Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Neurobiology

Learning rules for aversive associative memory formation

Takaaki Ozawa[1] and Joshua P Johansen[2,3]

For survival, organisms need the ability to flexibly modify their behavior. To achieve this, the brain is equipped with instructive brain circuits which trigger changes in neural connectivity and adaptive changes in behavior in response to environmental/internal challenges. Recent studies using a form of aversive associative learning termed fear conditioning have shed light on the neural mechanisms of instructive signaling. These studies demonstrate that fear learning is engaged through multiple, parallel aversive signaling pathways to the amygdala. Consistent with theoretical accounts of learning, activity in these circuits and behavioral learning is tightly regulated by the predictability of the aversive experience. However, in more complex learning conditions, these emotion circuits use a form of inference to approximate the appropriate reaction to danger. This suggests a revised view of how emotional learning systems represent aversive associations and how changes in these representations are instructed during learning.

Addresses
[1] Department of Basic Neurosciences, University of Geneva, Geneva, Switzerland
[2] RIKEN Brain Science Institute, Laboratory for Neural Circuitry of Memory, 2-1 Hirosawa, Wako-shi, Saitama, Japan
[3] Department of Life Sciences, Graduate School of Arts and Sciences, University of Tokyo, Tokyo, Japan

Corresponding author: Johansen, Joshua P (jjohans@brain.riken.jp)

Current Opinion in Neurobiology 2018, **49**:148–157

This review comes from a themed issue on **Neurobiology of behavior**

Edited by **Kay Tye** and **Naoshige Uchida**

https://doi.org/10.1016/j.conb.2018.02.010

To survive and flourish in an ever changing world organisms need to alter their behavior through experience. Distinct learning systems in the brain fine tune different aspects of behavior to optimize, for example, vocal/motor skills or perform survival functions such as obtaining food and avoiding danger [1–7]. To do this, dedicated neural circuits transduce salient experiences (e.g. motor errors, rewards, aversive stimuli) into neural instructive signals which alter connectivity in brain regions responsible for controlling these different behaviors. While there are distinct instructive signaling pathways for different types of learning, common themes appear to be shared across these systems including 'prediction error' coding whereby instructive signals become inhibited as outcomes become anticipated during learning (Figure 1). Thus by understanding the mechanisms of instructive signaling in one brain system, more general rules for how the brain learns can be gleaned.

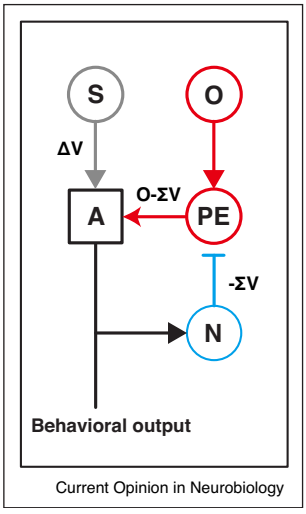
Great progress has been made in understanding how aversive emotional memories form and the instructive signaling circuits which trigger this type of learning. As an example of emotional learning, if one encounters a poisonous snake in the woods, strong emotional memories form of the encounter. Thus the next time one is walking in the woods and hears a rustling in the underbrush, a concerted emotional response is engaged including stress hormone release, changes in heart rate and freezing or avoidance responses. Together, this combined set of visceral reactions to threatening stimuli has been termed the fear response (but see [8] for alternate terminology). Fear learning is modeled in the laboratory using auditory fear conditioning (Figure 2) in which an auditory stimulus (also called a conditioned stimulus, CS) is paired with an aversive outcome (termed an unconditioned stimulus or US, such as mild electrical shock) [9,10]. Presentation of the auditory cue following learning elicits a set of behavioral and visceral fear responses. Thus, by defining the circuits that are activated by shock to alter brain connectivity and produce learning we can understand aversive instructive signaling in this learning system.

In this review we will discuss these recent discoveries on aversive instructive signaling circuits and neural coding in the fear learning system. In addition, we will highlight newer ideas about how information is represented in fear learning circuits and how more complex instructive signals are required for regulating these representations during learning. Because of the conserved nature of instructive signaling mechanisms across learning systems, we will draw parallels with other learning systems where appropriate.

