Associative versus Predictive Processes in Pavlovian Conditioning

Jérémie Jozefowiez

Université de Lille & Universidade do Minho

Correspond with:
Dr. Jérémie Jozefowiez
Laboratory of Cognitive and Affective Sciences (SCALab)
Université de Lille
Campus de Lille 3
Domaine Universitaire du Pont de Bois
B.P. 60149
59653 Villeneuve d'Ascq Cedex
France
Email: jeremie.jozefowiez@univ-lille3.fr
Phone: +33-320-041-6866
Fax: +33-320-041-6036

Submitted: December 2017

Abstract

While learning and memory are basically the same thing for the layman, they are considered distinct entities in psychology. The concept of association, which is central both to the field of conditioning and to that of retrieval and forgetting, could be used to bridge the gap between the two concepts. However, the concept is quite different in the fields of learning and memory, a situation for which this article argues that the Rescorla-Wagner model is mainly to blame. By viewing Pavlovian conditioning as the outcome of a predictive process but using the traditional associative language developed in memory studies to describe this process, it has introduced an unnecessary confusion between memory and prediction within the field of learning. This confusion needs to be acknowledged so that the concepts of associations and predictions can again be differentiated. This would allow for better integration of the fields of learning and memory.

Keywords: Learning, memory, conditioning, Rescorla-Wagner model, association, prediction.

Associative versus Predictive Processes in Pavlovian Conditioning

"First, then, one must consider what sort of things the objects of memory are, for this often leads people astray. For it is not possible to remember the future, which is instead an object of judgment and prediction. (…) Memory is of the past."

Aristotle, *De Memoria et reminiscientia* (trans. 1972)

1. Introduction

In psychology, learning and memory are studied by two different research communities. The reasons for this separation are mainly historical. For want of oversimplification, the learning tradition stems from the works of Thorndike and Pavlov and focuses on conditioning phenomena in non-human animals, while the memory tradition, stemming from the work of Ebbinghaus, focuses on verbal learning in human participants. However, at some point, these two subfields of psychology will have to merge. For anybody but the academic psychologist, learning and memory are so obviously related that it is hard to see how the understanding of one could proceed without an understanding of the other.

Within the learning tradition, a few attempts have been made to examine conditioning within the wider context of the study of memory. Wagner's (1981) SOP model explains conditioning by using an associative model, as is common in theories of conditioning, and the general architecture for memory proposed in cognitive psychology by Atkinson & Shiffrin (1968). Miller's comparator hypothesis (e.g. Denniston, Savastano & Miller, 2001) accounts for cue competition in Pavlovian conditioning in terms of interference in memory. Bouton (e.g. Bouton, 1993), and even more insistently Miller (e.g. Miller & Escorbar, 2002; Pollack, Jozefowiez & Miller, 2017), have argued that many basic conditioning phenomena, such as extinction, should be studied as instances of the interference phenomena studied in the memory literature. While I thoroughly agree with the need to better integrate studies on conditioning with studies on memory, I would like to draw attention herein to a potential obstacle to the successful integration of these two research fields.

2. Associative accounts of conditioning and memory

A good starting point for integrating research on memory and on conditioning is the concept of association, which is central to both fields. Ever since Aristotle, association has been critical to the understanding, if not of memory, at least of two of its most fundamental manifestations: remembering and forgetting. In a famous passage of his *Remembrance of things past,* French writer Marcel Proust tells us how the taste of a very specific cake, a madeleine, brought back to his mind strong and vivid memories of his childhood. Why did the taste of the madeleine bring back these specific memories and not others? The *only* explanation ever proposed for this basic phenomenon assumes the existence of an association between the two events. In the parlance of modern cognitive psychology, an association exists between the retrieval cue (the taste of the madeleine in Proust's case) and the target memory (the childhood event that Proust remembered), thus allowing the retrieval cue, when presented, to retrieve the target memory.

An association is not a thing but a property of a retrieval cue relative to a target memory. Its only intrinsic property is its strength, which is the likelihood that the presentation of the cue will retrieve the target memory. Both successes and failures of memory retrieval are explained by reference to the strength of the association. Whether a memory is retrieved depends on its level of activation which is itself a function of the retrieval cues currently presented to the subject and the strength of the association between them and the target memory. Oversimplifying a bit, the level of activation of a memory is the sum of the associative strength of all the retrieval cues currently present (see, for instance, Collins & Loftus, 1975 for a model explicitly built around those principles). As described by Baddeley (2015): "Retrieval, then, is a progression from one or more cues to target memory, via associative connections". Likewise, associations explain forgetting: interference occurs when a retrieval cue becomes associated with several target memories (Anderson & Neelie, 1996).

