Neurobiology of Learning and Memory 108 (2014) 14–21

Contents lists available at SciVerse ScienceDirect

Neurobiology of Learning and Memory

journal homepage: www.elsevier.com/locate/ynlme

Review

The conditions that promote fear learning: Prediction error and Pavlovian fear conditioning

Susan Shi Yuan Li, Gavan P. McNally *

School of Psychology, University of New South Wales, Australia

ARTICLE INFO

Article history:
Available online 16 May 2013

Keywords:
Conditioning
Fear
Prediction error
Associability
Blocking

ABSTRACT

A key insight of associative learning theory is that learning depends on the actions of prediction error: a discrepancy between the actual and expected outcomes of a conditioning trial. When positive, such error causes increments in associative strength and, when negative, such error causes decrements in associative strength. Prediction error can act directly on fear learning by determining the effectiveness of the aversive unconditioned stimulus or indirectly by determining the effectiveness, or associability, of the conditioned stimulus. Evidence from a variety of experimental preparations in human and non-human animals suggest that discrete neural circuits code for these actions of prediction error during fear learning. Here we review the circuits and brain regions contributing to the neural coding of prediction error during fear learning and highlight areas of research (safety learning, extinction, and reconsolidation) that may profit from this approach to understanding learning.

1. Introduction

In a standard Pavlovian fear conditioning experiment, a subject (typically a rat or mouse) receives presentations of a conditioned stimulus (CS) that co-terminate with presentations of an unconditioned stimulus (US). One consequence of these pairings is that the rat will display a variety of conditioned responses upon later presentations of the CS. These may include the species-typical defense of freezing, changes in heart rate, changes in blood pressure, ultrasonic vocalizations, among others. There are numerous questions that could be asked about this, but perhaps the most fundamental is, what are the circumstances that cause fear learning?

The answer to this question, provided by associative learning theorists in the 1970s and 1980s, was heavily influenced by analyses of empirical phenomena such as blocking (Kamin, 1968), unblocking (Kamin, 1968), relative validity (Wagner, Logan, Haberlandt, & Price, 1968), and overexpectation (Rescorla, 1970). For example, in the case of blocking, prior fear conditioning of CSA is able to block learning to CSB when CSA and CSB are presented in compound and followed by a US. In the case of overexpectation, fear to CSA is reduced when CSA, which is already established as predictor of shock, receives further fear conditioning in compound with another fear CS. Analyses of these effects show that temporal contiguity between a CS and US is not an adequate condition for fear learning. Instead, they encouraged the view that Pavlovian association formation depends on prediction error – a discrepancy between the predicted outcome of the conditioning trial and the actual outcome of that trial. This error determines whether the shock US is effective in supporting learning or not, so that unexpected USs are more effective in supporting learning than expected USs (Rescorla & Wagner, 1972). This error also determines whether the CS is effective in terms of entering into associations with the US, so that the associability of the CS is a function of how well it predicts the US, including relative to any other CSs present (Mackintosh, 1975; Pearce & Hall, 1980). The answer to this question could then be determined by describing the conditions under which such variations in CS and US processing can occur and describing the rules that govern these variations.

At the same time as associative learning theorists were refining and recasting the conditions for association formation, studies of the neurobiology of fear learning remained dominated by a different answer to this question. The Hebbian tradition identified and retained stimulus temporal contiguity as the critical determinant of learning. Within this tradition, learning depends on potentiation of synaptic communication produced by the co-occurrence of activity in neuron pairs, specifically the repeated and persistent activation of a post-synaptic neuron by a pre-synaptic neuron. The answer to the question could then be determined by identifying neurons where stimulus convergence may occur during a learning episode, such as identifying synapses for potential convergence of CS and US inputs, and then defining the changes that occur at such synapses at the cellular and molecular levels.

Both of these approaches led to remarkable successes. Much of the work in associative learning from the 1970s to the 1990s

* Corresponding author. Address: School of Psychology, The University of New South Wales, Sydney, NSW 2052, Australia. Fax: +61 2 93853641.
E-mail address: g.mcnally@unsw.edu.au (G.P. McNally).

1074-7427/$ - see front matter
http://dx.doi.org/10.1016/j.nlm.2013.05.002

successfully identified the actions of prediction error on association formation and provided formal learning rules to describe these actions. Likewise, fear learning yielded, as did other learning phenomena before it, to the power of molecular neuroscience. A wealth of information on the cellular and molecular mechanisms of synaptic plasticity during fear learning was generated. However, despite the fact that these two approaches were attempting to answer the same question, albeit at different levels of analysis, they proceeded relatively independently of each other. Neuroscientists came to a consensus view on the mechanisms for contiguity detection in Pavlovian fear conditioning: fear learning was due to stimulus contiguity causing NMDA receptor mediated synaptic plasticity at individual lateral amygdala neurons, and the subsequent intracellular cascades linked to this plasticity (Maren & Quirk, 2004; Schafe, Nader, Blair, & Ledoux, 2001). Associative learning theorists came to the view that such contiguity was the final step in a sophisticated stimulus selection process that allowed animals to use their past experiences with stimulus events to learn about and respond to those events in the future.

In this paper we review work that suggests reconciliation between these two approaches to understanding the conditions for fear learning. Our aim is to highlight how an understanding of the role of prediction error in association formation has yielded a richer understanding of the neural substrates of fear learning and to also highlight areas within the neurobiology of fear learning that may profit from this approach. The starting point for this work has been to accept the overwhelming evidence that individual neurons, in the lateral amygdala and elsewhere, encode the relationship between the CS and US during fear learning, and to also accept that there must be variations in the effectiveness of the CS and US in recruiting these neurons during fear conditioning. The questions then become: what is the neurobiological evidence for such variations in CS and US processing during fear learning? And, what are the mechanisms that cause these variations?

