CellPress

Categories, concepts, and conditioning: how humans generalize fear

Joseph E. Dunsmoor and Gregory L. Murphy

Psychology Department, New York University, New York, NY 10003, USA

During the past century, Pavlovian conditioning has served as the predominant experimental paradigm and theoretical framework to understand how humans learn to fear and avoid real or perceived dangers. Animal models for translational research offer insight into basic behavioral and neurophysiological factors mediating the acquisition, expression, inhibition, and generalization of fear. However, it is important to consider the limits of traditional animal models when applied to humans. Here, we focus on the question of how humans generalize fear. We propose that to understand fear generalization in humans requires taking into account research on higher-level cognition such as category-based induction, inferential reasoning, and representation of conceptual knowledge. Doing so will open the door for productive avenues of new research.

The problem of fear generalization

One of the most important challenges animals face is how to detect and react to threat. Classical conditioning is an elegant and evolutionarily conserved form of learning that animals possess to handle this challenge. In fear conditioning, a stimulus associated with threat begins to elicit a defensive response. However, if this process is overly specific, animals will later fail the challenge of facing threat in a dynamic environment where stimuli rarely assume the same exact form from one encounter to the next. Humans possess a remarkable ability to interpret the perceptual and conceptual details of a learning episode, allowing them to generalize learned behavior to a host of different stimuli. For example, being stung by a bee could lead one to avoid other bees and wasps that are similar to the original stinger. In this case, the generalization seems wise. In other cases, generalization may be maladaptive. For example, a harrowing automobile accident can lead to a fear and avoidance of driving or riding in cars, the neighborhood where the accident occurred, road signs or other symbols of driving, car chases in movies or TV shows, the sound of jingling keys, and other idiosyncratic associations of automobiles or accidents [1]. This is just one example of how fear is rarely confined to a specific object or event and how, when generalization goes awry, information that shares a seemingly irrelevant association can nonetheless provoke an emotional reaction.

In this article we discuss how understanding the complexity of human fear generalization demands going beyond traditional models of Pavlovian conditioning and stimulus generalization honed over the past century. We propose that fear conditioning research in humans should incorporate theoretical knowledge and experimental approaches from other domains of psychology, in particular the categories and concepts literature, where there is an established body of work investigating factors promoting the generalization of human knowledge. Integrating research on Pavlovian fear conditioning with theoretical knowledge and experimental approaches from other domains of psychology will provide a better framework to understand real-world generalization of fear learning. Fortunately, there is a rich theoretical and empirical foundation of research on conceptual processes in humans, and a number of useful approaches have been developed to examine how humans generalize knowledge.

Traditional models of fear learning and generalization

Pavlovian fear conditioning in laboratory animals is a productive area of research that continues to offer detailed insight into the behavioral and neurophysiological processes underlying how neutral conditioned stimuli (CS; e.g., a tone) become associated with aversive unconditioned stimuli (US; e.g., an electrical shock) to produce a conditioned fear response (CR; e.g., an increase in sweating or freezing in place). Research in the neuroscience of fear conditioning shows how simple sensory information from the CS and US converge in the lateral amygdala, leading to an increase in synaptic plasticity such that the CS itself evokes amygdala activity [2,3]. The amygdala initiates fear responses through output connections with the hypothalamus, brainstem, and other areas involved in responding to threat [4]. While neuroanatomical models of fear conditioning have been successfully extended to human research over the past several decades, advances in this line of research continue to rely overwhelmingly on rodent studies that incorporate simple stimuli like lights and tones.

A predominant concern since the earliest studies of classical conditioning is how conditioned learning generalizes [5]. Using appetitive cues, Pavlov long ago observed that the CR is not confined to the training CS, but instead

Corresponding authors: Dunsmoor, J.E. (joseph.dunsmoor@nyu.edu); Murphy, G.L. (gregory.murphy@nyu.edu).

1364-6613/
 http://dx.doi.org/10.1016/j.tics.2014.12.003

generalizes to other stimuli that have never been paired with the US (Figure 1A). Landmark studies in the mid-20th century turned to appetitive operant conditioning to reveal ordered gradients of generalized instrumental responses as a function of perceptual similarity to the CS [6].

In the past several years, models of stimulus generalization developed for animal learning studies have been adapted to the study of fear generalization in humans [7–9]. This research measures fear generalization by gradients of autonomic responses, like skin conductance responses (SCR, i.e., sweating) or fear-potentiated startle. Fear generalization research in humans provides important clinical translational value for evaluating overgeneralization of defensive responses characteristic of psychopathologies for which fear and anxiety are widespread, including post-traumatic stress disorder (PTSD), obsessive-compulsive disorder, and panic disorder [10,11].