Of course, association is also a key concept in theories of conditioning. Pavlov (1927) explained the development of the conditioned response (CR) to the conditioned stimulus (CS) by assuming that the pairing of the CS with the unconditioned stimulus (US) created an association between the representation of the CS and that of the US, thereby allowing the former to retrieve the latter when presented. With a few exceptions (for instance, see Mitchell, De Houwer & Lovibond, 2009 or Balsam, Drew, & Gallistel, 2010), contemporary theories of Pavlovian conditioning rely on the same assumptions. As summarized by Pearce & Hall (1980), "(i)t is usually assumed that an association is formed between the central representation of the CS and US so that activation of the first (…) arouses activity appropriate to the likely occurrence of the second (…). In consequence, the "associative strength" of the CS has become a central concept in classical conditioning theory, and the concern of the theorist has been largely to specify how various procedural manipulations work to determine this strength".

Hence, the two fields seem to rely on a similar concept: association. In such a perspective, the way conditioning relates to memory seems very straightforward, i.e. association is the building block of memory and research on conditioning attempts to reveal the conditions in which associations are created between two events.

3. Predictive accounts of conditioning

If this is the research program for the study of conditioning, it has been at least partially fulfilled by researchers since the days of Pavlov. For instance, the physical characteristics of the CS and US, notably their salience, have been shown to be critical. More important are the number of CS-US pairings and the spatial and temporal contiguity between them (see Escorbar & Miller, 2004 for reviews). In general, the role of these variables had been anticipated by the philosophical speculations of the British empiricists and their followers.

This was not so much the case for a variable whose importance started to be acknowledged only at the end of the sixties. Kamin (1968) showed that conditioning failed to occur if a CS was paired with the US in the presence of another CS already paired with the same US (the blocking effect). Rescorla (1968) demonstrated that conditioning was not only a function of the probability of the US in presence of the CS but also of the probability of the US in the absence of the CS (the contingency effect). These results along with many others showed that the predictive value of the CS relative to the US was just as important a determinant of conditioning as variables already identified by previous research, i.e. conditioning occurred only if the CS was a non-redundant reliable predictor of the US.

This led to a paradigmatic revolution in the study of conditioning. The process underlying conditioning was now seen as providing the basis for the ability of organisms to predict the future. Not only did this make sense of the various cue competition phenomena, such as blocking or the contingency effect, that were discovered at the time but it also provided a useful framework for understanding the topography of the CR. The behaviorist stimulus-response (S-R) theory of conditioning considered the CR as simply the unconditioned response (UR) triggered by the US now controlled by a new stimulus. This failed to account for situations where the CR and UR are topographically different, as in the case of CS paired with drug injections where the CR is usually the opposite of the UR (e.g. Mansfield & Cunningham, 1980; Siegel, 1975). This was not a problem for the new predictive account which held that the CR is an adaptive anticipatory response emitted in anticipation of a US to help the organism cope with it. This accounts both for situations where the CR and UR are similar and for those where they are different, as well as for changes in the topography of the CR with several variables, such as the CS-US interval (e.g.

Akins, 2000; Akins, Domjan, & Gutierrez, 1994).

Nevertheless, this radically new understanding of the process underlying conditioning did not lead to major changes at the theoretical level. Taking its inspiration from the Bush-Mosteller model of instrumental responding (Bush & Mosteller, 1951), the influential Rescorla & Wagner (1972) model, which provides the roadmap for all contemporary thinking on conditioning, stayed true to Pavlov's S-S account, i.e. conditioning was still about the learning of an association between the CS and the US allowing the latter to retrieve a representation of the former. The difference was that the rule for the formation of associations was updated so that the association would change only when the organism was surprised by the occurrence of the US following the CS. The change in the association always occurred in such a way that the organism would be less surprised by the next US occurrence (error-correction learning).

More precisely, let's write V_i the strength of the association between CS *i* and the US. On a specific trial *n,* the subject is presented with a set of CS *S(n). V(n)*, the activation of the US representation on that trial is the sum of the associative strength of the all the CSs present on that trial [$V(n) = \sum_{i \in S(n)} V_i$]. The actual intensity of the US on trial *n* is $\lambda(n)$. The associative strengths of all the CSs belonging to *S(n)* are then modified according to the equation

$$\Delta V_i(n) = \alpha_i \beta [\lambda(n) - V(n)] \qquad (1)$$

where $\Delta V_i(n)$ is the change in the associative strength of CS *i* on trial *n* while α_i and β are learning rate parameters respectively affected by the salience of CS *i* and of the US.

4. Central thesis

The similarity between the Rescorla-Wagner model and associative accounts of memory such as Collins & Loftus (1975) are clear: *V(n)* corresponds to the level of activation of a memory (in this case, one related to the US); it is the sum of the associative strength of the CSs acting as retrieval cues. Hence, despite the new emphasis on prediction, the Rescorla-Wagner model seems to have left intact the old associative account of conditioning and the way it relates to the study of memory.

