Neurobiology of Learning and Memory 108 (2014) 5–13

Contents lists available at ScienceDirect

Neurobiology of Learning and Memory

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

Isomorphisms between psychological processes and neural mechanisms: From stimulus elements to genetic markers of activity

Michael S. Fanselow [a,b,*], Moriel Zelikowsky [c], Jennifer Perusini [a], Vanessa Rodriguez Barrera [a], Sarah Hersman [a]

[a] Department of Psychology, University of California, 405 Hilgard Ave., Los Angeles, CA 90095, United States
[b] Department of Psychiatry & Biobehavioral Sciences, University of California, 405 Hilgard Ave., Los Angeles, CA 90095, United States
[c] Department of Biology and Biological Engineering, California Institute of Technology, 1200 E California Blvd., Pasadena, CA 91125, United States

ARTICLE INFO

Article history:
Available online 8 November 2013

Keywords:
Learning theory
Fear conditioning
Amygdala
ARC
catFISH
Error correction
Hippocampus

ABSTRACT

Traditional learning theory has developed models that can accurately predict and describe the course of learned behavior. These "psychological process" models rely on hypothetical constructs that are usually thought to be not directly measurable or manipulable. Recently, and mostly in parallel, the neural mechanisms underlying learning have been fairly well elucidated. The argument in this essay is that we can successfully uncover isomorphisms between process and mechanism and that this effort will help advance our theories about both processes and mechanisms. We start with a brief review of error-correction circuits as a successful example. Then we turn to the concept of stimulus elements, where the conditional stimulus is hypothesized to be constructed of a multitude of elements only some of which are sampled during any given experience. We discuss such elements with respect to how they explain acquisition of associative strength as an incremental process. Then we propose that for fear conditioning, stimulus elements and basolateral amygdala projection neurons are isomorphic and that the activational state of these "elements" can be monitored by the expression of the mRNA for activity-regulated cytoskeletal protein (ARC). Finally we apply these ideas to analyze recent data examining ARC expression during contextual fear conditioning and find that there are indeed many similarities between stimulus elements and amygdala neurons. The data also suggest some revisions in the conceptualization of how the population of stimulus elements is sampled from.

1. Introduction

Well into the last century, learning theorists have been developing models of the psychological processes underlying associative learning. These provide rules of how specific experiences change "associative strength" over the course of learning, and these rules provide powerful descriptions of both simple and complex forms of conditioning (Bush & Mosteller, 1951; Hull & et al., 1940; Mackintosh, 1975; Pearce & Hall, 1980; Rescorla & Wagner, 1972). These theories typically rely on several hypothetical constructs that, while not directly measurable, enhance the explanatory power of the theory (e.g., associative strength). More recently there have been explosive advances in our knowledge about the neural mechanisms required for learning (Nicoll & Malenka, 1999). For example, we know that glutamate's action on NMDA receptors at a set of synapses supports long-term potentiation of synaptic efficacy by increasing excitatory synaptic transmission at those synapses. It seems that the next step in developing our understanding of learning is to ask what, if any, isomorphisms exist between process and mechanism. This cross-level translation would likely be synergistic and drive each class of models (process and mechanism) beyond current understanding. Indeed, it would not be surprising if once such an isomorphism was identified it immediately suggested a modification to existing theories. Below we briefly review fear-conditioning data where there has been success in identifying such an isomorphism (error correction), and then introduce a hypothesis for isomorphisms relating to learning theories that assume that conditional stimuli are best decomposed into a set of primitive elements.

2. Error-correction: An example isomorphism between psychological process and neural mechanism

One example case of this synergy is the recognition that the teaching signal for conditioning is not the reinforcer but the degree to which the reinforcer received differs from what is typical in the current situation. Kamin (1968) first suggested that it was the surprisingness of reinforcement, not reinforcer magnitude, that

* Corresponding author at: Department of Psychology, University of California, 405 Hilgard Ave., Los Angeles, CA 90095, United States. Fax: +1 310 206 5895.
E-mail addresses: mfanselow@gmail.com, fanselow@ucla.edu (M.S. Fanselow).

1074-7427/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.nlm.2013.10.021

supported association formation. Then Rescorla and Wagner (1972) formalized this notion of surprise by saying "changes in associative strength of a stimulus as a result of a trial can be well-predicted from the composite strength resulting from all stimuli on that trial (p73)." Reinforcement then becomes the difference between the reinforcer delivered and the composite value for associative strength. Bolles and Fanselow (1980) elaborated these ideas into a prediction error framework saying that "any discrepancy between expected and perceived US (unconditional stimulus) features is fed back to alter future expectations (p293)" so that "any error in the expectation is fed back so as to reduce future errors (p293)." Importantly, Bolles and Fanselow (1980) described how a circuit capable of this function would look—it would have an inhibitory feedback signal that was a conditional response (CR), thereby proportional to associative strength that would subtract from the US experienced. Because that model was developed to specifically explain fear conditioning, the expectancy generated by the conditional stimulus (CS) was the expectation of pain and the perceived US was the pain caused by the US. This specificity at once suggested a mechanism—endogenous opioids and their descending analgesic influence was the embodiment of the negative feedback arm of the circuit. Rapidly, empirical data were generated that extensively supported that idea—treating animals with opioid antagonists turned them into animals that learned proportional to the actual US rather than its surprisingness (Fanselow, 1986a; Fanselow & Bolles, 1979; Helmstetter & Fanselow, 1987; McNally, Pigg, & Weidemann, 2004; Young & Fanselow, 1992; Zelikowsky & Fanselow, 2010). This pointed to an anatomy for this circuit (Fanselow, 1986a, 1998) that has now received extensive delineation (Johansen, Tarpley, LeDoux, & Blair, 2010; McNally, Johansen, & Blair, 2011). Additionally, eyeblink learning, which uses a very different circuit than fear conditioning, still contains a source for negative feedback that functions in a similar manner, that is, limiting the US's ability to drive changes in associative strength (Fanselow, 1998; Kim, Krupa, & Thompson, 1998). While still unknown, the circuitry mediating appetitive Pavlovian conditioning is likely to have a similar negative feedback arm (Schultz & Dickinson, 2000; Waelti, Dickinson, & Schultz, 2001).