2. Positive prediction error and fear learning

The Rescorla–Wagner learning rule (Rescorla & Wagner, 1972) asserts that fear learning is due to variations in the effectiveness of the shock US in supporting learning across the course of conditioning. Formally, this can be stated as:

$$\Delta V_X = \alpha\beta(\lambda - \Sigma V)$$

where the change in associative strength (V) of CSX is determined by the fixed saliences of the CS (α) and US (β), and the error term ($\lambda - \Sigma V$). This error term reflects the difference between the US that is delivered on a trial (λ), and the US that is expected or predicted to occur, based on the summed associative strengths (ΣV) of all CSs present on the trial. Prediction error, therefore, encodes the difference between the actual versus expected US. This error signal controls learning because it dictates variations in the effectiveness of the US in supporting learning. When $\lambda > \Sigma V$ (i.e. there is positive prediction error), the US is surprising and increments in fear learning occur. When $\lambda = \Sigma V$ (i.e. there is no prediction error), the US is not surprising and no increments in fear occur learning.

Fear learning depends on synaptic plasticity at neurons in the lateral amygdala (LA) and other brain regions (e.g., auditory cortex) that detect the temporal conjunction between the CS and US (Johansen, Hamanaka, et al., 2010; Letzkus et al., 2011; Mahan & Ressler, 2011; Maren & Quirk, 2004; McKernan & Shinnick-Gallagher, 1997; Rogan, Staubli, & LeDoux, 1998). A strong implication of the Rescorla–Wagner model is that the strength of US inputs to these LA and other neurons is not invariant across the course of conditioning. Rather, the strength of these inputs is a function of ($\lambda > \Sigma V$). Three lines of evidence suggest that the nervous system encodes variations in US effectiveness during Pavlovian fear conditioning in this manner. First, single unit recordings during Pavlovian fear conditioning show that US-evoked depolarization of LA neurons decreases across the course of CS–US pairings (Johansen, Tarpley, Ledoux, & Blair, 2010) as the US becomes expected and this decrease correlates with increases in expression of fear conditioned responses. US-evoked depolarization can be restored if US delivery is unsignalled. Second, imaging studies of cellular activity markers, such as the Fos protein, show that unexpected US presentations elicit significantly greater activity in LA neurons than expected US presentations (Furlong, Cole, Hamlin, & McNally, 2010). Finally, human neuroimaging studies show that US-evoked fMRI signals in amygdala are greater in response to unexpected than expected USs (Dunsmoor, Bandettini, & Knight, 2008; Dunsmoor & Schmajuk, 2009). These findings also extend, in human participants at least, to blocking paradigms. In a within subjects blocking design, amygdala fMRI BOLD responses to the blocked CS during Stage II fear learning are significantly smaller than those responses to the control CS, and these differences are preserved on a later test, mirroring the self-reported expectancy of the US and magnitude of the skin conductance CR (Eippert, Gamer, & Buchel, 2012) .

These findings are consistent with a prediction error interpretation derived from the Rescorla–Wagner model. At the start of conditioning, when prediction error is high because the US is unexpected, $\lambda > \Sigma V$, US-evoked activity in LA is high. Across the course of conditioning, as the CS comes to predict the US and hence the shock becomes expected, prediction error is low, $\lambda = \Sigma V$, and US evoked activity in LA is likewise low. Surprising presentations of the US, for example by omitting the CS, restore US evoked activity in LA neurons. It is highly probable, but yet to be shown that these variations in US-elicited activity determine the amount of fear learning that occurs on a trial.

These variations in US-elicited activity also occur in other brain regions and there is stronger evidence for a causal role of these other brain regions in contributing to prediction error during fear learning. The periaqueductal gray (PAG) is an important point of convergence between the output of the fear system (it is heavily innervated by central amygdala neurons) and the inputs to this system from the sensory systems (Carrive, 1993). Focal electrical stimulation of the PAG serves as an effective US during fear learning (Di Scala, Mana, Jacobs, & Phillips, 1987) and individual PAG neurons, like LA neurons, are responsive to the shock US during fear conditioning (Johansen, Tarpley, et al., 2010). This strongly implicates the PAG in US processing during fear learning. Indeed, individual PAG neurons show the same variations in US-evoked depolarization as LA neurons across the course of fear conditioning. A surprising US evokes depolarization of PAG neurons whereas an expected shock US does not. Moreover, US-elicited activity in PAG neurons is necessary for both US-elicited activity in LA neurons and fear learning (Johansen, Tarpley, et al., 2010). This last finding strongly implicates the PAG relaying variations in the effectiveness of US inputs to LA neurons.

Within the PAG, opioid receptors have been directly linked to learning in response to prediction errors. Systemic administrations of an opioid receptor antagonist prior to CS–US pairings augment the acquisition of Pavlovian fear conditioning in rats (Bolles & Fanselow, 1982; Fanselow & Bolles, 1979; Helmstetter & Fanselow, 1987; McNally, Pigg, & Weidemann, 2004a) and humans (Eippert, Bingel, Schoell, Yacubian, & Buchel, 2008) and also prevent reductions in amygdala BOLD responses to expected shock USs in humans (Eippert et al., 2008). PAG opioid receptors, especially those in the vlPAG, are causally implicated in this process. For example, using a within-subjects design, we (McNally & Cole, 2006) trained rats to fear CSA in Stage I. In Stage II, CSA and CSB were conditioned in compound (AB+) as were CSC and CSD (CD+). On test, there was less fear to CSB than CSD – the prior training of CSA blocked fear