The central thesis of this article is that this is not the case. Equation (1) fundamentally alters the concept of association to the point that it does not correspond anymore to the concept of association used to account for retrieval in memory. Because of the paradigmatic nature of the Rescorla-Wagner model, this altered concept of association carries over to most contemporary models of conditioning. This is hidden by the fact that the same associative language is used by both students of conditioning and students of memory, hence creating an unnecessary confusion between two different cognitive processes: memory and remembering on one hand, and prediction and expectation on the other.

For most researchers studying conditioning, this has little importance as they do not try to connect their work to research on memory. But it becomes critical when one tries to articulate the relation between the study of conditioning and the study of memory. In my opinion, it constitutes one of the major roadblock to a successful integration of those two fields of research.

5. Memory and prediction

To reach this conclusion, let's start with the obvious: remembering the past and predicting the future are two different cognitive processes. A cognitive event cannot be both the retrieved memory of a past event and, *at the same time*, the expectation of a future event. Otherwise, you would not be able to tell the future from the past as the same mental event could be either one of them (more likely, you would not even have a concept of past and future). Obviously, this is not the case as you can always tell whether you are remembering the past or expecting the future. Viewing a dark sky, I might either remember that it rained yesterday or expect that it will rain in a short time but the *same* mental event cannot be both my memory of yesterday's weather and my expectation of the upcoming one.

Indeed, most instances of daily remembering have nothing to do with prediction. I can bring back to mind memories of a unique event (my Ph.D. defense for instance, or the wedding of my best friend). Prediction does not play any part of such instances of remembering.

On the other hand, predictions rely on memory but are not identical to them. Imagine you visit a new restaurant for the first time and that the meal is absolutely delicious. The two events (the restaurant and the meal) are now associated in your memory: thinking of one (perhaps hearing the name of the restaurant) will bring back the other (the memory of the meal). The next time you decide to go out for lunch, remembering that meal might lead you to expect a similar experience if you visit that restaurant again. If you act on that prediction, visit the restaurant again and instead find that the food is atrocious, this will not erase the memory of the initial meal, which you will still be able to remember. But the fact that you can now recollect two discrepant memories regarding the quality of the food in that restaurant will certainly affect your future predictions regarding the quality of the food there. In this example, your ability to predict the quality of a meal at this restaurant depends on your ability to remember the quality of a past meal at this restaurant or at similar places. In other words, in this example, your ability to predict the future depends on your ability to remember the past. Though related, they are clearly different processes.

6. Associations in the Rescorla-Wagner model: memory or prediction?

A corollary of the fact that remembering and predicting are two different phenomena is that it is not possible to account for both of them with the same theoretical concept. As we have seen, association is the key concept to explain remembering. Hence, it cannot be used at the same time to account for prediction. Yet, this is exactly what the Rescorla-Wagner model seems to do: it provides an explanation of the various predictive phenomena evidenced in Pavlovian conditioning (blocking, contingency effect, etc.) within the context of the associative account of memory already used to frame older theories of conditioning. This is clearly not possible: again, a single process cannot account for both memory and prediction. Hence, either the Rescorla-Wagner model does not deal with prediction, or it does. But if it does, it does not deal with memory. If so, the concept of association in the Rescorla-Wagner model is different from the one used to explain remembering.

There is little doubt that the Rescorla-Wagner model deals with prediction. Rescorla & Wagner (1972) themselves acknowledge it: “The central notion suggested here can also be phrased in somewhat more cognitive terms. One version might read: organisms only learn when events violate their expectations. Certain expectations are built up about the events following a stimulus complex; expectations initiated by that complex and its component stimuli are then modified when consequent events disagree with the composite expectation” (Rescorla & Wagner, 1972, p. 75). In Equation (1), it is clear that $V(n)$ is the “composite expectation” which builds up “about the event following a stimulus complex”, the later corresponding to $\lambda(n)$.

But even if Rescorla and Wagner did not agree with this, it would not matter as this interpretation is imposed by the very logic of Equation (1). This equation describes how the associative strengths are modified according to surprise, which is represented by the difference $[\lambda(n) - V(n)]$. By definition, surprise is the difference between what is expected and what actually happens. The latter is $\lambda(n)$, the actual intensity of the US. This means that $V(n)$ cannot be anything but the expected intensity of the US. But, if $V(n)$ is the prediction of the US intensity, it cannot be the activation of a memory of the US.

This conclusion has implication for the individual associative strength V_i as $V(n)$ is nothing but a sum of the associative strength. From a theoretical point of view, it is not clear why summing associative strengths would lead to a prediction of the US intensity except if the strength of the CS-US association can be interpreted as being a prediction of the intensity of the US in the presence of the CS. In this case, it is perfectly understandable that summing the associative strength of all the CSs present in the environment will lead to an aggregate prediction of the US intensity. Moreover, the strength of a CS-US association will converge on the ΔP value existing between these two stimuli (e.g. Rescorla & Wagner, 1972; Danks, 2003). ΔP, the difference between the probability of the US in the presence of the CS and its probability in the absence of the CS, is an objective measure of the predictive value of the CS regarding the US. Again, this makes it difficult to interpret the strength of the CS-US association as nothing but the prediction of the US intensity in the presence of the CS.

