

Associationism

Robin A. Murphy¹

¹University of Oxford

MIT Press

Published on: Sep 24, 2026

URL: <https://oecs.mit.edu/pub/xyhzydtg>

License: [Creative Commons Attribution-NonCommercial 4.0 International License \(CC-BY-NC 4.0\)](https://creativecommons.org/licenses/by-nc/4.0/).

Associationism proposes that most (if not all) animal—including human—understanding, and the ability to act upon that understanding, is the result of learned mental associations. Associations can be thought of as theoretical connections between remembered or imagined experiences. Although associationism may seem a limited psychological perspective on the contents and mechanisms of thought, research suggests it is richer than might be expected. Associations may reflect everything we know: the strength of the learning of an experience, the physical intensity of experience, the summation of similar previous experiences, and the attention devoted to those experiences as well as their perceived value. From these connections, individuals construct an understanding of their environment, helping them understand how things relate to one another, and this provides the data employed to interact, explore, decide, and achieve goals. Associationism explains how a dog acts and learns regarding the sights, smells, and sounds relevant for dinner, just as it does for a person. The goals for a psychological theory of association are threefold. First, they aim to describe the conditions that encourage association formation. Second, they aim to describe the content of those associations, what is remembered and linked in the mind. Third, theories of association aim to explain how associations translate into action. Finally, moving beyond theory to describe the workings of the brain, some associationists attempt to relate the dynamics and content of associations to brain cells (i.e., neurons) and their connections.

History

Associationism is perhaps the oldest characterization found in writings on the nature of memory. Aristotle proposed a correspondence between experience, memory, and thought driven by association ([Beare, 2010](#); [Sorabji, 2006](#)). When someone perceives two things, they are likely to remember those two things as associated. Aristotle argued that association is supported by four conditions: similarity, contrast, contiguity, and frequency of occurrence. David [Hume \(1789\)](#) proposed the more radical idea that associations do not simply relate memories, but they form the basis of complex understanding. The sense humans have of detecting the most likely cause of an event requires nothing more than relying on previous association. According to Hume, humans never perceive causation directly; rather, causal learning is a species of associative learning [see [Causal Learning](#)]. Further, John Stuart Mill (1806–1873; [Ball, 2016](#)) suggested that there might be rules of association, a form of “mental chemistry” that explains how associations interact in predictable ways. Chemistry is an especially useful analogy, given that chemical elements can combine to make compounds with novel properties, as when gaseous oxygen and hydrogen combine to make liquid water. A general theory of association might thus show how complex thoughts with novel properties might arise from the combination of the simple atoms of thought and their associations.

Modern scientific theories of association took inspiration from this early work but were advanced by discoveries using microscopy that visualized the mental filaments that inhabit the brain. Santiago Ramón y Cajal, who claimed that, “to know is to associate” ([Ramón y Cajal, 1954](#)), based his observations on the physical associative structure of the cerebral cortex—which, according to Ramón, was “made of

association cells” [see [Neuroplasticity](#)]. His demonstration that neurons in the central nervous system have directionality suggested to him that they could form an associative network to house thought. In the 20th century, this study of the interface between neuronal action and mental association became strongly related to the work of Ivan [Pavlov \(1927/2010\)](#). Pavlov studied how food caused digestive responses (e.g., salivation, gastric secretions) in dogs. However, once Pavlov’s dogs had experience with his experimental apparatus, the dogs were as likely to salivate to food in their mouth as they were to the anticipation of food if they simply saw, smelled, or heard the stimulus of the experimental apparatus. Pavlov outlined how much of his dogs’ behavior could be accounted for by the learning acquired during *Pavlovian conditioning*. Any stimulus is an event that generates a perceptual reaction via sight, taste, hearing, smell, and touch with the functional reaction to any stimulus being the encoding of that experience. During conditioning, an initially neutral stimulus, like a brief sound (the *conditioned stimulus*), could be frequently paired contiguously with a biologically relevant stimulus, like meat powder (the *unconditioned stimulus*), in a way that the conditioned stimulus could come to cause responses, like salivation (a *conditioned response*), initially only caused by the unconditioned stimulus. This work demonstrated an empirical approach for how principles of experience could generate long-lasting changes in behavior and was proposed to also explain human behavior.

Core concepts

Preparedness

As research explored the range of flexible associations that can be formed to support behavior, doubts emerged regarding whether nature provided anything other than the general associative architecture of the brain. Localization of function in the brain suggested minimally that certain types of knowledge might reside easily in different brain locations. Nature may have prepared the brain to deal with certain types of experience before the experience occurred. A further set of discoveries found that some associations were more easily formed than others. Sensory modalities were not equivalent in their ability to support associations. For example, tastes are more easily associated with subsequent nausea than sounds, whereas sounds are more easily associated with pain than tastes ([Garcia & Koelling, 1966](#)). The conclusion was that some perceptual modalities provide the basis for similarity and allow some associations to be more easily formed than others.

Inhibition

Although the notion of associations encoding real experiences may seem a plausible idea, associations can also form with the specific *absence* of a stimulus. For example, although associating a particular virus with illness involves direct causal understanding, it is also possible to learn that a vaccine inhibits that outcome. Acquiring an association between the stimulus (e.g., vaccine) and the absence of illness requires a different type of association, one that suppresses rather than activates the representation for illness. A mental representation for the absence of an event as an association requires elaborating the conditions that support inhibition and the content of that association. Pavlov demonstrated inhibition by showing

that dogs could learn to associate a neutral stimulus to the absence of a second stimulus (e.g., food). Like the virus and vaccine example, an association is first created between one stimulus (e.g., a tone) and the associate (e.g., food). Having learnt this first association, the subsequent presentation of the tone and a new stimulus (e.g., a light) together, but without the food, violates the expectation for food that the tone association generates. This violation defines the light as a signal for the absence of otherwise expected food. Conditioned inhibition is important because it cannot be explained simply by looking at what the animal is currently perceiving; it requires combining information over time.

Similarity and strength

For association to be more than the simple memory of two events co-occurring, they need to be flexibly activated, recovered, and combined. Similar experiences contribute to single association, but any two experiences are never exactly the same. If a tone signals stomach illness, how similar does a new experience of the tone and illness need to be in order to be combined with the previous learning rather than treated as a novel experience? Clearly, any two instances are never the same (even the same object presented twice does not look exactly the same, given lighting, viewing angle, and distance, etc.), so it raises the following question: What conditions support treating two different experiences as similar? One simple answer is that similarity is related to the physical properties of the stimuli ([Pearce, 1987](#)). For instance, if an animal learns that an intense or loud tone is associated with food, varying the intensity of the tone might be expected to elicit a different response—or the same response if tone frequency is the defining associative variable. Associations are activated by similar stimuli but often with some decrement in activity (what is called the *generalization decrement*). The further the stimulus departs from the originally trained stimulus in both frequency and intensity, the weaker the conditioned response is likely to be ([Van Osselaer et al., 2004](#)).

Configurations

One stimulus may enter multiple associations with many other stimuli, which allows chains or networks of associations. Bread may become associated with tomato (bread → tomato), and tomato may become associated with lettuce (tomato → lettuce). Because tomato is not restricted to a single association, it may set up the activation of lettuce following the presentation of bread; a new idea may emerge from separate but related experiences. The mental concept of tomato then has two associates or meanings following from experience.

Once multiple associations are formed in a mental network, it is also possible that accumulated experience could set up opposing associations. This means a stimulus could be associated with both the presence and absence of its associate. The idea that a stimulus can have more than one meaning, including opposing meanings, is demonstrated clearly with a *negative patterning discrimination*. Consider how two different stimuli, a tone and a light, might independently be trained as associates for food. If both are presented simultaneously, the experimenter's expectation might be that the subject anticipates food presentation or even that twice as much food might be expected. However, if training includes

explicit experience that the compound (tone and light together) is never followed by food even if each element alone is followed by food, both meanings will be retained. Here, the tone is both relevant for anticipating food and not anticipating food. The mechanism for solving a negative patterning discrimination seems to require that the content of the association contains a unique representation for the configuration of the two separately trained events as well as their compound. Although negative patterning discriminations illustrate the flexibility of associations, they also demonstrate how associations can be formed by complex configurations of stimuli. Associations between stimuli can define the collective meaning of an object or stimulus or can support stimuli having different meanings when in different combinations [see [Face Perception](#)].

