Published in final edited form as:
Neurobiol Learn Mem. 2014 February ; 108: 5–13. doi:10.1016/j.nlm.2013.10.021.

Isomorphisms between Psychological Processes and Neural Mechanisms: From Stimulus elements to genetic markers of activity

Michael S. Fanselow[1], **Moriel Zelikowsky**[2], **Jennifer Perusini**[1], **Vanessa Rodriguez Barrera**[1], and **Sarah Hersman**[1]

[1]Department of Psychology, Department of Psychiatry & Biobehavioral Sciences; University of California, Los Angeles; 405 Hilgard Ave, Los Angeles, CA 90095

[2]Department of Biology and Biological Engineering; California Institute of Technology; 1200 E California Blvd, Pasadena, CA 91125

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.

Keywords

Learning Theory; Fear Conditioning; Amygdala; ARC; catFISH; error correction; hippocampus

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, 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.

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 supported association formation. Then Rescorla & 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 & 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 & Bolles, 1979; Fanselow, 1986a; Helmstetter & Fanselow, 1987; McNally et. Al., 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 et al., 2010; McNally et al., 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 et al., 1998). While still unknown, the circuitry mediating appetitive Pavlovian conditioning is likely to have a similar negative feedback arm (Schultz & Dickinson, 2000; Waelti et al., 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 (Lubow, 1973; Hall & Pearce, 1979). 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; Pare et al., 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.

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; Weber, 1834; Stevens, 1957). One important example of this is Edwin R. Guthrie's (1935) suggestion 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 (Wagner, 2003; 2008; Rudy & Wagner, 1975; Rescorla, 1976; Rescorla & Wagner, 1972). 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.

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 population of elements for the CS enters into association. Whether incremental learning curves represent continuous variation in an individual's associative strength or are an artifact of averaging multiple subjects with different learning rates is a debate that has raged as long as there has been learning theory (Krechevsky, 1932; 1938; Lashley, 1929; Spence, 1940). Indeed, some have challenged the entire approach of breaking down learning into trials altogether and suggested that the formation of a representation of the environment is abrupt and what changes with experience is confidence in the certainty of the representation (Gallistel, 1990; 2007). Others have suggested that there are several factors that shape learning curves other than CS-US pairings (i.e., trials; Gottlieb, 2008). It is likely that there are different answers to this question that depend on the specific task and even method of training (Riley, 1968). If our goal is to track stimulus elements during acquisition of fear, first we must show that learning in this task is incremental.

Most presentations of the acquisition of fear conditioning show curves composed of group averages, which do not allow one to determine whether or not an individual 's learning is incremental or stepwise. Gottlieb & Rescorla (2010), using a within-subjects design found little evidence that the number of Light-Shock pairings was an important factor when they examined acquisition of fear. However, clear differences emerged when testing of these rats was carried out during extinction of the CS-US association. Gottlieb & Rescorla (2010) therefore suggested that the failure to see an effect of the number of CS-US pairings during acquisition was a sensitivity issue. There are two factors that may have reduced Gottlieb & Rescorla's (2010) sensitivity to observing trial-by-trial changes during acquisition. First, all training and testing occurred in the same context. Such tests naturally confound conditioning to the explicit CS (light) and the conditioning context. While they reported responding to the CS in terms of a ratio of CS to preCS responding (suppression ratio) such measures do not eliminate the confound that baseline conditional responding creates for accurate assessment of fear (Jacobs et al., 2010). Second, and possibly more important, the measure of fear was suppression of a food reinforced lever press. The motivation of these food-deprived rats may have masked expression of small changes in fear that occur from trial-to-trial.

To avoid these sensitivity issues, we wished to use the simplest example of fear acquisition for these experiments. In some ways, contextual fear conditioning is the simplest case of acquisition as there are only two stimuli, the CS (context) and the US (shock). Auditory conditioning adds a CS and thereby complicates the picture because we know that CSs interact with each other (e.g., they contribute to Rescorla & Wagner's composite stimulus, and they enter into a Rescorla-Wagner-like competition, where overshadowing between stimuli influences the degree to which a stimulus will reach its asymptotic conditioning

value). Additionally, rather than leverpress suppression in food-deprived rats, we examined freezing in nondeprived animals. The test of the incremental nature of learning would be whether individual animals adopt intermediate levels of responding during the course of learning. Below we show that learning appears incremental with two very different procedures, one in rats and one in mice.

Figure 1 presents data from an experiment conducted for another purpose. Rats received a single session of 15 (1 mA, 1.0 sec) shocks occurring randomly over 90 min. The data present freezing for 5 rats during one minute periods prior to each shock. Freezing is a reflection of the context-shock association in this sort of situation. As is typical, the average learning curve negatively accelerates reaching asymptotic freezing levels near 100% in about 4 trials. All rats contributing to the curve adopted intermediate levels of performance prior to asymptote, except one who went from 0 to 100 after one shock. Importantly, on trial two, these 4 rats (C–F) had a level of freezing that was greater than the first trial but was less than the asymptotic level indicating that freezing behavior was acquired in an incremental manner.

Figure 2A–E asks a similar question of mice using a one trial a day procedure carried out over 4 days. Mice were placed in a context once a day for 3 min. Every session ended in a single shock (0.65mA, 2.0 sec). Each line on an individual graph represents a different trial. The composite graph (Fig 2E) suggests that each successive trial caused an increase in freezing. Individual curves clearly indicate that all mice adopted intermediate levels of freezing prior to the 4^{th} trial. Indeed, on trial 2, every mouse froze more than it did on trial one but less than trial 3. Again, the acquisition of contextual fear was graded. Thus, incremental learning is a robust behavioral phenomenon that can be observed both within and across species. Such profound behavioral effects lead one to question the neural analogs for the hypothetical stimulus elements that account for such incremental learning.

The Neurobiology of Stimulus Elements

First, we must preface our attempt to define a stimulus element in neurobiological terms by acknowledging Estes' insistence that the elements were entirely hypothetical and never directly measurable (Estes, 1950). For Estes, this was a mathematically tractable approach, as the terms of theory could be expressed as a proportion of the entire set of elements even if the set size itself was unknown. Here we will violate this hypothetical assumption with reckless abandon and develop a hypothesis based on the assumption that we can not only count the elements but also know whether or not they were sampled.

The first question is where should we look for stimulus elements? If we are interested in the ability for the CS to acquire a fear CR the obvious place to look is the BLA.

The next step is to identify the population of elements for a given CS. Stimulus sampling theory assumes that the total population of elements for a given stimulus is fixed and that these elements are in one of two states, sampled or not. Estes (1950) also assumed that, "in the familiar buzz-shock conditioning experiment, for example, S c (the set of elements for the CS) would represent the population of stimulus elements emanating from the sound source (p284)." Some BLA neurons meet these 3 requirements. There are BLA neurons that receive synaptic input from the CS (i.e., they emanate from the sound source; e.g., Romanski et al., 1993), and they are potentially countable. Action potentials, which drive the neurons output, are all-or-none (i.e., they are in one of two states). Given the absence of adult neurogenesis in the amygdala, the number of BLA neurons is more-or-less fixed.

We will further limit the population of elements to BLA pyramidal cells. First, because they are the excitatory projection neurons that communicate with the downstream structures that

drive behavior. Second, the high tonic firing rate of inhibitory neurons does not correspond with something that is probabilistically activated and only at times when the CS is presented. This should not be taken to mean that BLA interneurons do not play an important role; they clearly regulate BLA physiology and fear-related behavior (Ehrlich et al., 2009). They certainly serve to coordinate the firing of principle neurons (Ryan et al., 2012). One role interneurons may play within the current schema is to control the probability of an element (i.e., principle neuron) being sampled (i.e., activated).

In Estes' theory, the population of elements for the CS was not constrained by the US. This worked for stimulus sampling theory because it assumed that any CS element could become associated with any response element. This "equipotentiality" assumption has been contradicted empirically in that not every US can become associated with every CS, some CS-US combinations work well (Taste-Illness, Tone-Shock) and some work poorly (Tone-Illness, Taste-Shock, Light- Shock; Chung et al., 2011; Garcia & Koelling, 1966; Newton et al., 2004). A good argument can be made that when translating stimulus elements into neurons, a US-related constraint must be placed on what neurons contribute to the population because an element that is never activated by the US cannot be part of the population representing the CS-US association. In this way, Estes' original idea of stimulus sampling is constrained to elements that have the potential to neurally converge with the US. The BLA contributes to several types of Pavlovian conditioning, including appetitive conditioning and taste aversion learning (Chung et al., 2011). Separate populations of BLA neurons contribute to these different types of learning and they appear randomly distributed rather than topographically organized throughout the region (Chung et al., 2011). If a shock US cannot activate a food-related neuron, that neuron has no chance in entering into a CS-Shock association and thus cannot be thought of as part of the relevant population of elements. However, at this point in time, it is technically difficult to determine whether an individual neuron has the potential to be activated by both the CS and the US. Therefore, we will define the total population of elements as the population of BLA neurons and recognize that this is likely an overestimate that occurs because we are including some neurons in our analysis that the shock can not activate.

Stimulus sampling in the basolateral amygdala

Now that we have limited the set of elements to specific principle neurons in the BLA, the most critical question emerges. How do we know that an element was sampled during stimulus exposure? In stimulus sampling theory an element can hold one of two states, sampled or not, so we are looking for something binary. To be able to relate to stimulus sampling theory a large number of elements, ideally the entire population, would be tracked with some measure of neural activity that corresponds to being sampling. Here we propose that transcription of the immediate early gene coding for activity-regulated cytoskeletal protein (Arc), as identified through fluorescent in situ hybridization (FISH), is a useful indicator of this activational status.

The selection of Arc is based on several features of this immediate early gene. First, it is expressed in pyramidal neurons but not interneurons. Arc is an effector gene whose RNA is transported from the nucleus to the dendrites, where it regulates postsynaptic changes necessary for both learning and synaptic plasticity (Bramham et al., 2008). It has an advantage over electrophysiology for our purposes, because electrophysiology at its best can record only from a few hundred neurons, while post-mortem processing of the Arc signal can potentially assess the entire population of relevant neurons. The advantage of Arc over other immediate early genes such as cfos is twofold. First, as mentioned above it is limited to projection neurons. Second, and most important, the expression and movement of Arc mRNA from the nucleus to the cytoplasm has a distinct and rapid time course that allows

identification of neural activity limited to brief epochs (Figure 3). Almost immediately upon relevant neural activity Arc signal can be identified as transcription foci inside the nuclei of neurons (Pevzner et al., 2012). If that neuron does not remain active transcription stops, and the transcription foci disappear 15 min later (Vazdarjanova et al., 2002). The newly transcribed Arc mRNA moves into the cytoplasm and can be observed as a penumbra that surrounds the nucleus starting about 20 min after the initial activity (Guzowski & Worley, 2001). Therefore, when examining images on a confocal microscope, any neuron with transcription foci (nuclear) was active 3 sec to 5 minutes before the brains were taken. Any neuron showing a cytoplasmic signal surrounding the nucleus was active 20–25 minutes ago. Neurons active at both time points will be positive for Arc in both compartments. If one strategically places two events with respect to what cellular compartment should contain Arc, one can determine what neurons were active during the first, the second or both events (Figure 3).

This imaging technique is called **c**ellular compartmental analysis of **t**emporal activity using **FISH** (catFISH). Using this technique, Guzowski et al. (1999) gave rats two 5 min exposures to a context. The exposures were separated by 20 min and were either to the same or different environments. When the environments were the same about 40% of hippocampal CA1 neurons showed double labeling with very few showing signal in only one compartment. On the other hand, rats going into different compartments showed only 15% double labeling. The greater overlap in activation for the same condition is consistent with the idea that each environment activates a unique, but partially overlapping, set of elements, which is a core assumption of stimulus sampling theory. Another aspect of Guzowski et al's finding is that the number of neurons responding to both contexts (15%) was what would be expected by chance if both contexts randomly sampled 40% of CA1 neurons. Thus Arc is behaving in a manner that matches what would be expected from a stimulus element being sampled. Guzowski's work focused on the hippocampus rather than the amygdala. While the hippocampus is important for building a representation of the context, it does not represent the context-shock association; hence our focus on the BLA. Importantly, using Arc catFISH, Barot et al (2008; 2009), found that BLA neurons respond to both CS and US.

Therefore, we used catFISH to examine the change in Arc expression as learning progresses (Zelikowsky et al., submitted). Since catFISH is limited to two time points, we can only look at two points along the learning curve. Both in theory (assuming a negatively accelerated learning curve) and based on the data described above (Figures 1 and 2) the greatest change in behavior should occur between the first two trials. Therefore, we looked at one trial contextual fear conditioning, when the first event for catFISH was a 5 min context exposure that ended with a single shock (Zelikowsky et al., submitted). The second event was a 5 min test in the same context, 20 min later. An ideal control group should have the same CS (context) and US (shock) exposure but in a manner that should not foster a CS-US association. There is extensive data on the immediate shock deficit, which shows that a shock given simultaneously with placement in the chamber does not result in learning a context-shock association, so this condition provided an ideal control group for our purposes (Fanselow, 1986b; 1990). Figure 3A shows the design and its relationship to the catFISH protocol. Freezing behavior scored during the test provided evidence that the delayed-shock rats acquired an association, freezing at 36% (±4.4%0), while the immediate-shock rats did not, freezing at 1.0% (± 0.7%). Indeed, there was no overlap in the distribution of freezing scores between the two groups ($U(8,8)=64$, $p<.001$).

Immediately after behavioral testing, brains were removed for catFISH processing (Figure 3B). Quantification of the Arc data for the immediate-shock control group is in figure 3C. For the immediate shock rats each context exposure activated Arc in about 14% of the

neurons. This was split almost equally between cells activated just during the first exposure (8%, Cytoplasmic) or just during the second exposure (6%, Nuclear) with very few neurons expressing Arc to both events (0.3%, double). The tendency in this control group seems to be that if a BLA neuron is active on the first exposure (during conditioning) it is less likely to be reactivated on the second exposure.

A totally different pattern emerged in the conditioned (Delayed-shock) rats (Figure 3D). There were no neurons activated only during the training event (0%, cytoplasmic); every neuron that was active during the effective context-US pairing reactivated during test (8.3%). Thus while both conditions **activated** a similar total number of neurons during training (8.4%) the likelihood of **reactivation** at test was far greater in the group that acquired fear. Interestingly, during the test both groups recruited a new population of neurons (5.6%, nuclear only). The BLA differentially reactivates neurons based on their past learning history.

Discussion

Arc expression patterns indicated that amygdala neurons are activated by exposure to a novel environment. If that exposure is effectively reinforced, as it was in the delayed shock condition, there is complete reactivation of those previously reinforced neurons. Such reactivation may reflect increased attention to the cues that predict shock (Mackintosh, 1975) or possibly the formation of associations between stimulus elements that were concurrently active (e.g., McLaren et al., 1989; Wagner, 1981). However, it is important to note that the amygdala only resampled elements that were reinforced. In the immediate shock group, activation of neurons during the first exposure reduced the likelihood of reactivation on the second exposure.

These BLA data are strikingly different than what happens in hippocampus with 2 placements in the same environment. There, the majority of activated cells show double labeling regardless of whether the experience was reinforced or not (Guzowski et al., 1999, Zelikowsky et al., submitted). This is consistent with the hippocampus rapidly forming an integrated, Gestalt-like representation of the current environment (Fanselow, 2000). The hippocampus provides this information to the amygdala (Maren et al., 1995) so that it can be used as a CS to associate with the US. One complication for this view is that it suggests that while the hippocampus' representation of a place is configural, the amygdala's representation of context-US associations is elemental. The difference in the nature of processing between the amygdala and hippocampus may reflect a difference in what each structure evolved to accomplish. The hippocampus creates a detailed representation of a place so that it can be successfully remembered and navigated. On the other hand, the amygdala attaches affective salience to the best predictors of biologically significant events.

This distinction between the hippocampus as a configural processor, but the amygdala as an elemental processor for Pavlovian conditioning harkens back to Sutherland and Rudy's configural model of the hippocampus (1990). In Pavlovian conditioning elemental error-correction processes will tend to limit associative strength to the most salient and/or best predictors (e.g., Rescorla & Wagner, 1972). The nature of episodic memory, which has been strongly linked with the hippocampus, is quite different (Bierley et al., 1983; Eldrige et al., 2000; Tulving & Thompson, 1973). Episodic memory is rich in detail containing aspects of place, time and specific events. It is defined by having all of these components included in the same representation. Rather than being inclusive, Pavlovian representations are exclusive and limited to either the best predictor or the most salient of equally predictive stimuli. Contexts, which are comprised of individual elements that are not particularly predictive, would not fare well in the competitive environment of error correction (Rescorla,

1972). The hippocampus may overcome this lack of salience of the contextual elements by binding them together into a more salient whole that is better able to compete with explicit cues such as a tone. Certainly, further research is needed to see how the hippocampal representation of place is transformed to the amygdala's representation of that place being dangerous.

Incremental Learning

Our freezing data strongly suggest that fear learning is an incremental process (Figures 1 and 2). Based on the molecular imaging data we saw that a conditioning trial recruits BLA neurons into a network that is responsive to the CS. So a possibility is that as learning progresses, additional neurons are recruited to the memory representation. As more neurons are recruited a stronger fear response can be generated. Ultimately, asymptote will be reached when all the neurons that have the potential to encode the CS-US association are recruited into the representational network. To determine if individual neurons are reactivated as part of a memory representation a within-subjects approach is necessary. The catFISH technique allows such within-subjects contrast but is limited to two time points. We chose to image the first 2 trials of acquisition because that is the point at which we observed the largest increment in behavior (Figure 1 & 2). Future research should compare other points along the learning curve. Techniques that allow retrospective analysis of neuronal activity at multiple time points would facilitate such efforts.

One can make some interesting extrapolations from these data. For example, during the second context exposure, both groups activated (sampled) a number of neurons that had not been activated previously. These neurons are available to encode the outcome of trial 2. For the conditioned rats, if this happens on subsequent trials with a similar outcome, a larger and larger percentage of the population of BLA neurons would become associated with shock. The learning curve increments as a greater proportion of the total population of elements (neurons) becomes associated with shock. This is exactly the pattern that stimulus sampling theory predicts. It also suggests that the number of BLA neurons activated may be isomorphic with associative strength.

In marked contrast to the conditioned animal's perfect (100%) reactivation of neurons there was virtually no reactivation of neurons in the controls (0.3%). If this happens for many successive nonreinforced trials it may leave few neurons to activate if a shock were to occur, meaning that nonreinforced trials could impede subsequent conditioning. This would provide a ready explanation of the ability for CS pre-exposure to retard learning that is referred to as latent inhibition (Lubow, 1973). However, it should be noted that latent inhibition has been difficult to demonstrate with contextual cues (Young & Fanselow, 1992).

Implications for Stimulus Sampling Theory

Earlier we suggested that discovering an isomorphism between psychological process and neural mechanism might immediately suggest a modification to existing theories. And indeed, if one accepts the premise that BLA neurons are stimulus elements and Arc indicates recent sampling of those elements, the catFISH data suggest changes in the assumptions of stimulus sampling theory. For example, Estes assumed that the probability of sampling an element did not change from trial to trial. Within the amygdala, however, reinforcement of a neuron led to an increase in the likelihood of being resampled (indicated by high double labeling in the conditioned group) and/or a decrease in the probability of being sampled when there was no outcome (indicated by little double labeling in the immediate shock group). Thus the random sampling of elements does not occur with replacement—resampling is altered based upon past experience. Our two epochs are somewhat close to

each other so we cannot tell if these altered probabilities of resampling are permanent or dissipate with time. In this regard, Wagner has proposed an elemental model that specifically states that once a stimulus element has been activated it temporally becomes refractory to reactivation and that administration of a new stimulus (e.g., shock) may reduce this refractory period (Wagner, 1981; Wagner et al., 1973).

A fundamental assumption in Estes theory is that that on any given trial some elements are sampled (active) and some are not. When we suggest that a BLA neuron is isomorphic with a stimulus element we can ask the question if there is some mechanistic account of why one element is sampled and one is not. Recent work by the Josselyn and Silva laboratories suggest that levels of the transcription factor cAMP response element-binding protein (CREB) determine whether or not a BLA neuron will get incorporated into a memory by enhancing the neuron's excitablility (Josselyn, 2010; Zhou et al., 2009; Won & Silva, 2008). Importantly, they showed that increasing CREB in a neuron increases the probability that the neuron will express Arc (Han et al., 2007). Thus CREB levels may predispose a neuron to be sampled and Arc can be used to assess whether or not the neuron was sampled.

Typically Psychological Process models of learning are cast in generalist terms; they are assumed to work more-or-less identically in all learning systems. The work of Bolles (e.g., Bolles, 1970), Garcia (Garcia & Koeling (1966)) and others (Timberlake & Fanselow, 1994) argued that a better understanding comes when we think of a type of learning as specialized for its functional context. For example, Bolles moved our understanding of fear and avoidance from a general reinforcement process view to one of fear as a defensive behavior system (Bolles, 1972; 1975). It could be argued that the reconceptualization of learning into separable systems serving different functions is what allowed the neuroscience of fear to develop the way it did (LeDoux, 2012). Indeed, the finding that different types of learning (e.g., fear, eyeblink, reward) rely on different neural circuits (e.g., amygdala, cerebellum, striatum, respectively) offers direct support for the specialist position. Having knowledge of the structure and function of the circuit led us to identify the amygdala as a place to look for something that corresponded to Guthrie and Estes hypothetical stimulus elements. In turn, cellular imaging of a putative neural mechanism for this hypothetical construct led to the possible refinement of the psychological process model. This back and forth between process and mechanistic accounts holds great promise for deepening our understanding of learning generally. This special issue shows that this exciting effort is underway.

Acknowledgments

Supported by NIMH Grant RO1 MH62122 to MSF.

References

Barot SK, Kyono Y, Clark EW, Bernstein IL. Visualizing stimulus convergence in amygdala neurons during associative learning. Proceedings of the National Academy of Sciences of the United States of America. 2008; 105(52):20959–20963. [PubMed: 19091953]

Barot SK, Chung A, Kim JJ, Bernstein IL. Functional imaging of stimulus convergence in amygdalar neurons during Pavlovian fear conditioning. PLoS ONE. 2009; 4(7):e6156. [PubMed: 19582153]

Bierley RA, Kesner RP, Novak JM. Episodic long-term memory in the rat: effects of hippocampal stimulation. Behavioral Neuroscience. 1983; 97:42–8. [PubMed: 6838725]

Bolles RC. Species-specific defense reactions and avoidance learning. Psychological Review. 1970; 77(1):32–48.

Bolles RC. Reinforcement, expectancy, and learning. Psychological Review. 1972; 79(5):394–409.

Bolles, RC. Theory of motivation. 2. New York: Harper & Row; 1975.

Bolles RC, Fanselow MS. A perceptual-defensive-recuperative model of fear and pain. Behavioral & Brain Sciences. 1980; 3(2):291–301.

Bramham CR, Worley PF, Moore MJ, Guzowski JF. The immediate early gene arc/arg3.1: regulation, mechanisms, and function. Journal of Neuroscience. 2008; 28(46):11760–11767. [PubMed: 19005037]

Bush RR, Mosteller F. A mathematical model for simple learning. Psychological Review. 1951; 58:313–323. [PubMed: 14883244]

Chung A, Barot SK, Kim JJ, Bernstein IL. Biologically predisposed learning and selective associations in amygdalar neurons. Learning and Memory. 2011; 18(6):371–374. [PubMed: 21576517]

Cushman JD, Maldonado J, Kwon EE, Garcia AD, Fan G, Imura T, et al. Juvenile neurogenesis makes essential contributions to adult brain structure and plays a sex-dependent role in fear memories. Frontiers in Behavioral Neuroscience. 2012; 6(3):1–17. [PubMed: 22279431]

Ehrlich I, Humeau Y, Grenier F, Ciocchi S, Herry C, Lüthi A. Amygdala inhibitory circuits and the control of fear memory. Neuron. 2009; 62(6):757–771. [PubMed: 19555645]

Eldridge LL, Knowlton BJ, Furmanski CS, Bookheimer SY, Engel SA. Remembering episodes: a selective role for the hippocampus during retrieval. Nature Neuroscience. 2000; 3:1149–52.

Estes WK. Toward a statistical theory of learning. Psychological Review. 1950; 57:94–207.

Fanselow MS. Conditioned fear-induced opiate analgesia: A competing motivational state theory of stress-analgesia. Annals of the New York Academy of Sciences. 1986a; 467:40–54. [PubMed: 3524387]

Fanselow MS. Associative vs. topographical accounts of the immediate shock freezing deficit in rats: Implications for the response selection rules governing species specific defensive reactions. Learning & Motivation. 1986b; 17:16–39.

Fanselow MS. Factors governing one trial contextual conditioning. Animal Learning & Behavior. 1990; 18:264–270.

Fanselow MS. Pavlovian conditioning, negative feedback, and blocking: Mechanisms that regulate association formation. Neuron. 1998; 20:625–627. [PubMed: 9581755]

Fanselow MS. Contextual fear, gestalt memories, and the hippocampus. Behavioural Brain Research. 2000; 110:73–81. [PubMed: 10802305]

Fanselow MS, Bolles RC. Triggering of the endorphin analgesic reaction by a cue previously associated with shock: Reversal by naloxone. Bulletin of the Psychonomic Society. 1979; 14:88–90.

Fanselow MS, LeDoux JE. Why we think plasticity underlying Pavlovian fear conditioning occurs in the basolateral amygdala. Neuron. 1999; 23:229–232. [PubMed: 10399930]

Fechner, GT. Elemente der Psychophysik. Vol. 2. Leipzig: Breitkopf und Härtel; 1860. p. 559(Reprinted, Bristol: Thoemmes Press, 1999)

Fendt M, Fanselow MS. The neuroanatomical and neurochemical basis of conditioned fear. Neuroscience & Biobehavioral Reviews. 1999; 23:743–760. [PubMed: 10392663]

Gallistel, CR. The organization of learning. Cambridge, MA: MIT Press; 1990.

Gallistel CR. Flawed foundations of associationism? Comment on Machado and Silva (2007). American Psychologist. 2007; 62:682–685. [PubMed: 17924751]

Garcia J, Koelling RA. Relation of cue to consequence in avoidance learning. Psychonomic Science. 1966; 4(3):123–124.

Gottlieb DA. Is the number of trials a primary determinant of conditioned responding? Journal of Experimental Psychology: Animal Behavior Processes. 2008; 34:185–201. [PubMed: 18426303]

Gottlieb DA, Rescorla RA. Within-subject effects of number of trials in rat conditioning procedures. Journal of Experimental Psychology: Animal Behavior Processes. 2010; 36:217–231. [PubMed: 20384402]

Guthrie, ER. The psychology of learning. New York: Harper & Brothers; 1935.

Guzowski JF, McNaughton BL, Barnes CA, Worley PF. Environment-specific expression of the immediate-early gene Arc in hippocampal neuronal ensembles. Nature Neuroscience. 1999; 2(12): 1120–1124.

Guzowski JF, Worley PF. Cellular compartment analysis of temporal activity by fluorescence in situ hybridization (catFISH). Current Protocols in Neuroscience. 2001:1.8.1–1.8.16.

Hall G, Pearce JM. Latent inhibition of a CS during CS-US pairings. Journal of Experimental Psychology: Animal Behavior Processes. 1979; 5:31–42. [PubMed: 528877]

Han JH, Kushner SA, Yiu AP, Cole CJ, Matynia A, Brown RA, et al. Neuronal competition and selection during memory formation. Science. 2007; 316(5823):457–460. [PubMed: 17446403]

Haubensak W, Kunwar P, Cai H, Ciocchi S, Wall N, Ponnusamy R, et al. Genetic dissection of an amygdala microcircuit that gates conditioned fear. Nature. 2010; 468:270–276. [PubMed: 21068836]

Helmstetter FJ, Fanselow MS. Effects of naltrexone on learning and performance of conditional fear-induced freezing and opioid analgesia. Physiology & Behavior. 1987; 39:501–505. [PubMed: 3575497]

Hull, C., et al. Mathematico-Deductive Theory of Rote Learning. New Haven, NJ: Yale University Press; 1940.

Jacobs N, Cushman J, Fanselow MS. The Accurate Measurement of Fear Memory in Pavlovian Conditioning: Resolving The Baseline Issue. Journal of Neuroscience Methods. 2010; 190:235–239. [PubMed: 20471421]

Johansen JP, Tarpley JW, LeDoux JE, Blair HT. Neural substrates for expectation-modulated fear learning in the amygdala and periaqueductal gray. Nature Neuroscience. 2010; 13(8):979–986.

Josselyn SA. Continuing the search for the engram: examining the mechanism of fear memories. Journal of Psychiatry & Neuroscience. 2010; 35(4):221–228. [PubMed: 20569648]

Kamin, LJ. “Attention-like” processes in classical conditioning. In: Jones, MR., editor. Miami symposium on the prediction of behavior: Aversive Stimulation. Coral Gables, Florida: University of Miami Press; 1968. p. 9-31.

Kim JJ, Krupa DJ, Thompson RF. Inhibitory cerebello-olivary projections and blocking effect in classical conditioning. Science. 1998; 279(5350):570–573. [PubMed: 9438852]

Krechevsky I. ”Hypotheses” in rats. Psychological Review. 1932; 39(6):516–532.

Krechevsky I. A study of the continuity of the problem-solving process. Psychological Review. 1938; 45(2):107–133.

Lashely, KS. Brain Mechanisms and Intelligence. Chicago: The University Press; 1929.

LeDoux J. Rethinking the emotional brain. Neuron. 2012; 73(4):653–676. [PubMed: 22365542]

Lubow RE. Latent inhibition. Psychological Bulletin. 1973; 79(6):398–407. [PubMed: 4575029]

Mackintosh NJ. A theory of attention: Variations in the associability of stimuli with reinforcement. Psychological Review. 1975; 82(4):276–298.

Maren S, Fanselow MS. Synaptic plasticity in the basolateral amygdala induced by hippocampal formation stimulation in vivo. Journal of Neuroscience. 1995; 15:7548–7564. [PubMed: 7472506]

McLaren, IPL.; Kaye, H.; Mackintosh, NJ. An associative theory of the representation of stimuli: Applications to perceptual learning and latent inhibition. In: Morris, RGM., editor. Parallel distributed processing: Implications for psychology and neurobiology. Oxford: Oxford University Press; 1989. p. 102-130.

McNally GP, Johansen JP, Blair HT. Placing prediction into the fear circuit. Trends in Neuroscience. 2011; 34(6):283–292.

McNally GP, Pigg M, Weidemann G. Blocking, unblocking, and overexpectation of fear: A role for opioid receptors in the regulation of Pavlovian association formation. Behavioral Neuroscience. 2004; 118(1):111–120. [PubMed: 14979787]

Moga DE, Calhoun ME, Chowdhury A, Worley P, Morrison JH, Shapiro ML. Activity-regulated cytoskeletal-associated protein is localized to recently activated excitatory synapses. Journal of Neuroscience. 2004; 125(1):7–11.

Newton JR, Ellsworth C, Miyakawa T, Tonegawa S, Sur M. Acceleration of visually cued conditioned fear through the auditory pathway. Nature Neuroscience. 2004; 7(9):968–973.

Nicoll RA, Malenka RC. Expression mechanisms underlying NMDA receptor dependent long-term potentiation. Annals of the New York Academy of Sciences. 1999; 868:515–525. [PubMed: 10414328]

Paré D, Quirk GJ, LeDoux JE. New vistas on amygdala networks in conditioned fear. Journal of Neurophysiology. 2004; 92(1):1–9. [PubMed: 15212433]

Pearce JM, Hall G. A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. Psychological Review. 1980; 87(6):532–552. [PubMed: 7443916]

Pevzner A, Miyashita T, Schiffman AJ, Guzowski JF. Temporal dynamics of Arc gene induction in hippocampus: relationship to context memory formation. Neurobiology of Learning & Memory. 2012; 97(3):313–320. [PubMed: 22390855]

Rau V, Decola JP, Fanselow MS. Stress-induced enhancement of fear learning: An animal model of posttraumatic stress disorder. Neuroscience and Behavioral Research. 2005; 29:1207–1223.

Rescorla, RA. Informational variables in Pavlovian conditioning. In: Bower, GH., editor. The Psychology of Learning & Motivation: Advances in research and theory. New York: Academic Press; 1972. p. 1-46.

Rescorla RA. Stimulus generalization: Some predictions from a model of Pavlovian conditioning. Journal of Experimental Psychology: Animal Behavior Processes. 1976; 2(1):88–96. [PubMed: 1249526]

Rescorla, RA.; Wagner, AR. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In: Black, AH.; Prokasy, WF., editors. Classical Conditioning II: current research and theory. New York: Appleton-Century-Crofts; 1972. p. 64-99.

Riley, DA. Discrimination Learning. Boston: Allyn & Bacon; 1968.

Romanski LM, Clugnet MC, Bordi F, LeDoux JE. Somatosensory and auditory convergence in the lateral nucleus of the amygdala. Behavioral Neuroscience. 1993; 107(3):444–450. [PubMed: 8329134]

Rudy, JW.; Wagner, AR. Stimulus selection in associative learning. In: Estes, WK., editor. Handbook of learning and cognitive processes. Vol. 2. Hillsdale, NJ: Erlbaum; 1975. p. 269-303.

Ryan SJ, Ehrlich DE, Jasnow AM, Daftary S, Madsen TE, Rainnie DG. Spike-timing precision and neuronal synchrony are enhanced by an interaction between synaptic inhibition and membrane oscillations in the amygdala. PLoS ONE. 2012; 7(4):e35320. [PubMed: 22563382]

Schultz W, Dickinson A. Neuronal coding of prediction errors. Annual Review of Neuroscience. 2000; 23:473–500.

Setlow B, Gallagher M, Holland PC. The basolateral complex of the amygdala is necessary for acquisition but not expression of CS motivational value in appetitive Pavlovian second-order conditioning. European Journal of Neuroscience. 2002; 15(11):1841–1853. [PubMed: 12081664]

Spence KW. Continuous versus non-continuous interpretations of discrimination learning. Psychological Review. 1940; 47(4):271–288.

Stevens SS. On the psychophysical law. Psychological Review. 1957; 64(3):153–181. [PubMed: 13441853]

Sutherland RJ, Rudy JW. Configural association theory: The role of the hippocampal formation in learning, memory, and amnesia. Psychobiology. 1989; 17:129–144.

Timberlake W, Fanselow MS. Symposium on behavior systems: Learning, neurophysiology, and development. Psychonomic Bulletin & Review. 1994; 1:403–404.

Tulving E, Thomson DM. Encoding specificity and retrieval processes in episodic memory. Psychological Review. 1973; 80:352–373.

Vazdarjanova A, McNaughton BL, Barnes CA, Worley PF, Guzowski JF. Experience-dependent coincident expression of the effector immediate-early genes arc and Homer 1a in hippocampal and neocortical neuronal networks. Journal of Neuroscience. 2002; 22(23):10067–10071. [PubMed: 12451105]

Waelti P, Dickinson A, Schultz W. Dopamine responses comply with basic assumptions of formal learning theory. Nature. 2001; 412(6842):43–48. [PubMed: 11452299]

Wagner AR, Rudy JW, Whitlow JW. Rehearsal in animal conditioning. Journal of Experimental Psychology. 1973; 97(3):407–426. [PubMed: 4705247]

Wagner, AR. SOP: A model of automatic memory processing in animal behavior. In: Spear, NE.; Miller, RR., editors. Information processing in animals: Memory mechanisms. Hillsdale, NJ: Erlbaum; 1981. p. 5-47.

Wagner AR. Context-sensitive elemental theory. Quarterly Journal of Experimental Psychology. 2003; 56B:7–29. [PubMed: 12623534]

Wagner AR. Evolution of an elemental theory of Pavlovian conditioning. Learning & Behavior. 2008; 36(3):253–265. [PubMed: 18683469]

Weber, EH. De pulsu, resorptione, auditu et tactu. Anatationes anatomicae et physiologicae. Leipzig: Koehler; 1834.

Won J, Silva AJ. Molecular and cellular mechanisms of memory allocation in neuronetworks. Neurobiology of Learning and Memory. 2008; 89(3):285–292. [PubMed: 17962049]

Young SL, Fanselow MS. Associative regulation of Pavlovian fear conditioning: US intensity, incentive shifts and latent inhibition. Journal of Experimental Psychology: Animal Behavior Processes. 1992; 18:400–413. [PubMed: 1402696]

Zelikowsky, M.; Chawla, MK.; Hersman, S.; Barnes, CA.; Fanselow, MS. Neuronal ensembles in amygdala, hippocampus and prefrontal cortex track differential components of contextual fear. (submitted)

Zelikowsky M, Fanselow MS. Opioid regulation of Pavlovian overshadowing in fear conditioning. Behavioral Neuroscience. 2010; 124:510–519. [PubMed: 20695650]

Zhou Y, Won J, Karlsson MG, Zhou M, Rogerson T, Balaji J, Neve R, Poirazi P, Silva AJ. CREB regulates excitability and the allocation of memory to subsets of neurons in the amygdala. Nature Neuroscience. 2009; 12(11):1438–1443.

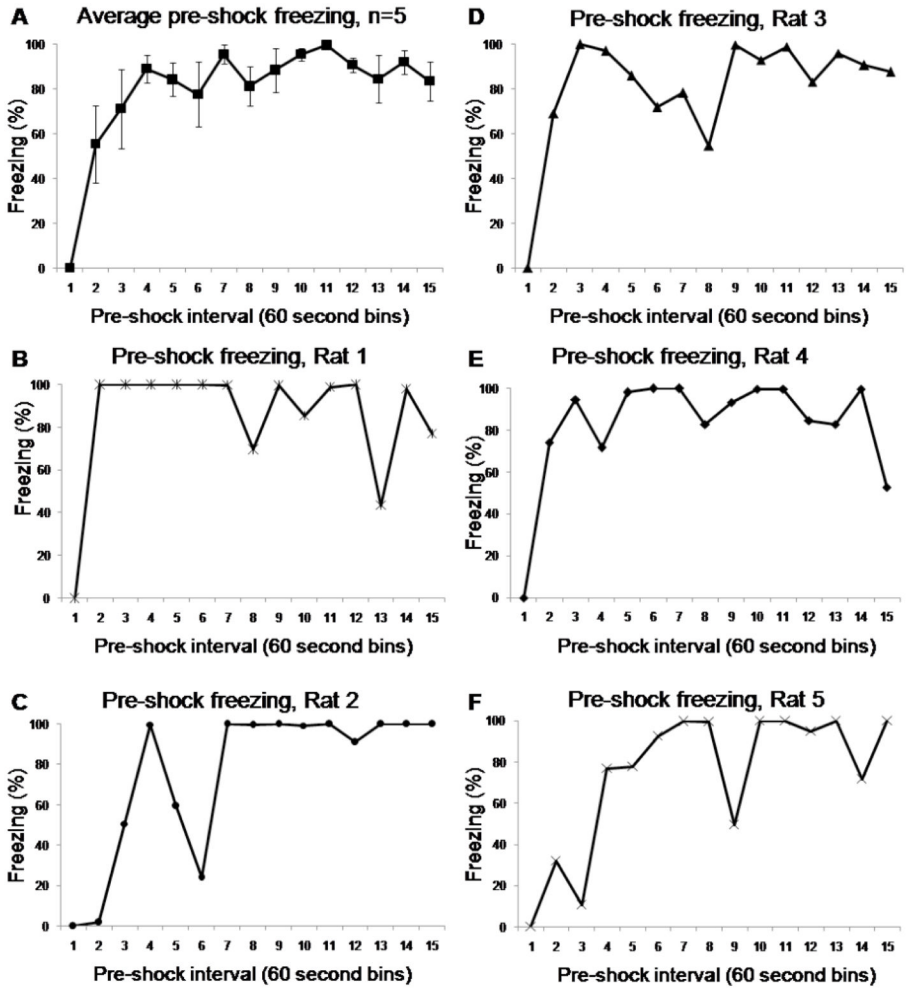

Figure 1.

Rats received 15 shocks spaced randomly over 90 min (see Rau et al., 2005 for a full description of the procedure). The data presented are for time spent freezing during a 1 min period prior to each shock, which is a measure of context fear acquisition. Panel A presents the average acquisition curve of 5 rats. Panels D–F show the data of each individual subject that contributed to the composite curve of Panel A.

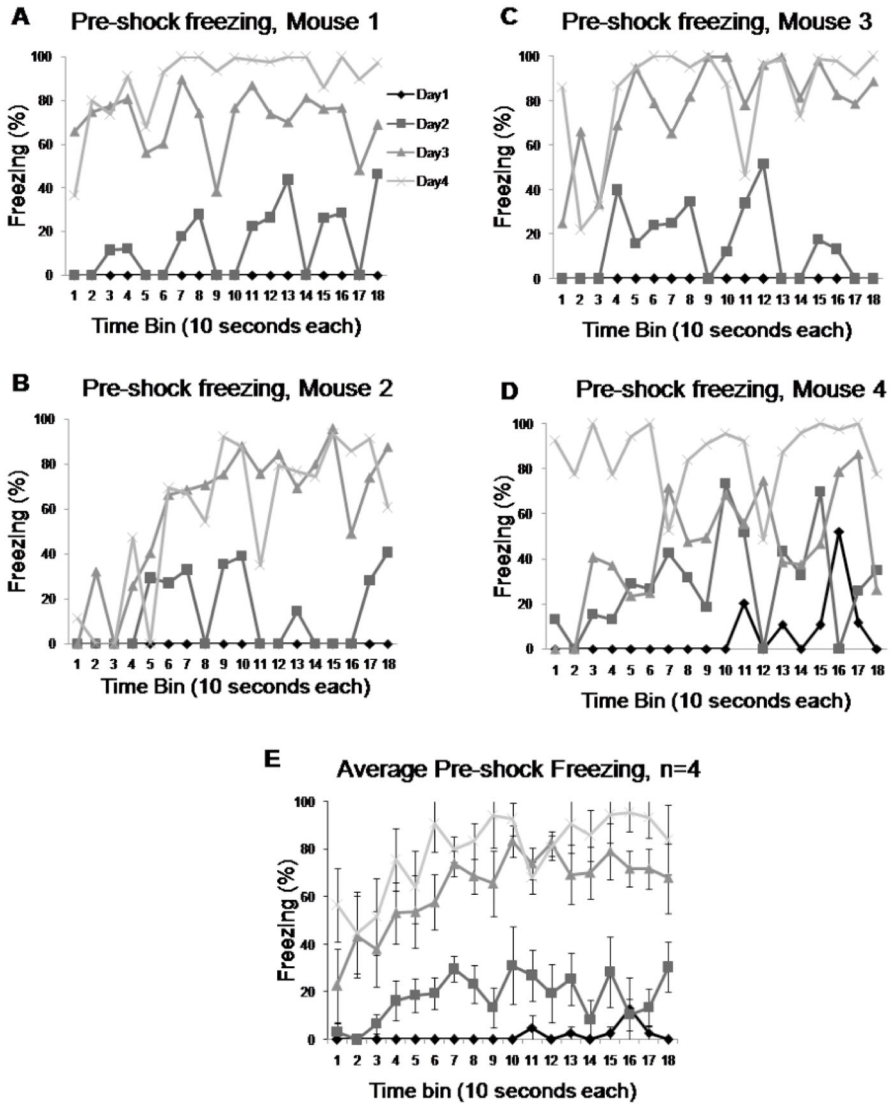

Figure 2.

Mice (c57Bl6) were placed in a context once a day for 4 Days. Three min after placement in the chamber they received a shock (see Cushman et al., 2012 for a more complete description). Each figure shows freezing during 10 sec bins taken through the 3 min context exposure. Each of the 4 trials is plotted separately on the same graph. Panels A–D show individual mice and Panel E shows the composite curve for those 4 animals.

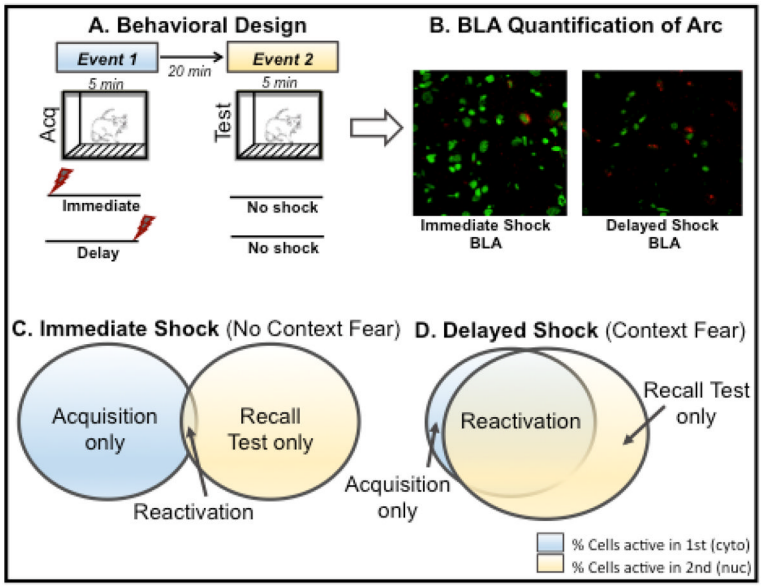

Figure 3.

Panel A: The catFISH technique (***c***ellular compartment ***a***nalysis of ***t***emporal activity by ***f***luorescence ***i***n ***s***itu ***h***ybridization) takes advantage of the unique timecourse of Arc mRNA as it translocates from nucleus (0–5 min after activity) to cytoplasm (20 min after activity) to quantify the entire population of neurons activated by the first behavioral episode (cytoplasmic), a second episode (nuclear), or both (cytoplasmic and nuclear). Comparing rats shocked immediately after placement in a context (Immediate) with rats given time to form a contextual representation (Delay) is a method to demonstrate associative fear memory formation while controlling for the effects of the shock (see text for data). Panel B: A confocal image taken from an immediate (left) and a delayed shock rat showing BLA nuclei in green (sytox) and Arc mRNA in red. Panels C and D: Venn diagrams indicating the proportion of neurons active at the first event (training), second event (testing), or both (overlap in circles). In the BLA of the immediate group, independent populations of neurons were active with each contextual experience (Panel C). In contrast, the Delay group has almost complete reactivation of the first cohort of neurons, with the addition of a new population of units during recall (Panel D). These two populations in the learning group have characteristics of Estes' theory of stimulus sampling, and may relate to new aspects of the context sampled during the second experience.