Hence, the Rescorla-Wagner model is undeniably a predictive account. This implies that it is not a memory account. The associative language it uses is a vestigial remnant of its origin in associative accounts of conditioning. This is unfortunate because the meaning of an association has been altered by Equation (1) to the point that the concept of an association in the Rescorla-Wagner model has nothing in common with the older concept still used by memory researchers. Associations in the Rescorla-Wagner model are predictions in disguise.

Another way to show that the Rescorla-Wagner model does not deal with memory is to point out that, according to the model, an association between a CS and a US is created only if the CS is a non-redundant reliable predictor of the US (every associative model of conditioning since makes the same assumption, with the notable exception of Miller's comparator hypothesis. See, for instance, Denniston et al, 2001). Applied to remembering, this would imply that a stimulus can act as a retrieval cue for a memory only if it is a non-redundant reliable predictor of the event represented by the memory. Though, to my knowledge, there has been no systematic experimental investigation of this issue, it does not fit with everyday observation of the way our memory works. A stimulus can clearly evoke a memory even when there is no predictive relation between them. Contiguity, not contingency, seems to link a retrieval cue to the memory it brings back to mind in most instances of everyday remembering. The very fact that we can remember accidental co-occurrence of two events without concluding that they are linked in a causal or predictive fashion points in the same direction as well as evidence that associations in memory are bidirectional (if E1 is a retrieval cue for a memory of E2, then E2 is a retrieval cue for a memory of E1, even if E1 temporarily precedes E2. See, for instance, Ebbinghaus, 1885 or the review in Murdock, 1973, chapter 5). For sure, associative models of memory such as the one by Collins & Loftus (1975) are built with the implicit assumption that only contiguity matters. Hence, the variable necessary for the creation of an association between the CS and the US according to the Rescorla-Wagner model is not necessary for the creation of associations in memory. The conclusion is again that the Rescorla-Wagner model does not deal with memory and that the concept of association in the Rescorla-Wagner model does not correspond to the concept of association used in research on memory.

7. Disentangling memory and prediction

The discussion has focused on the Rescorla-Wagner model, despite the fact that there are other models of conditioning available which often contradict it on some important points (for instance, Denniston et al, 2001; Mackintosh, 1975; Pearce & Hall, 1980; Wagner 1980) and despite the many empirical challenges it faces (see Miller, Barnet & Grahame, 1995 for a review), because it has a special status within the field of conditioning. The Rescorla-Wagner model is much more than an important theory. It is truly a Kuhnian paradigm (Kuhn, 1962) in the sense that it has set the question on which researcher should focus and the concepts that can be used to answer those questions. Part of its legacy is the almost exclusive focus in current research on cue competition phenomena (to the detriment of other variables such as temporal factors) and the emphasis on prediction and the role of surprise as well as much

Deleted: ; Denniston et al., 2001

closer attention given to context effect. Moreover, beyond the field of conditioning, the extension of the Rescorla-Wagner model by Sutton and Barto underpins the field of reinforcement learning (RL. See Sutton & Barto, 1998 for a review and an introduction and Sutton & Barto, 1990 for the links between RL and the Rescorla-Wagner model) in artificial intelligence, which aims at creating learning algorithms for adaptive artificial agents. The discovery by Wolfram Schultz that the midbrain dopamine neurons, whose role in reward learning is well-known, might implement the surprise learning signal postulated by the Rescorla-Wagner model and reinforcement learning algorithms (see reviews in Schultz, 2015) has made the idea of surprise-driven error-correction learning a central concept in contemporary behavioral neurosciences.

The altered concept of association is also part of the legacy of the Rescorla-Wagner model. It is the least noticeable as the change from the old concept of association to the new one of association as prediction is concealed and implicit. Comparing the Rescorla-Wagner model to previous models, it seems that, besides the rules for updating associative strengths (Equation 1), nothing has changed while, in reality, everything has changed. This has far reaching consequences. It creates an unnecessary confusion by conflating the concept of association with that of prediction. It makes it look as if the model is saying something about association, i.e. about memory, while it is really talking about something totally different, i.e. prediction.

Let me take two examples. The Rescorla-Wagner explanation of cue competition phenomena such as blocking is often contrasted with those provided by, for instance, Miller's comparator hypothesis by saying that, while the later explains cue competition in term of a failure of retrieval, the former does it in term of a failure of encoding. Encoding, retrieval: all those terms refer to memory phenomena. But the Rescorla-Wagner model deals with prediction: *it has nothing to say about memory*. Claiming that the model views cue competition as an encoding is a perfect example of the kind of confusion that it has created in the researchers' mind by blending memory with prediction.