Knowledge about the neural mechanisms of error correction can, in turn, provide insights into psychological processes. One example is the finding that learning is slowed by both nonreinforced and weakly reinforced pre-exposure to a CS (Hall & Pearce, 1979; Lubow, 1973). Initially, it was thought that both effects were caused by similar reductions in attention to the CS (Pearce & Hall, 1980). However, using pharmacological manipulation of the error-correction circuit it was found that US-related error correction mechanisms account for the slower conditioning after a shift from a weak to a strong shock but cannot account for the slowed learning after nonreinforced CS preexposure (Young & Fanselow, 1992).

Interestingly, while the mechanisms of error-correction have yielded to such an analysis of US processing, we still do not know if there are process-mechanism isomorphisms for the concepts of associative strength and CS representation. These hypothetical constructs are as key to an understanding of learning processes, as is error correction. The purpose of this paper is to speculate about a potential isomorphism for the representation of the CS and how that representation comes to track associative strength. Fear conditioning again holds promise because the underlying circuitry has been fairly well characterized (Fendt & Fanselow, 1999; Haubensak et al., 2010; Paré, Quirk, & LeDoux, 2004). Importantly, sensory information about the CS is relayed to the basolateral amygdala (BLA) and substantial evidence suggests that it is within this structure that the CS–US associations underlying fear conditioning are formed (Fanselow & LeDoux, 1999). But first, we must briefly look at some assumptions about the representation of the CS that have been made by learning theorists.

3. Stimulus-sampling theory

When we talk about CS we typically refer to the objective external stimulus such as a tone or light. However, mechanistically, it is either the neural representation of that stimulus, or what that neural representation engenders, that must change during associative learning. Psychologists have long recognized that the objective stimulus and its neural representation are not identical (Fechner, 1860; Stevens, 1957; Weber, 1834). One important example of this is Edwin R. Guthrie's (1935) that the stimulus is really a dynamic set of a very large number of elements only a proportion of which are active at any given moment. William Estes (1950) quantitatively formalized Guthrie's view in his stimulus-sampling theory. A critical aspect of the model is that a stimulus is made of a large set of primitive elements and only a subset (sample) of elements can be active at any point in time. Therefore, every time a stimulus is experienced it is represented by a somewhat different set of elements. With each experience a new sample is taken randomly, with replacement, from the total population of elements. Two experiences are similar to the extent that they contain common elements. This elemental view of associative learning has been incorporated into many associative models because it has tremendous explanatory power especially for phenomena such as acquisition, discrimination and generalization, core aspects of any learning theory (Rescorla, 1976; Rescorla & Wagner, 1972; Rudy & Wagner, 1975; Wagner, 2003, 2008). As illustration, generalization of responding to stimuli in a manner proportional to their similarity is easily explained by the degree to which the finite populations of elements representing the individual stimuli share common elements.

A second key aspect to the Guthrie-Estes elemental view is that associations are formed in an all-or-none manner to the individual elements. If, on a given trial, an element is present, it may enter into association but if it is not present it cannot enter into association. Thus an incremental learning curve occurs because on trial one, associations are formed only to the limited set elements present. On trial two, a randomly determined set of elements is sampled, of which only a small proportion of which were present on the first trial. Those resampled elements drive a small CR and the new elements sampled can now become associated with the trial's outcome (e.g., shock in fear conditioning). With each trial a greater proportion of the entire population of elements will have entered into association and therefore a stronger and more consistent CR will emerge.

Is there something in the nervous system that corresponds to a stimulus element? Can we in some way track whether an element has been sampled in the sense offered by Estes? If possible, we would have another critical lynchpin for understanding the neural basis of learning. Additionally, being able to track a stimulus element during learning would provide data that could test and sharpen our models of the psychological processes describing association formation. Indeed, measurement of such elements could potentially be isomorphic with associative strength. Below we entertain one potential neural candidate for a stimulus element in the context of the learning curve. Our first consideration is behavioral; we need to select an appropriate learning task for analysis, which would be one where there are clear increments in learning over trials. Next, we speculate on the neural basis of a stimulus element and then examine how those elements behave.

4. The incremental nature of fear conditioning

Above we argued, using the vantage of stimulus-sampling theory, that an animal shows increasing amounts of fear to a CS as conditioning proceeds because a greater proportion of the total