Much of the research in the nascent field of human fear generalization uses simple sensory cues like colors or shapes. This approach is in line with historical studies of stimulus generalization in laboratory animals and allows generalization to be measured as a function of similarity to the original CS along a definable sensory dimension. Yet, real-world fear learning situations tend to involve complex stimuli with multiple dimensions (e.g., a dog), rather than simple unidimensional sensory cues. Moreover, humans routinely incorporate prior conceptual knowledge and apply inductive reasoning to infer unobserved properties and causal structure of details surrounding an emotional event ('Your policy is the cause of this whole fiasco!'). Such processes bring added meaning to emotional experiences by determining our emotional reactions to similar experiences in the future. In this way, traditional models of stimulus generalization underserve the complexity inherent to fear generalization in real-world situations.

The predominant strategy with controlled laboratory paradigms has been to study conditioning with unfamiliar or simple stimuli so that prior experience will not influence learning or generalization. Fear generalization based on the perceptual regularities of unfamiliar or simple stimuli could in fact rely on basic low-level processes devoid of higher-order reasoning, and is already well described by traditional models of Pavlovian conditioning (Figure 1), for example, freezing to a tone of 1000 Hz after being shocked to a tone of 800 Hz [12–14]. However, for humans, most feared stimuli are familiar and are semantically connected to bodies of knowledge (guns, speeding vehicles, criminals, etc.). What is the effect of such knowledge? Traditional approaches to the study of conditioned learning that employs lights and tones cannot tell us how to account for these factors (Box 1). We contend that fear generalization based on real-world events about which people have knowledge will necessarily incorporate higher-order processes, which are not easily accounted for by traditional models of stimulus generalization along a single dimension. Such processes are accounted for in other domains of psychology, which could be used to make predictions for how humans will generalize fear expression following aversive learning experiences.

Categorization

Physically similar objects often share similar underlying properties, explaining why animals ought to generalize what they have learned about one object to other physically similar objects [14]. Humans also transfer knowledge between physically dissimilar objects that are conceptually related—the process of induction. For example, knowledge that dogs and cats give birth to live young can be extended to other mammals, like whales or bats, whose births have never been observed. This conceptual path of generalization could be used in the transfer of conditioned fear behaviors as well, from the CS to other stimuli from the same category that may vary considerably in physical form but could also pose a threat. There have been historically few attempts, however, to connect the literature on the

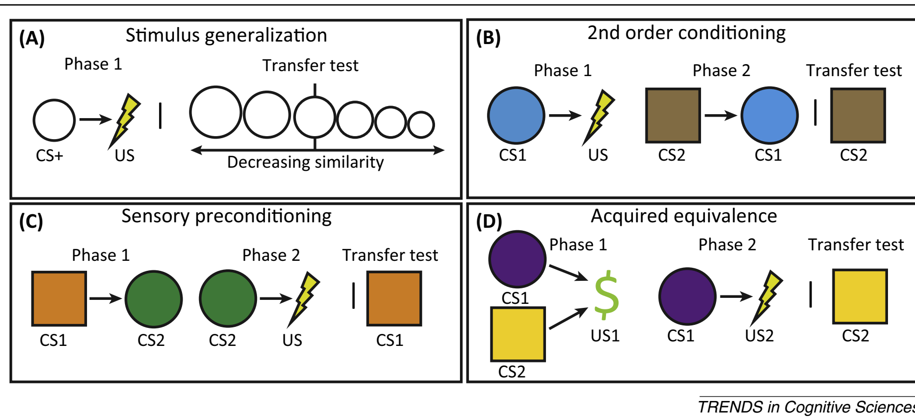

Figure 1. Examples of Pavlovian conditioning techniques traditionally used to investigate the transfer of conditioned learning. In each case, learned (conditioned) responses transfer from one conditioned stimulus (CS) to other stimuli that have not before predicted an unconditioned stimulus (US) – depicted here as an aversive electrical shock (lightning bolt). These techniques have been used to investigate generalization of conditioned learning in a number of different species, including rodents, pigeons, zebra fish, and humans. **(A)** In traditional stimulus generalization paradigms, the response initially conditioned to the conditioned stimulus (CS+) transfers as a function of physical similarity to other stimuli that have not previously predicted the US. **(B)** In second-order conditioning, a CS (CS1) is first paired with the US. CS1 is then paired with another stimulus (CS2), leading to the transfer of the conditioned response from CS1 to CS2. **(C)** Sensory preconditioning involves an initial pairing between two stimuli (CS1 and CS2) in the absence of reinforcement. CS2 is then paired with a US. The initial association between CS1 and CS2 promotes the transfer of conditioned responding when CS1 is later presented alone. **(D)** In acquired equivalence paradigms, dissimilar stimuli (CS1 and CS2) will be treated similarly if they predict the same outcome (US1). In this case, the US1 is rewarding, establishing an approach response. If CS1 is then paired with a different outcome that produces a new conditioned response, such as freezing in anticipation of an electric shock, then CS2 may take on properties associated with the new CS1–US2 relationship as well.