learning-performance problem to the analysis of habituation in developing zebrafish. Grau describes how many of the key behavioral and neurobiological criteria for learning can be met in the study of Pavlovian conditioning in the spinal cord, a domain not typically associated with high-level learning theory. This demonstrates that the capacity to learn exists in even relatively simple systems, meaning that apparent differences in learning between systems may not reflect qualitatively distinct processes. Davidson, Sample, and Swithers review how an approach that is strongly rooted in learning theory has led to some novel insights about obesity. Finally, McLaren, Forrest, McLaren, Jones, Aitken, and Mackintosh end the special issue with a two-process theory of human learning. In view of current ideas to the contrary, they make a strong case to suggest that learning in humans occurs through the operation, on the one hand, of basic associative processes that, essentially, is preserved across species, as well as, on the other hand, more flexible propositional processes involved in inferential reasoning.

We believe the papers in this special issue provide a nice survey of current thinking in the neuroscience of associative learning theory. They lay the groundwork for the future of learning theory and provide a starting point for discovery of unifying principles.

Acknowledgments

The authors gratefully acknowledge Ted Abel for suggesting that this special issue be assembled, and for his support and encouragement in its development. In addition, the Elsevier production and editorial staff has been extremely helpful. Finally, we would like to thank Bob Rescorla, Allan Wagner, Nick Mackintosh, and Tony Dickinson for their inspiring careers and for helping us to see the way forward. Preparation of this article was supported by NIH Grants DA 034995 to ADR and DA018165 and DA025922 to KML.

References

- Bitterman ME. The comparative analysis of learning. *Science*. 1975; 188:699–709. [PubMed: 17755167]
- Boakes, RA. From Darwin to behaviourism: Psychology and the minds of animals. Cambridge [Cambridgeshire], New York: Cambridge University Press; 1984.
- Davis HP, Squire LR. Protein synthesis and memory: A review. *Psychological Bulletin*. 1984; 96:518–559. [PubMed: 6096908]
- Fanselow MS. Learning theory and neuropsychology: Configuring their disparate elements in the hippocampus. *Journal of Experimental Psychology: Animal Behavior Processes*. 1999; 25:275–283.
- Hall, G. Associative structures in Pavlovian and instrumental conditioning. In: Gallistel, C.; Pashler, H., editors. *Stevens' handbook of experimental psychology Learning, motivation, and emotion*. Vol. 3. New York: John Wiley & Sons Inc.; 2002. p. 1-46.
- Hull, CL. *Principles of behavior: An introduction to behavior theory*. New York: D. Appleton-Century Company; 1943.
- Konorski, J. *Conditioned reflexes and neuron organization*. Cambridge: University Press; 1948.
- Konorski, J. *Integrative activity of the brain, an interdisciplinary approach*. Chicago: University of Chicago Press; 1967.
- Mackintosh, NJ. *Conditioning and associative learning*. Oxford [Oxfordshire], New York: Clarendon Press, Oxford University Press; 1983.
- Pavlov, IP. *Conditioned reflexes, an investigation of the physiological activity of the cerebral cortex*. London: Oxford University Press; 1927.
- Pavlov IP. The reply of a physiologist to psychologists. *Psychological Review*. 1932; 39:91–127.
- Pearce JM. A model for stimulus generalization in Pavlovian conditioning. *Psychological Review*. 1987; 94:61–73. [PubMed: 3823305]
- Rescorla, RA.; Holland, PC. Some behavioral approaches to the study of learning. In: Rosenzweig, MR.; Bennett, EL., editors. *Neural mechanisms of learning and memory*. Cambridge, MA: MIT Press; 1976. p. 165-192.
- Rescorla, RA.; Wagner, AR. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In: Black, AH.; Prokasy, WF., editors. *Classical conditioning II: Current research and theory*. New York: Appleton-Century-Crofts; 1972. p. 64-99.
- Rudy JW. Context representations, context functions, and the parahippocampal–hippocampal system. *Learning & Memory*. 2009; 16:573–585. [PubMed: 19794181]
- Schultz W. Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*. 1998; 80:1–27. [PubMed: 9658025]
- Skinner, BF. *The behavior of organisms*. New York: Appleton-Century-Crofts; 1938.
- Skinner BF. Are theories of learning necessary? *Psychological Review*. 1950; 57:193–216. [PubMed: 15440996]
- Spence, KW. *Behavior theory and conditioning*. New Haven: Yale University Press; 1956.
- Thorndike EL. *Animal intelligence. An experimental study of the associative processes in animals*. *Psychological Review*. 1898; 5:551–553. Abstract.

- Thorndike, EL. Animal intelligence; experimental studies. New York: The Macmillan Company; 1911.
- Tolman, EC. Purposive behavior in animals and men. New York: Century; 1932.
- Tolman, EC. Operational behaviorism and current trends in psychology. In: Tolman, EC., editor. Collected papers in psychology. Berkeley: University of California Press; 1936.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript