# Different learning architectures contribute to value prediction in human cognition

Frédéric Alexandre

HAL Id: hal-01931492

https://inria.hal.science/hal-01931492

Submitted on 22 Nov 2018

# Different learning architectures contribute to value prediction in human cognition

Frédéric Alexandre[1,2,3]

[1]Inria Bordeaux Sud-Ouest, 200 Avenue de la Vieille Tour, 33405 Talence, France
[2]LaBRI, Université de Bordeaux, Bordeaux INP, CNRS, UMR 5800, Talence, France
[3]Institut des Maladies Neurodégénératives, Université de Bordeaux, CNRS, UMR 5293, Bordeaux, France
frederic.alexandre@inria.fr

**Abstract.** Pavlovian conditioning plays a fundamental role in our cognitive architecture, by its capacity to bind values to stimuli. Due to the multifarious characteristics of this learning mode, many approaches in machine learning have been proposed to implement it in artificial intelligence models. Considering the complementary properties of these models and inspired by biological evidences, we propose not to select the best model but rather to combine them, thereby forming a cognitive architecture. From a functional point of view, we report the good properties and performances of this architecture. From a methodological point of view, this work highlights the interest of defining a cognitive function at the algorithmic level, binding its general properties to its implementation details. It is also proposed that it is a fruitful approach to decipher and organize many information extracted by modern approaches in neuroscience, towards the definition of a global cognitive architecture.

**Keywords:** Value prediction, Bio-inspiration, Competiton, Cognitive architecture

## 1 Introduction

Having particularly in mind the brain performing a cognitive function, D. Marr [28] has argued that, to be fully understood, an information processing system must be examined at three different levels of description. The function itself can be globally described at the computational level. The representations and the processes used to implement the function are described at the algorithmic level. The way of computing them are described at the implementational level. This later level, refering to the brain, corresponds to the brain circuitry but Z. Pylyshyn [35] has proposed a similar tri-level description in a purely cognitive purpose and has simply defined the lower level as the functional architecture.

The decomposition is interesting because it makes it clear that not a unique algorithm can carry out the function and not a unique circuit can implement the algorithm. This decomposition is also interesting in the framework of artificial

intelligence because it helps structuring steps to build a model. On one side is the cognitive function to model. On the other side are formalisms available (e.g. neuronal computations). In between are probably the more critical and important steps: defining the structures and the sequences of information processing to bridge both sides (e.g. neuronal architecture and learning rules). The classical approach in artificial intelligence is top-down: from a cognitive function to be modeled, an algorithm and its implementation in a certain formalism are proposed. The approach can be also bottom-up, particularly if the problem is set in computational neuroscience: from experimentally observed neuronal circuits and mechanisms, it is questioned which ones are active to emulate the cognitive function generally associated to the neuronal region.

In this paper, having in mind this decomposition, I extend a series of works performed in my team to model pavlovian conditioning from different points of view, in order to propose a new view of this fundamental learning mode, with an impact on cognitive modeling as well as artificial intelligence and machine learning.

## 2   Pavlovian conditioning

Pavlovian conditioning [33] is an elementary learning process, present in most species, by which an animal learns to predict biologically significant stimuli announcing pain or pleasure (called Unconditioned Stimuli, US, e.g. some food) by some other stimuli (called Conditional Stimuli, CS, e.g. a bell ringing before food delivery) that anticipates its venue.

This learning paradigm has been extensively studied in behavioral neuroscience because it is easy to observe (US anticipation is associated to other so-called pavlovian responses like salivation in the appetitive case and freezing in the aversive case) and not so simple as it might appear: Many experiments have demonstrated that pavlovian learning is not a simple associative process measuring the contingencies between the US and the CS [24]. For example, when an association is learnt and extinguished, it is more rapid to be learnt again. Some authors have proposed that during the process, latent causes are extracted [13]. Others have stressed the role of the context [39]. As a results, after a century and thousands of publications, many data have been accumulated, many models have been proposed but pavlovian learning remains not fully understood.

It is also important to underline that pavlovian conditioning has a prominent role in cognitive modeling and in the domain of Artificial General Intelligence, because it gives the capacity to link values to objects and events of the environment and this capacity is important in a cognitive-oriented description of intelligent behavior. For example, in a general description of cognition [4], one central module is concerned with declarative memory, providing facts and knowledge about the world, in which goals can be defined and sensorimotor associations can be elaborated, thus forming other central modules. All these modules heavily rely on the elementary capacity to predict aversive and appetitive values for CS from the innate detection of US. Some cues to elaborate the CS might come

from declarative memory and the US correspond to the main behavioral goals for most animals. This capacity is then extended to conditioned reinforcers [7], corresponding to intermediate goals like, in humans, tools or money, and at the basis of categorization [5] and of judgments [22] used in strategies and more generally in decision making and planning.

In this paper, we propose to study pavlovian conditioning from two different points of view. On the one hand, from a top-down machine learning point of view, we will wonder how it is possible to associate the CS and the US. In other words, the question is to associate the perception of external stimuli like the CS, also called exteroception, to the internal perception of pain or pleasure, also called interoception, as it is the case with the US. We summarize some more or less classical strategies at the algorithmic level in the next section.

On the other hand in computational neuroscience, we have modeled some neuronal structures and circuits that are reported to play a fundamental role in pavlovian conditioning. This particularly concerns the amygdala, a primitive neuronal structure in the medial temporal lobe and its main afferences. We summarize in the following the main results of our study in a bottom-up way and explain how we integrate these pieces of information towards a more global understanding of pavlovian conditioning, at the algorithmic level and how this can be more generally profitable for cognitive modeling.

## 3    Modeling pavlovian conditioning in Machine Learning

In the domain of machine learning, pavlovian conditioning has often been associated to supervised learning. In its simplest aspect, it corresponds, when a stimulus is perceived, to predict the occurrence of the US and the learning can consequently be supervised by the fact that, some milliseconds later, the US is perceived or not. Even if some binary models have been proposed, predicting the magnitude of the US (for example the size of the reward or the intensity of a painful electric shock) is the main purpose of the learning rule. In that aspect, it is interesting to note that one of the main learning rules proposed for pavlovian conditioning by Rescorla and Wagner [36] is similar to the main learning rule setting the bases of supervised learning in neural networks, proposed by Rosenblatt for the Perceptron [38]. They both set the focus on a multiplicative term, called error for the perceptron and surprise for pavlovian learning, corresponding to the difference between the predicted and the computed US. This term (null if there is no error, thus ensuring the stability of the rule) also explains that there is no learning when the US is already predicted, including by other CS. Importantly, this has been observed in experimental situations in a famous paradigm called blocking, increasing the interest for this rule.

Nevertheless, for the same kind of structural reasons as the perceptron, the Rescorla-Wagner learning rule has some limitations, particularly related to the complexity of the decision that can be taken in the input space. Concerning the perceptron, it has been shown that its learning rule guarantees the convergence towards a unique minimum value of error, but this results in a good quality of

the learnt solution only in case of linear separability in the input space [30]. Concerning pavlovian conditioning, it has been observed that some mammals can answer to some combination of two variables (x and y) called positive (x AND y) et negative (x XOR y) patterning, differently than the corresponding combined effect of each variable, raising the idea to create configural representation (representing the co-occurrence of x and y), able to provide a different effect. This principle of configural learning proposed in [39] resulted in that paper in the same idea of adding hidden units as in the multilayer perceptron, to propose a model of pavlovian conditioning able to perform such complex decision.

The idea of performing error-driven learning in multilayered networks has been made possible in Machine Learning by the back-propagation learning algorithm which allows to make emerge internal representations in the hidden layers. Even if some biologically plausible learning rules have also been proposed [31], the main visible effect of this algorithm is rather today in the huge developments about deep learning [25], which make the learning of any complex association affordable by a deep network. Nevertheless, as it can be seen in the difficult steps of parameter tuning in these networks, there is no proof of convergence in this case and, instead of looking for the best solution, mastering this technology rather corresponds to obtaining a satisfactory solution in a satisfactory time length.

It is also often forgotten that the only strong theoretical result about layered network is the fact that removing one layer can require to exponentially increase the size of the previous layer [16]. To tell it differently, it is possible to approximate any separation with a unique hidden layer and a random fixed connection from the input to the hidden layer (but possibly with a very large hidden layer). This has for example been experimentally studied in [6], showing that the function learnt by complex deep networks can be efficiently approximated by a simple network with a unique (and large) hidden layer. It is only supposed that the projection to the hidden layer is likely to make emerge the useful internal representations, to be linearly separated for the targeted decision. A strategy when the corresponding knowledge is available, is to directly send the pertinent information to the hidden layer, else the solution is to use random projections and the possibility to create pertinent information becomes very likely when the layer is very large. In summary, the main advantage of deep networks is to have what is called a compact representation (because requiring less neurons), when the representation is elaborated hierarchically, as compared to a flat representation, but this is at the price of a less stable representation.

In a completely different view, it is often neglected in Machine Learning that associative memories can provide a kind of supervised learning. Associative memories are these kinds of recurrent networks, like the Hopfield network [18], that are able to learn "by heart" some prototypes as stable states and, in a recall phase, to consider a new input and stabilize on the closest state stored beforehand. This process is called auto-association, since the network is shown to be able to associate a pattern (the new input) with itself (seen as the closest prototype learnt by heart before). This is generally considered useful in Machine

Learning when the new input has been corrupted by noise and when the recall is a way to denoise this input.

This process can easily be extended to hetero-association. In this case, just consider a series of $i$ vectors $V_i$ as the concatenation of subvectors $X_i$ and $Y_i$ and consider that these vectors are learned as prototypes in an associative memory. The principle of this learning process is that a noisy $V_j$ will be reconstructed to the closest $V_i$ learned before. If, in the vector $V_i$, the noise corresponds to set $Y_i$ to zero and let $X_i$ unchanged, then the recall will reconstruct the missing part $Y_i$ and will consequently act in hetero-association, which can be seen as a kind of supervised learning, even if in this case there is no generalization and the network converges towards a pattern already stored before.

Several other characteristics of this learning process must be underlined. On the one hand, it is interesting because it is very rapid and presenting only one time a prototype is sufficient to memorize it (as opposed to layered perceptrons learning slowly). On the other hand, it can be considered weak because it is subject to interference (close prototypes can create spurious states combining these prototypes) and saturation (learning too many prototypes can lead to a catastrophic forgetting) [12].

## 4  Modeling pavlovian conditioning in Computational Neuroscience

Concerning the implementation of pavlovian conditioning in the brain, the amygdala is a neural structure generally considered to play a central role in this process [26]. The structure is composed of a set of nuclei, including the basolateral nucleus for sensory inputs and the central nucleus expressing pavlovian motor responses. Observing direct plastic connections between these sensory and motor nuclei [11] is a good argument to propose a shallow layered architecture to model this network. Particularly, the lateral part of the basolateral nucleus has been reported to represent elementary CS, selected in its afferences from the thalamus and primary sensory cortex [26], whereas the basal part receives afferences from more integrated regions like associative sensory cortex and hippocampus [14].

A functional analogy with a model like the simple perceptron can be also evoked for the amgdala, since signals corresponding to error of prediction have been observed in the network [27]. Such error signals are known to be carried by neuromodulators, which can act at the learning and functioning levels in the amygdala but also in other cerebral structures. This is the case for acetylcholine, considered to tag errors due to stochasticity in the environment [42] and acting to balance the activity between the lateral and basal sensory nuclei [8], but also to trigger memorization in the hippocampus [15] and shift selective attention in the higher regions of the sensory cortex [32]. To better understand the putative role of neuromodulation on these structures, let us first evoke the participation of the structures to pavlovian conditioning.

The basal nucleus of the amygdala receives inputs from the associative sensory cortex [26]. In the visual case for example and in a large region going from

the primary visual cortex to the inferotemporal cortex, this path has been described as a layered architecture, extracting more and more abstract cues to create categories for visual recognition [41]. Based on biological observations, it has also been proposed that the release of acetylcholine plays a prominent role for transmitting an error signal, as an attentional signal fed back throughout the architecture and some models have been proposed accordingly, as a biological implementation of a kind of error backpropagation [37].

The hippocampus is also known to be a major input of the basal nucleus of the amygdala [26], generally considered to provide information about the context of the task, that can play an important role in discrimination. This is consistent with the fact that the hippocampus is known to perform arbitrary binding of its inputs, collected in the entorhinal cortex from main sensory regions of the cortex, and to learn these associations through a recurrent network to form a so-called episodic memory [40]. Recently, we have also proposed that the hippocampus might in fact associate interoception and exteroception in its inputs, making it a perfect candidate for learning the vector CS-US in heteroassociation [19] and recalling US when the CS is presented alone. In a more computational study [20], we have also shown that this organization is in fact beneficial to resist to interference.

The complementary roles in learning of the cortex and the hippocampus have already been theoretized [29]. Whereas the cortex is able to learn slowly cues important for discrimination and generalization, the hippocampus has a different learning mode, allowing to learn quickly, by heart, some specific configuration to be remembered. In the case of the amygdala, this should correspond either to extract cues in the cortex to learn some features of the CS or, when this results in an error of prediction, to learn by heart the global sensory configuration in the hippocampus and to store it as a special case to be remembered. It is consequently interesting to remind now that acetylcholine release triggers storing in the hippocampus.

With the aim of integrating all these elements in a unique picture, we begin to report preliminary work we did to model pavlovian conditioning in the amygdala [9]. Our main goal was to evaluate which of the many standard rules proposed for pavlovian conditioning [24] was the most faithful, as compared to biological data. What we proposed, in short, is that several populations with different behavior, recently reported in the basolateral nucleus [17], receive different kinds of cues, each kind reported to be important in some classic rule of pavlovian learning. In addition, we proposed that, consistent with the fact that several populations with different excitatory and inhibitory roles have also been recently described in the central nucleus of the amygdala [11], the behavior actually expressed by the amygdala is in fact the result of the competition between different lines of elementary sensorimotor association in the amygdala. In conclusion, contrarily to rules from experimental psychology, becoming more and more complex as they want to model pavlovian conditioning, taking more and more facts into account, we propose in [9] that the same complexity of behavior can be expressed by the simple competition between different lines of association, underlying the role of

neuromodulation to favor some lines over others, depending on the context or the recent history of behavior [3].

One step further, what we propose in this paper is not to forget that cues arriving in the sensory input of the amygdala are themselves learnt in other regions. This is consequently a wider view, proposing to model pavlovian conditioning in a network of neuronal structures and to consider that critical events in the network can trigger important learning or functioning steps in different locations throughout the network. In this view, as it was already proposed in [9], the amygdala seen as a simple perceptron decides which US to predict from a variety of inputs received in the basolateral nucleus. These inputs can correspond to simple elementary CS coming from the thalamus or to more elaborated hints or configurations learnt in other cortical areas or in the hippocampus. Possibly, a release of acetylcholine will give a stronger weight to complex inputs, meaning that the environment is too noisy to rely on simple cues.

This mode of learning is incremental in the sense that, at the beginning, it can be supposed that hardly no complex cues have been extracted in the hippocampus or in the highest levels of the cortex. Errors of prediction will result in neuromodulation triggering learning in the hippocampus of the specific cases poorly considered before. Later on, if the same case occurs again, it will trigger recall in the hippocampus that will reconstruct the corresponding US and send it back to the amygdala. Concerning learning in the layered architecture of the cortex and in addition to bio-plausible mechanisms proposed in [37] and mentioned above, the phenomenon of consolidation must be also evoked. In this process, also reported in neuroscience and described in [29], the hippocampus can act as a supervisor to the cortex and send it some examples previously learned by heart in episodic memory, in order to transfer it to semantic memory in the cortex. As far as we know, no stable learning rules have been proposed so far to implement this important learning process, but this will be very important to consider in the future because it might allow to remove examples in the hippocampus and prevent it from saturation.

## 5   Discussion

The capacity for an intelligent agent to associate a sensory state (taken in the widest sense, including more or less abstract multimodal and contextual information) to a value (particularly representing its possible contribution to achieving the current goal of the agent) is very desirable to define in a cognitive architecture. It can be paralleled, in natural cognition, with pavlovian conditioning linking a state (more or less abstract stimulus, context) to the unconditioned stimulus it predicts, annoucing a pain or a pleasure. In the living, this learning has been shown to contribute both to declarative and implicit memories [40].

In this paper, we have summarized pavlovian conditioning as the capacity to associate a vector representing the sensory state with a vector representing the anticipated internal state and we have evoked three classical strategies in machine learning to learn this kind of association, each with its respective strengths

and weaknesses. Shallow layered networks ensure a quick processing and optimal learning but only for simple direct associations. In contrast, deep layered networks can learn complex associations, which is particularly useful to elaborate internal representations, concepts and abstractions, but at the price of a very long learning that can fall in local minima and must choose pertinent configurations. Of completely different nature is the elaboration of associative memories in recurrent networks. Prototypes can be learned by heart, possibly in one shot but are sensitive to interference and globally an associative memory can be subject to saturation.

From this global analysis, one strategy is to look for the best compromise; another one is to try to benefit from the strengths of all. In the previous section, we have argued to show that the brain exploits this latter strategy. Whereas pavlovian conditioning can be summarized as an implicit sensorimotor learning taking place in the amygdala and associating simple cues to pavlovian responses related to the predicted US, this cannot explain all the observed pavlovian associations and other inputs to the amygdala carrying other kind of information must be also considered. This is particularly the case for inputs coming from the hippocampus and from high cortical areas and learned and processed in these structures with different criteria. What we have also made clear here is that these structures are also dependent on the processing made in the amygdala and particularly on neuromodulation released from its errors of prediction.

This system of combined learning occuring in different sites in parallel but also for common reasons can be interpreted from a phylogenetic point of view. In primitive animals, only simple associations can be learnt [23]. With evolution, experience from recent episodes and association of multimodal information can be integrated to the pavlovian association.

It is of course also very tempting to extrapolate this interpretation and the related observations to the domains of cognitive modeling and artificial intelligence. Proposing models based on the cooperation of memories is not new in these domains [1, 29] and relates to the design of ensemble or multi-method approaches. This must be more and more seriously considered, since new technologies in experimental neuroscience allow to observe that most evolved cognitive functions are in fact elaborated from an interplay between neuronal structures [34]. In the present case, this gives an elegant way to introduce specific cases to an associative process with no loss of generalisation. Using hidden layers to elaborate configural representations is not mandatory, as it has been reminded that a simple layer could perform the same task (at the price of a larger hidden layer), but it is interesting for using internal representations for other purposes. This hierarchy of hidden layers provides a level of abstraction (not obtained by a shallow network) that could be exploited for other cognitive function and particularly for declarative memory and for executive functions.

We have also mentioned up to which point neuromodulation plays an important role in associating the processing and learning phases in the neuronal structures, and can be seen as a rare ground of coordination between the otherwise parallel processes. In the future, we plan to better understand their functions

[2] and we have already proposed some mechanisms to decipher some temporal aspects of pavlovian conditioning [21] and to go beyond, towards operant conditioning and associated high level cognitive functions [10].

To conclude on a more methodological point of view and coming back to Marr's levels of description [28], we have presented some top-down approaches in Machine Learning and bottom-up approaches in computational neuroscience, forming the state of the art on the current understanding of pavlovian conditioning. We demonstrate here that synthetizing these analyses by the algorithmic level is a unique way to propose new interpretations and to revisit cognitive modeling on this topic.

# References

1. Alexandre, F.: Biological Inspiration for Multiple Memories Implementation and Cooperation. In: In V. Kvasnicka P. Sincak, J. Vascak and R. Mesiar, editors, International Conference on Computational Intelligence (2000)
2. Alexandre, F., Carrere, M.: Modeling Neuromodulation as a Framework to Integrate Uncertainty in General Cognitive Architectures. In: The Ninth Conference on Artificial General Intelligence. New-York, United States (Jul 2016), https://hal.inria.fr/hal-01342902
3. Alexandre, F., Carrere, M., Kassab, R.: Feature, Configuration, History : a bio-inspired framework for information representation in neural networks. In: International Conference on Neural Computation Theory and Applications. Rome, Italy (Oct 2014), https://hal.inria.fr/hal-01095036
4. Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., Qin, Y.: An integrated theory of the mind. Psychol Rev 111(4), 1036–1060 (Oct 2004), http://dx.doi.org/10.1037/0033-295x.111.4.1036
5. Ashby, F.G., Alfonso-Reese, L.A., Turken, A.U., Waldron, E.M.: A neuropsychological theory of multiple systems in category learning. Psychological review 105(3), 442–481 (Jul 1998), http://view.ncbi.nlm.nih.gov/pubmed/9697427
6. Ba, J., Caruana, R.: Do Deep Nets Really Need to be Deep? In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q. (eds.) Proceedings of the Neural Information Processing Systems Conference, NIPS 2014, pp. 2654–2662. Curran Associates, Inc. (2014), http://papers.nips.cc/paper/5484-do-deep-nets-really-need-to-be-deep.pdf
7. Belova, M.A., Paton, J.J., Morrison, S.E., Salzman, C.D.: Expectation modulates neural responses to pleasant and aversive stimuli in primate amygdala. Neuron 55(6), 970–984 (2007)
8. Calandreau, L., Trifilieff, P., Mons, N., Costes, L., Marien, M., Marighetto, A., Micheau, J., Jaffard, R., Desmedt, A.: Extracellular hippocampal acetylcholine level controls amygdala function and promotes adaptive conditioned emotional response. The Journal of neuroscience : the official journal of the Society for Neuroscience 26(52), 13556–13566 (Dec 2006)
9. Carrere, M., Alexandre, F.: A pavlovian model of the amygdala and its influence within the medial temporal lobe. Frontiers in Systems Neuroscience 9(41) (2015)
10. Carrere, M., Alexandre, F.: Modeling the sensory roles of noradrenaline in action selection. In: The Sixth Joint IEEE International Conference Developmental Learning and Epigenetic Robotics (IEEE ICDL-EPIROB). Cergy-Pontoise / Paris, France (Sep 2016), https://hal.inria.fr/hal-01401882
11. Ciocchi, S., Herry, C., Grenier, F., Wolff, S.B.E., Letzkus, J.J., Vlachos, I., Ehrlich, I., Sprengel, R., Deisseroth, K., Stadler, M.B., Muller, C., Luthi, A.: Encoding of conditioned fear in central amygdala inhibitory circuits. Nature 468(7321), 277–282 (Nov 2010), http://dx.doi.org/10.1038/nature09559
12. French, R.M.: Semi-distributed representations and catastrophic forgetting in connectionist networks. Connection Science 4(3-4), 365–377 (1992)
13. Gershman, S.J., Blei, D.M., Niv, Y.: Context, learning, and extinction. Psychological review 117(1), 197–209 (Jan 2010), http://view.ncbi.nlm.nih.gov/pubmed/20063968
14. Goosens, K.A., Maren, S.: Contextual and Auditory Fear Conditioning are Mediated by the Lateral, Basal, and Central Amygdaloid Nuclei in Rats. Learning & Memory 8(3), 148–155 (May 2001)

15. Hasselmo, M.E.: The role of acetylcholine in learning and memory. Curr Opin Neurobiol 16(6), 710–715 (Dec 2006)
16. Hastad, J., Goldmann, M.: On the power of small-depth threshold circuits. Computational Complexity 1, 113–129 (1991), `https://doi.org/10.1007/BF01272517`
17. Herry, C., Ciocchi, S., Senn, V., Demmou, L., Muller, C., Luthi, A.: Switching on and off fear by distinct neuronal circuits. Nature 454(7204), 600–606 (Jul 2008), `http://dx.doi.org/10.1038/nature07166`
18. Hopfield, J.J.: Neural networks and physical systems with emergent collective computational abilities. In: Proceedings of the National Academy of Sciences, USA. pp. 2554–2558 (1982)
19. Kassab, R., Alexandre, F.: Integration of exteroceptive and interoceptive information within the hippocampus: a computational study. Frontiers in Systems Neuroscience 9(87) (2015)
20. Kassab, R., Alexandre, F.: A Modular Network Architecture Resolving Memory Interference through Inhibition. In: et al., J.M. (ed.) Computational Intelligence, Studies in Computational Intelligence, vol. 669, pp. 407–422. Springer (2016), `https://hal.inria.fr/hal-01251022`
21. Kaushik, P.S., Carrere, M., Alexandre, F., Raju, S.B.: A biologically inspired neuronal model of reward prediction error computation. In: 2017 International Joint Conference on Neural Networks, IJCNN 2017, Anchorage, AK, USA, May 14-19, 2017. pp. 3577–3584 (2017), `https://doi.org/10.1109/IJCNN.2017.7966306`
22. L., A.G.: Human contingency judgments : Rule based or associative. Psychological Bulletin 114(3), 435–448 (1993)
23. Laberge, F., Muhlenbrock-Lenter, S., Grunwald, W., Roth, G.: Evolution of the Amygdala: New Insights from Studies in Amphibians. Brain Behav Evol 67(4), 177–187 (2006), `http://dx.doi.org/10.1159/000091119`
24. Le Pelley, M.E.: The role of associative history in models of associative learning: a selective review and a hybrid model. The Quarterly Journal of Experimental Psychology 57(3), 193–243 (Jul 2004), `http://dx.doi.org/10.1080/02724990344000141`
25. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature 521(7553), 436–444 (May 2015), `http://dx.doi.org/10.1038/nature14539`
26. LeDoux, J.: The amygdala. Current Biology 17(20), R868–R874 (Oct 2007)
27. Li, S.S., McNally, G.P.: The conditions that promote fear learning: Prediction error and Pavlovian fear conditioning. Neurobiology of Learning and Memory 108(0), 14–21 (2014)
28. Marr, D.: Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. Henry Holt and Co., Inc., New York, NY, USA (1982)
29. McClelland, J.L., McNaughton, B.L., O'Reilly, R.C.: Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychological review 102(3) (Jul 1995), `http://view.ncbi.nlm.nih.gov/pubmed/7624455`
30. Minsky, M., Papert, S.: Perceptrons: An Introduction to Computational Geometry. MIT Press, Cambridge (1969)
31. O'Reilly, R.C., Munakata, Y.: Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain. The MIT Press, 1 edn. (Sep 2000), `http://www.worldcat.org/isbn/0262650541`
32. Pauli, W.M., O'Reilly, R.C.: Attentional control of associative learning–a possible role of the central cholinergic system. Brain Research 1202, 43–53 (Apr 2008)

33. Pavlov, I.P.: Conditioned Reflexes (V.Anrep, trans.). London: Oxford University Press" (1927)
34. Paz, R., Bauer, E.P., Paré, D.: Measuring correlations and interactions among four simultaneously recorded brain regions during learning. Journal of neurophysiology 101(5), 2507–2515 (2009), `http://dx.doi.org/10.1152/jn.91259.2008`
35. Pylyshyn, Z.W.: Computation and Cognition. MIT Press (1984)
36. Rescorla, R., Wagner, A.: A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In: Classical Conditioning II: Current Research and Theory, pp. 64–99. Appleton Century Crofts (1972)
37. Roelfsema, P.R., van Ooyen, A.: Attention-gated reinforcement learning of internal representations for classification. Neural computation 17(10), 2176–2214 (2005), `http://dx.doi.org/10.1162/0899766054615699`
38. Rosenblatt, F.: The perceptron: a probabilistic model for information storage and organization in the brain. In: Anderson, J.A., Rosenfeld, E. (eds.) Neurocomputing: Fundations of Research (1989), pp. 89–92. The MIT Press (1958)
39. Schmajuk, N., DiCarlo, J.: Stimulus configuration, classical conditioning and the hippocampus. Psychological Review 99, 268–305 (1992)
40. Squire, L.: Declarative and nondeclarative memory: multiple brain systems supporting learning and memory. Journal of cognitive neuroscience 4(3), 232–243 (1992), `http://view.ncbi.nlm.nih.gov/pubmed/23964880`
41. Thorpe, S.J., Fabre-Thorpe, M.: Seeking Categories in the Brain. Science 291(5502), 260–263 (Jan 2001), `http://dx.doi.org/10.1126/science.1058249`
42. Yu, A.J., Dayan, P.: Uncertainty, Neuromodulation, and Attention. Neuron 46 (2005)