Blocking

The failure of simple principles of contiguity and frequency to predict learning is highlighted by selective learning experiments demonstrating that contiguity is impacted by what the organism has already associated. For instance, the phenomenon of *blocking* is an example of how contiguity fails to predict association formation because of the presence of another associated stimuli ([Kamin, 1969](#)). If two stimuli are presented simultaneously with food, an animal will learn that each stimulus is independently associated with food; however, if either of the two stimuli have previously been paired with food in the absence of the other, then that prior association can “block” the new association that would have formed to the unpreexposed stimulus. Contiguity as a principle of learning will fail if this new experience is linked to an event that is already activated by other learnt associations (e.g., [Murphy et al., 2001](#)). These effects are referred to as selective associations; prior experience automatically selects what to learn without any directed or deliberate decision. The meaning or value of a stimulus is not based on its absolute experience but rather its experience relative to other experienced stimuli.

The failure of contiguity as a general condition for learning inspired the development of new theories of learning. The Rescorla–Wagner model, for example, explains both association and attentional phenomena like blocking and the similar relative validity effect as the result of a competition between associates ([Rescorla & Wagner, 1972](#)). Like the blocking effect, the relative validity effect demonstrates that it is not the absolute value of a predictor for its consequences that determines the association formed, but rather associations form on the basis of relative value. On the surface, selecting information on the basis of relative importance seems a complex cognitive function. The Rescorla–Wagner model demonstrated that this function could emerge from a simple learning algorithm.

Questions, controversies, and new developments

Associationism is an attractive and perhaps beguiling theory because it provides a simple yet powerful explanation of thinking and behavior for biologically relevant actions such as eating. However, it is also proposed as a useful starting point for understanding memory, attention, concept formation, causal reasoning (e.g., [Baker et al., 1996](#); but see [Gallistel, 2007](#)), and perhaps even all aspects of human thought, like language [see [Language](#)]. The consideration of these ideas has led to enhanced theories of association,

and how to constrain the theory to be useful scientifically has been an important question ([Dickinson, 2001](#)).

The search for underlying neural mechanisms that support associations is also ongoing. Neuroscientists connected association to what they saw in the brain, but the details describing the content of associations in neural terms remain elusive. Modern theories of synaptic plasticity and neuroplasticity seek to explain how the phenomenon of association can be implemented in neural tissue [see [Neuroplasticity](#)].

Finally, the unique personalities of individuals seem too difficult to explain via a general-process associative account ([Byrom & Murphy, 2018](#)). For instance, two similar people can both be bitten by the same dog with similar vigor, but the appearance of a dog phobia in one person and not the other is a challenge to associative theories. This difference can be attributed to genetic or dispositional differences that remove the effect from an associative explanation. Another possibility is that previous associations influence learning during the potentially phobic experience. If all humans possess the same learning mechanism and most people have somewhat similar daily experiences, it can be difficult to understand why they seem to diverge in so many ways. It may be that individual differences and a person-centered psychology reflect everyone's willful contribution to their psychology. Alternatively, an individual's unique differences emerge because of the complex, lifelong experience that creates a unique network of associations unique to each person. The dog bite is the same absolute stimulus but is different relative to previous associations. Further research requires mapping how complex experience contributes to individual differences in thinking and behavior as well as the neurochemical substrates that support associations.

Broader connections

Theories of association have been important for the development of theories of *reinforcement learning* in both psychology and computer science (e.g., [Kang et al., 2024](#); [Sutton & Barto, 1999](#)) [see [Reinforcement Learning](#)]. Reinforcement learning is a mechanistic model of learning in which associations can be strengthened or weakened by a reward signal. It is a very general theory that shows how teacherless systems can produce intelligent behavior.

A long-standing dispute within psychology has been the extent to which simple theoretical entities like associations can account for the most complex cognitions such as reasoning or language. Historically, research into language and language acquisition proposed that there was not enough primary linguistic input for an associative process to be sufficient to capture features of syntax [see [Language Acquisition](#)]. The conclusion drawn was that nature provided specific linguistic tools to generate language. However, with the advent of large language models, this view has been challenged [see [Large Language Models](#)]. Because large language models are trained in a broadly associative way, associative processes might be more powerful than has been supposed. Whether these models can capture the full range of human experience is currently under study ([Wang et al., 2025](#)). Similarly, complex abstract concepts like human creativity or our sense of justice might seem outside the scope of an associative theory. However,

associations between, say, the events and comparative value of stimuli related to justice-perceived experiences and the consequences of those experiences might allow even complex, abstract concepts to be amenable to an associative account.