The second example is even more telling as it concerns a claim, found in many articles and textbooks, according to which the Rescorla-Wagner account of extinction as a form of unlearning (i.e. as a loss of the associative strength between the CS and the US) is inadequate because it fails to account for recovery phenomena, such as spontaneous recovery, renewal or reinstatement. This is correct only if the associations in the Rescorla-Wagner model are assumed to be associations in the old fashioned sense of the term, that it is to say something that represents the ability of the CS to retrieve a representation of the US. If they are recognized for what they *really* are, i.e. predictions of the US intensity, then the Rescorla-Wagner model's account of extinction is perfectly adequate. Certainly, during extinction, the prediction of the US intensity in presence of the CS goes down to 0. Whether this is because memories are erased or not is out of the scope of the model. It just appears not to be the case because the model somehow has tricked us into thinking it deals with memory by inappropriately couching its predictive account in the language of associations.

Associations are not predictions. We need to disentangle the concept of association from the one of prediction and stop talking about association when really what we mean is prediction. A first step in the right direction would be to strip the Rescorla-Wagner model of its associative language. This language is not essential to the model. The reader can try the following experiment for himself: whenever the model uses the words "association" or "associative", substitute the words "prediction" and "predictive". I think he will find out that nothing is lost in the process. On the contrary, once we start speaking of "predictive strength" instead of "associative strength", the real scope of the model becomes clearer and the confusion between memory and prediction is avoided. A similar analysis should also be carried out for other models of conditioning.

This is important because Pavlovian conditioning is probably about prediction. The case made by Rescorla and Wagner that, in a Pavlovian task, the animal is trying to predict the US based on the CS is extremely compelling and has become even more so with new discoveries made since Rescorla and Wagner's original proposal. This implies that, if Pavlovian conditioning is about prediction, it is about memory in the sense that, in some cases, predictions depend on associations (remember the restaurant example) but not only and certainly not in the straightforward way suggested by Pavlov's S-S account. That means that to successfully integrate research on conditioning with research on memory, we need to identify which part of a conditioning phenomenon is due to the memory process, and which part is due to the predictive process. For instance, is blocking a failure to remember or a failure to predict? This is only possible if we acknowledge that these two processes are not the same. I would go further and argue that any model where two processes, one corresponding to memory, the other to prediction, cannot be identified is necessarily false or at least, seriously incomplete.

This is why the confusion introduced by the Rescorla-Wagner model between memory and prediction is roadblock to a successful synthesis of research on memory and research on conditioning. It frames things in such a way that it is not even possible to ask the question: a failure to remember is necessarily a failure to predict and vice versa.

However, if Pavlovian conditioning is about predictions, there is also another possibility: it might on occasions have nothing to do with memory at all, at least if by memory, we mean remembering, that is to say the form of memory for which the concept of association is highly relevant. This is because there are at least three different ways in which an organism might be able to predict the future and only one of them, the one illustrated in the restaurant example, is based on retrieved episodic memories.

Another way would be for the organism to run a kind of mental simulation of the process it is trying to predict. Think of when you are trying to predict how your spouse might react to something you are about to do, of a scientist trying to figure out what a model predicts in a specific experimental situation or maybe of a rat navigating through a maze using a cognitive map à la Tolman. One could argue that the mental model is still stored in memory from which it is retrieved, but that would be more in semantic than in episodic memory and the memory retrieval part would be the least important part of the process. Hence, the concept of association would be much less relevant to understanding predictions generated in this way.

Finally, a mechanism where predictions are directly generated in the presence of a stimulus without any mediation from retrieved memories or a mental model is still possible. This is how some of the artificial agents that learn to interact with their environment through RL work (Sutton & Barto, 1998). The two previous mechanisms for predictions rely on either episodic or semantic memory, i.e. explicit declarative memory. Perhaps more implicit procedural forms of learning, as in some instances of motor learning, would rely more on direct predictions.

8. Related proposals

This three-fold classification of mechanisms for prediction mirrors that for RL algorithms recently proposed by Gershman & Daw (2016). As already indicated, RL algorithms allow artificial agents to learn through interactions with their environment how to behave optimally in it. Though focusing on decision-making problems, the first RL algorithms, like TD learning (Sutton, 1988. This is the algorithm used to model the learning signal generated by dopamine neurons) or Q-learning (Watkins & Dayan, 1991), actually took their clues from the Rescorla-Wagner model of which they are real-time extensions. Such algorithms are used with agents performing direct predictions. When confronted with a stimulus, the agent generates a prediction of the amount of reward that should follow the emission of all the behavior it can emit in presence of that stimulus. These predictions are then updated by the RL algorithm in a Rescorla-Wagner fashion according to the mismatch between the predicted consequences and the actual ones. This kind of model-free algorithm contrasts with subsequent model-based algorithms (such as the Dyna algorithm proposed by Sutton, 1991) where the agent builds an internal model of the environment which it uses to simulate runs in the environment on which its choices will be based. This corresponds to the mental model strategy for prediction described above.

Following these two well-established classes of RL algorithms, Gershman & Daw (2016) advocated the need for a third known as episodic RL, in which predictions are based on the remembrance of past interactions with the environment. Of course, this corresponds to the memory-mediated mechanism for prediction discussed in the restaurant example. Their proposal is based not only on formal grounds, as they see episodic RL as providing a potential solution to some of the computational shortcomings of model-free and model-based algorithms, but also on recent experimental results showing the importance of retrieved episodic memories in human decision-making. For instance, Ludvig, Madan, & Spetch (2015), Madan, Ludvig, & Spetch (2014) and Bornstein, Khaw, Shohamy, & Daw (2015. see also Wimmer & Buchel, 2016. Gershman & Daw, 2016 provides a more extensive review) have shown that reminding participants of a specific outcome associated with a choice option using a stimulus prime influences the choice for that option. In the study by Ludvig et al. (2015), participants became more risk-prone after being reminded of the greater gain they could obtain from a risky option and more risk-averse when reminded of the lower gain they could get from that same option. Murty, et al. (2016) showed that choices between lotteries were affected by the amount of money the participants received during a one-off encounter with those lotteries only when they could explicitly remember that encounter. These

Deleted: (see Gershman & Daw, 2016 for a more extensive review)

Deleted: FeldmanHall, Hunter, Phelps, & Davachi developments parallel work in behavioral economics that also emphasizes the impact of retrieved episodic memories on individual choice (availability heuristics: Tvesky & Kahneman, 1973; decision by sampling: Stewart, Chater, & Brown, 2006; contingent sampling: Biele, Erev, & Ert, 2009; instance theory: Gonzalez & Dutt, 2011)

9. Conclusion

Memory and conditioning are obviously related and an integrated approach to these two phenomena is certainly needed. It is clear that, as Ralph Miller has repeatedly advocated (e.g. Miller & Escobar, 2002; Pollack, Jozefowiez, & Miller, 2017), many fundamental conditioning phenomena can be very usefully conceptualized as instances of memory interference. Also, the concept of association is central to theoretical treatments of both conditioning and memory.

But, as I tried to argue in this paper, the relation between conditioning and memory might be more complicated than initially expected, and acknowledgment of that complexity is necessary for a successful integration of the two research fields. If conditioning is about predictions, as most of us has come to believe since the advent of the Rescorla-Wagner model, then association and remembering are only part of the story. Sometimes, they might not be even part of the story at all: predictions of the US might be achieved through mental simulations or direct predictions, without any mediation from episodic memory. There might even be situations where the reverse is true: the development of the CR could entirely be explained by an associative process, without any reference to a predictive one. Such a case has been made, for instance, for evaluative conditioning (e.g. De Houwer, Thomas, & Baeyens, 2001) as well as for taste preference and aversion (e.g. Higgins & Rescorla, 2004; Myers & Sclafani, 2006). This ties to another point I have been arguing elsewhere (Jozefowiez, 2014): it is unlikely that a single process explains the development of the CR to the CS in all Pavlovian experimental preparations. Predictions can be achieved in many ways.

Identifying how they are achieved by a specific organism in a specific situation and how it relates to memory processes might be necessary to clarify the relations between conditioning and memory. All variables affecting conditioning might not be affecting memory. Variables affecting memory in a specific Pavlovian preparation might affect another process in a different Pavlovian preparation.

One could oppose two arguments to this conclusion. The first would be to deny that the Rescorla-Wagner model is about prediction: it is just an equation whose terms should not be interpreted or, if they are, we should be aware that the terms we use in those interpretations ("prediction", "surprise", etc.) are just metaphors and should not be taken literally. In this case, articles such as this one would be pointless. I personally find this kind of restriction on theorizing unsatisfying if the goal of science is to provide us with an explanation of nature. Maybe we are taking risks when interpreting the Rescorla-Wagner model or the equations of quantum physics, but, at least from my point of view, these interpretations are the whole point of the scientific enterprise. Moreover, such a position makes it difficult to see how various theoretical concepts are related, notably in the case of this paper, how the Rescorla-Wagner model relates to an associative account of remembering and forgetting. Moreover, based on the arguments provided in section 6, it is very difficult, at least for me, to imagine that terms like "prediction" and "surprise" are used but metaphorically when applied to the Rescorla-Wagner model. There is also no indication that neuroscientists like Wolfram Schultz speaks metaphorically when claiming to have found evidence of error-correction learning in the brain. In RL, this is clearly not the case as the equations of RL are supposed to provide a rigorous mathematical framework for prediction and surprise (see, for instance, Equation 4 and following in Sutton & Barto, 1990).

A second argument would be to say that the Rescorla-Wagner model and similar models where associations are in the end prediction in disguise still describe one form of memory as it describes how links in an associative network are modified by experience. I did not rule out this possibility in this article. There are at least three ways by which an organism could achieve predictions: inference from memory which corresponds to episodic RL; simulation which corresponds to model-based RL; and direct prediction which corresponds to model-free RL. The Rescorla-Wagner model could be seen as providing an account of this third type of prediction mechanism. But it remains to be seen whether predictions are ever achieved in that way, and if so, in which circumstances. In the end, the issue boils down to what we mean by "memory". Any system modified by its past experience could be described as having a memory. In this case, DNA, as it reflects eons of natural selection, could be considered as a kind of memory. But such a definition of memory is so broad that it is unlikely that there is a single concept of memory. In the present paper, I used a much restrictive definition of memory, basically using it as a synonym of remembering, the most common and fundamental manifestation of memory in human beings and the one for which the concept of association is critical. The whole point of this paper was to examine whether the Rescorla-Wagner model has anything to say about remembering and the conclusion was that it does not. If the Rescorla-Wagner model corresponds to direct prediction, it might describe a type of memory if memory is undersood as a general term encompassing any system which can be affected by its past. But it has nothing to do with remembering. The later involves recovering a representation of a past event perceived as such. Direct prediction does not.

References

Atkins, C. K., (2000). Effect of species-specific cues and the CS-US interval in the topography of the sexually conditioned response. *Learning and Motivation, 31,* 211-235.

Atkins, C. K., Domjan, M., & Gutierrez, G. (1994). Topography of sexually conditioned behavior in male Japenese quail (*Coturnix japonica*) depends on the CS-US interval. *Journal of Experimental Psychology: Animal Behavior Processes, 20,* 199-209.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89-195). New York: Academic Press.

Anderson, M. C., & Neely, J. H. (1996). Interference and inhibition in memory retrieval. In E. L. Bjork & R. A. Bjork (Eds), *Handbook of perception and cognition* (2nd edition, pp. 237-313). San Diego, CA: Academic Press.

Baddeley, A. (2015). Retrieval. In A. Baddeley, M. W. Eysenck, & M. C. Anderson (Eds), *Memory (2nd edition)*. New York: Psychology press.

Balsam, P. D., Drew, M. R., & Gallistel, C. R. (2010). Time and associative learning. *Comparative Cognition & Behavior Reviews, 5,* 1-22.

Biele, G., Erev, I., & Ert, E. (2009). Learning, risk attitude and hot stoves in restless bandit problems. *Journal of Mathematical Psychology, 53,* 155-167.

Boenstein, A. M., Khaw, M. W., Shohamy, D., & Daw, N. D. (2015). Reminders of past choices bias decisions for reward in humans. *bioRxiv* 033910. doi: 10.1101/033910

Bouton, M. E. (1993). Context, time and memory retrieval in the interference paradigms of Pavlovian conditioning. *Psychological Review, 114,* 80-99.

Bush, R. R., & Mosteller, F. A. (1951). A mathematical model for simple learning. *Psychological Review, 58,* 313-323.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82,* 407-428.

Danks, D. (2003). Equilibria of the Rescorla-Wagner model. *Journal of Mathematical Psychology, 47,* 109-121.

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Associative learning of likes and dislikes: a review of 25 years of research on human evaluative conditioning. *Psychological Review, 127,* 853-869.

Denniston, J. C., Savastano, H. I., & Miller, R. R. (2001). The extended comparator hypothesis: Learning by contiguity, responding by relative strength. In S. Klein & R. Mowrer (Eds), *Handbook of contemporary learning theories* (pp. 65-117). Mahwah, NJ: Erlbaum.

Ebbinghaus, H. (1885). *Über das gedächtnis: untersuchungen zur experimentellen psychologie* [On memory: a contribution to experimental psychology]. Leipzig: Dunker.

Escobar, M., & Miller, R. R. (2004). A review of the empirical laws of basic learning in Pavlovian conditioning. *International Journal of Comparative Psychology, 17,* 279-303.

Gershman, S. J., & Daw, N. D. (2016). Reinforcement learning and episodic memory in humans and animals: An integrative framework. *Annual Review of Psychology, 68,* 101-128.

Gonzalez, C., & Dutt, V. (2011). Instance-based learning: Integrating sampling and repeated decision from experience. *Psychological Review, 118,* 523-551.

Higgins, T., & Rescorla, R. A. (2004). Extinction and retraining of simultaneous and successive flavor conditioning. *Learning & Behavior, 32,* 213-219.

Jozefowiez, J. (2014). The many faces of Pavlovian conditioning. *International Journal of Comparative Psychology, 27,* 526-536.