Depolarization and neuromodulatory signaling trigger amygdala neural plasticity during fear learning

Critical to understanding any instructive signaling system is identifying a site of neural plasticity where the association between predictive cues and outcomes are encoded. For fear conditioning, the amygdala is a known site of

Figure 1



Conceptual schematic diagram of prediction error coding for sensory-outcome associative learning. Across many learning systems, instructive signals are inhibited when salient outcomes become predicted by other sensory or motor cues. In some systems learning (ΔV) has been modeled using a prediction error term ($\Delta V = O - \Sigma V$), with O = reinforcing outcome and ΣV = predictive strength of all given sensory predictive stimuli. Thus learning only occurs when the outcome is not well predicted. In this example schematic model of sensory-outcome associative learning, an originally neutral stimulus (S) becomes associated (A) during learning (ΔV) with a salient outcome (O) through activation of an instructive signal signaling pathway (prediction error, PE) by O. Once S fully predicts O, it can activate A by itself and drive behavioral responses. In addition, S can now activate a negative feedback system (N) which inhibits ($-\Sigma V$) the ability of O to activate PE thereby setting prediction error coding ($O - \Sigma V$) in PE.

plasticity (Figure 2) [4,9,11–15]. Auditory input synapses from the thalamus and cortex to the lateral nucleus of the amygdala (LA) are strengthened during fear learning, likely onto specific populations of LA or basal (B) nucleus of the amygdala neurons [16–18]. This occurs in parallel with strengthening of LA/B inputs to the central nucleus of the amygdala (CeA) [19,20]. Following fear learning, presentation of the auditory cue now produces stronger activity in the LA and CeA [21,22,23••] which, in coordination with the medial prefrontal cortex (mPFC) [24–30], generates fear responses through projections to the peri-aqueductal gray (PAG, for freezing) [22,31] and other brainstem and hypothalamic sites mediating other fear responses [31–33].

So what are the aversive shock evoked signals within the amygdala which trigger this plasticity? LA neurons are activated by aversive shocks and recent work has found that shock evoked depolarization of LA pyramidal neurons during the shock period of fear conditioning is necessary for plasticity of auditory processing in the LA and fear learning [34–36]. Further support comes from an *in vivo* intracellular recording study in LA neurons which reported that depolarization during the shock is required for learning induced potentiation of sensory responses [37]. Together with data showing coincidence of tone and shock information in some LA neurons, this suggests that Hebbian mechanisms are important in producing fear learning. However, under normal learning conditions depolarization is not sufficient to produce learning unless β-adrenergic receptors are co-activated [34,38]. This idea that Hebbian plasticity alone is not sufficient to produce amygdala plasticity during fear learning was also suggested by a recent paper using *in vivo* calcium imaging of large populations of LA/B neurons [39••]. Through population level analyses of stimulus processing before, during and after fear conditioning the authors showed that auditory CS-evoked responses became more similar to the shock US-evoked response across the population of recorded cells. Notably, learning induced changes in auditory processing were not well correlated with shock responsiveness, prompting the authors to suggest that Hebbian mechanisms alone were not sufficient to explain amygdala associative plasticity.

Several other issues with the pure Hebbian interpretation have also been proposed (see [40] for detailed discussion of these issues). One is that extracellularly recorded LA neurons exhibit phasic spiking responses to auditory stimulus onset and this response habituates, suggesting that the CS inputs to these cells may not be active at the time of US occurrence. Furthermore, standard fear conditioning can still be induced when shocks occur milliseconds-seconds after tone offset at a time when auditory inputs may no longer be active. Thus the timing of the shock after tone offset is outside the window for coincident Hebbian LTP as defined in *ex vivo* slice physiology experiments. However, plasticity mechanisms *in vivo* are likely more temporally flexible as there are many network level processes available in intact systems which are not present in excised brain slices. For example, soma targeting, PV interneurons in LA are activated by tones and shocks during fear learning [35] and this could clamp somatically generated action potential firing while plateau potentials in the dendrites are still active. Thus associative Hebbian-like plasticity could occur independently of action potential firing, similar to a mechanism reported for an *in vivo* somatosensory cortex synaptic plasticity [41]. Providing support for this idea, many cells in auditory thalamic regions which provide input to LA exhibit sustained auditory evoked responses and others display stimulus offset responses [42,43]. Thus auditory inputs to LA cells may be active throughout the auditory cue period and even exhibit offset responses which could lengthen the time these inputs are active to coincide with shock occurrence, though this may not always be evident in the extracellularly recorded action potential firing rate of LA neurons. Furthermore, although many LA neurons show phasic firing rate increases to auditory stimulus onset [21], some exhibit sustained responses [44–46], albeit at lower levels