Further reading

- Dickinson, A. (2012). Associative learning and animal cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1603), 2733-2742. <https://doi.org/10.1098/rstb.2012.0220>
- Rescorla, R. A. (1988). Pavlovian conditioning: It's not what you think it is. *American Psychologist*, 43(3), 151-160. <https://doi.org/10.1037/0003-066X.43.3.151>
- Wasserman, E. A., & Castro, L. (2022). The Rescorla-Wagner model: The culmination of Hume's theory of causation. *Journal of Experimental Psychology: Animal Learning and Cognition*, 48(4), 315-320. <https://doi.org/10.1037/xan0000325>

References

- Baker, A. G., Murphy, R. A., & Vallée-Tourangeau, F. (1996). Associative and normative models of causal induction: Reacting to versus understanding cause. In D. R. Shanks, K. Holyoak, & D. L. Medin (Eds.), *Causal learning* (pp. 1-45). Academic Press.

↩

- Ball, T. (2016). Psychology, associationism, and ethology. In C. Macleod & D. E. Miller (Eds.), *A companion to Mill* (pp. 143-159). Wiley.

↩

- Beare J. I. (2010). On memory and reminiscence Aristotle (ca. 350 b.c.). *Annals of Neurosciences*, 17(2), 87-91. <https://doi.org/10.5214/ans.0972-7531.1017208>

↩

- Byrom, N., & Murphy, R.A. (2018). Individual differences are more than a gene x environment interaction: The role of learning. *Journal Experimental Psychology: Animal Learning & Cognition*, 44(1), 36-55. <https://doi.org/10.1037/xan0000157>

↩

- Dickinson, A. (2001). Causal learning: Association versus computation. *Current Directions in Psychological Science*, 10(4), 127-132. <https://doi.org/10.1111/1467-8721.00132>

↩

- Gallistel, C. R. (2007). Flawed foundations of associationism? Comment on Machado and Silva (2007). *American Psychologist*, 62(7), 682-685. <https://doi.org/10.1037/0003-066X.62.7.682>

↩

- Garcia, J., & Koelling, R. A. (1966). Relation of cue to consequence in avoidance learning. *Psychonomic Science*, 4(1), 123-124. <https://doi.org/10.3758/BF03342209>

↩

- Hume, D. (1789). Of probability; and of the idea of cause and effect. In L. A. Selby-Bigge (Ed.), *A treatise of human nature* (pp. 73-78). The Clarendon Press.

↩

- Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior* (pp. 279-296). Appleton-Century-Crofts.

↩

- Kang, P., Tobler, P. N., & Dayan, P. (2024). Bayesian reinforcement learning: A basic overview. *Neurobiology of Learning and Memory*, 211, 107924. <https://doi.org/10.1016/j.nlm.2024.107924>

↩

- Murphy, R. A., Baker, A. G., & Fouquet, N. (2001). Relative validity effects with either one or two more valid cues in Pavlovian and instrumental conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, 27(1), 59-67. <https://doi.org/10.1037//0097-7403.27.1.59>

↩

- Pavlov, P. I. (2010). Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex. *Annals of Neurosciences*, 17(3), 136-141. <https://doi.org/10.5214/ans.0972-7531.1017309> (Original work published 1927)

↩

- Pearce, J. M. (1987). A model for stimulus generalization in Pavlovian conditioning. *Psychological Review*, 94(1), 61-73. <https://doi.org/10.1037/0033-295X.94.1.61>

↩

- Ramón y Cajal, S. (1954). *Neuron theory or reticular theory? Objective evidence of the anatomical unity of nerve cells*. Consejo Superior de Investigaciones Científicas.

↩

- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II* (pp. 64–99). Appleton-Century-Crofts.

↩

- Sorabji, R. (2006). *Aristotle on memory* (2nd ed.). University of Chicago Press.

↩

- Sutton, R. S., & Barto, A. G. (1999). Reinforcement learning. *Journal of Cognitive Neuroscience*, 11(1), 126-134.

[↩](#)

- Van Osselaer, S. M., Janiszewski, C., & Cunha, M, Jr. (2004). Stimulus generalization in two associative learning processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(3), 626-638. <https://doi.org/10.1037/0278-7393.30.3.626>

[↩](#)

- Wang, R., Liu, J., Qian, C., Shen, Y., Pan, Y., Xu, Z., Abbasi, A., Ji, H., & Zhang, D. (2025). Rescorla-Wagner steering of LLMs for undesired behaviors over disproportionate inappropriate context. In C. Christodoulopoulos, T. Chakraborty, C. Rose & V. Peng (Eds.), *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing* (pp. 19821-19856). Association for Computer Linguistics.

[↩](#)