Kamin, L. J. (1968). “Attention-like” processes in classical conditioning. In M. R. Jones (Ed), *Miami symposium on the prediction of behavior: aversive stimulation* (pp. 9-33). Miami, FL: University of Miami Press.

Kuhn, T. S. (1962). *The structure of scientific revolutions.* Chicago, IL: University of Chicago Press.

Ludvig, E. A., Madan, C. R., & Spetch, M. L. (2015). Priming memories of past wins induces risk seeking. *Journal of Experimental Psychology: General, 144,* 24-29.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review, 82,* 276-298.

Madan, C. R., Ludvig, E. A., & Spetch, M. L. (2014). Remembering the best and worst of times: Memories for extreme outcomes bias risky decisions. *Psychonomic Bulletin & Review, 21,* 629-636.

Mansfield, J. G., & Cunnimgham, C. L. (1980). Conditioning and extinction of tolerance to the hypothermic effect of ethanol in rats. *Journal of Comparative and Physiological Psychology, 94,* 962-969.

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences, 32,* 183-198.

Miller, R. R., & Escobar, M. (2002). Associative interference between the cues and between outcomes presented together and presented apart: An integration. *Behavioural Processes, 57,* 163-185.

Miller, R. R., Barnet, R. C., & Grahame, N. J. (1995). Assessment of the Rescorla-Wagner model. *Psychological Bulletin, 117,* 363-386.

Murdock, B. B. (1973). *Human memory: theory and data.* New York: Wiley.

Murty, V.P, FeldmanHall, O., Hunter, L. E., Phelps, E. A., & Davachi, L. (2016). Episodic memories predict adaptive value-based decision making. *Journal of Experimental Psychology: General, 145,* 548-558.

Myers, K. P., & Sclafani, A. (2006). Development of learned flavor preferences. *Developmental Psychobiology, 48,* 380-388.

Pavlov, I. P. (1927). *Conditioned reflexes.* New York: Dover.

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review, 87,* 532-552.

Pollack, C. W., Jozefowiez, J., & Miller, R. R. (2017). Stepping back from "persistence and relapse" to see the forest: Associative interference. *Behavioural Processes, 141,* 128-136.

Rescorla, R. A. (1968). Probability of shock in the presence and absence of CS in fear conditioning. *Journal of Comparative and Physiological Psychology, 66,* 1-5.

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: Current Research and Theory* (pp. 64-99). New York: Appleton Century Croft.

Tvesky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 9,* 207-232.

Schultz, W. (2015). Neuronal reward and decision signals: From theories to data. *Physiological Review, 95,* 853-951.

Siegel, S. (1975). Evidence from rats that morphine tolerance is a learned response. *Journal of Comparative and Physiological Psychology, 89,* 498-506.

Sutton, R. S. (1988). Learning by the method of temporal difference. *Machine Learning, 3,* 9-44.

Sutton, R. S. (1991). Dyna, an integrated architecture for learning, planning, and reacting. *ACM SIGARD Bulletin, 2,* 160-163.

Sutton, R. S., & Barto, A. G. (1990). Time-derivative models of Pavlovian reinforcement. In M. Gabriel & G. Moore (Eds), *Learning and computational neurosciences: Foundations of adaptive networks* (pp. 497-537). Cambridge, MA: MIT Press.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction.* Cambridge, MA: MIT Press.

Stewart, N., Chater, M., & Brown, G. D. (2006). Decision by sampling. *Cognitive Psychology, 53,* 1-26.

Watkins, C. J. C. H., & Dayan, P. (1992). Q-learning. *Machine Learning, 8,* 279-292.

Wagner, A. R. (1981). SOP: A model of automatic memory processing in animal behavior. In N. E. Squire & R. R. Miller (Eds), *Information processing in animals: memory mechanisms* (pp. 5-47). Hillsdale, NJ: Erlbaum.

Wimmer, G. E., & Buchel, C. (2016). Reactivation of reward-related patterns from single past episodes support memory-based decision making. *Journal of Neuroscience, 36,* 2868-2880.

Acknowledgements

Jérémie Jozefowiez, Laboratory of Cognitive and Affective Sciences (SCALab), Université de Lille (Villeneuve d'Ascq, France) & Centro de Investigação em Psicologia (CIPSI), Universidade do Minho (Braga, Portugal). This manuscript is based on a talk given at the Comparative Cognition Conference (April 2017, Melbourne, FL), part of a session honoring the contribution of Dr. Ralph R. Miller to the study of learning. Correspondence concerning

this article should be addressed to Jérémie Jozefowiez, Laboratory of Cognitive and Affective Sciences (SCALab), Université de Lille, Campus de Lille 3, Domaine Universitaire du Pont de Bois, BP 60149, 58653 Villeneuve d'Ascq Cedex, France. E-mail: jeremie.jozefowiez@univ-lille3.fr.