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].

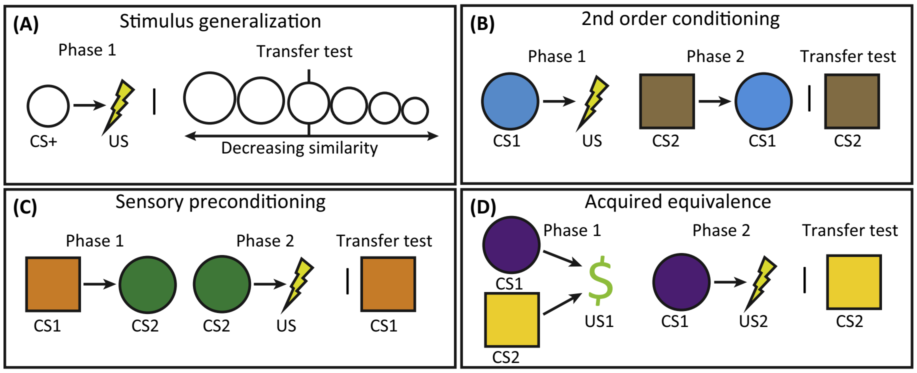
In the past several years, models of stimulus generalization developed for animal learning studies have been adapted to the study of fear generalization in humans [7–9]. This research measures fear generalization by gradients of autonomic responses, like skin conductance responses (SCR, i.e., sweating) 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



TRENDS in Cognitive Sciences

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.

Box 1. Beyond lights and tones: factors known to influence conditioned learning

It has long been recognized that traditional models of classical conditioning that emphasize CS–US pairing do not explain all the ways humans and other species acquire and express learned fear [28]. For instance, fear can be communicated (or pre-empted) via verbal instructions or observation, circumventing the need for direct CS–US pairing altogether [29]. The nature of the CS also determines the strength and persistence of fear conditioning. Although animals can acquire conditioned responses to simple stimuli, like lights and tones, conditioning occurs faster and persists longer if the CS is from the class of fear-relevant stimuli [30]. Fear-relevant stimuli include prepotent threats, such as snakes, spiders, or emotional faces that signal danger or anger. A number of cognitive factors can affect the strength and persistence of fear conditioning as well [31], including cognitive reappraisal strategies, perceived controllability or predictability, and expectancy biases.

A controversial account of human associative learning has proposed that it is a reasoning process involving beliefs about the relationship between stimuli and events, rather than a purely low-level automatic process [32]. Much of the evidence supporting this view comes from human conditioning studies in which only subjects with awareness of the CS–US association exhibit a conditioned response [33]. These modifications of traditional forms of classical conditioning are in line with Rescorla's [28] remark that 'Pavlovian conditioning is not a stupid process by which the organism willy-nilly forms associations between any two stimuli that happen to co-occur. Rather, the organism is better seen as an information seeker using logical and perceptual relations among events, along with its own preconceptions, to form a sophisticated representation of its world' (p. 154).

representation of categories and concepts with classical conditioning.

An exception is category- or semantic-based fear conditioning in humans. In category-based conditioning, subjects acquire fear to a class of stimuli rather than to a specific CS. For example, subjects can learn through experience that members of a natural category (e.g., pictures of different animals) predict an aversive US, like an electric shock to the wrist, and will then express fear in response to novel category members that have never been paired with shock. Although examples of category conditioning are sparse compared to direct forms of fear conditioning, studies have shown that subjects exhibit category-specific anticipatory SCRs, ratings of shock expectancy, and category-selective increases in brain activity in visual cortex and fear-learning networks (e.g., the amygdala and insula) within the first few category conditioning trials [15,16]. Such findings show that humans can use category-level knowledge to associate a US to the entire category despite considerable variation in physical form between stimuli.

Category-based fear conditioning demonstrates that production of defensive responses can be mediated by the principles of categorization elegantly detailed in the 1970s by Eleanor Rosch and others [17]. In real-world situations, category-level fear generalizations explain why someone with a strong fear of dogs is also frightened by other types of animals or items associated with dogs (dog collars or veterinarians), or might avoid places associated with dogs (parks or hiking trails). Even though a particular park has never been entered, knowledge that dogs may run loose in parks could cause that park to be feared and avoided. Thus, networks of interconnected concepts and knowledge provide a route of fear generalization from a known threat to other stimuli connected to it.

Category-based induction and stimulus typicality

Within the concepts literature, the generalization of knowledge is often studied by category-based induction tasks in which properties are stated as being true of some premise categories, and then subjects judge whether they are also true of conclusion categories. For example, if salmon and trout have a mandibular reflex, do sharks also have a mandibular reflex? Many phenomena of this type of generalization have been identified [18]. One important finding is that the more typical the premise items are, the more likely their properties will be judged as true of other category members (other things equal). For example, drawing an inference from sparrows to other birds would be stronger than drawing an inference from loons or peacocks to other birds. Indeed, induction is stronger from sparrows to peacocks than from peacocks to sparrows [18].

A recent study examined whether a similar effect occurs in fear conditioning (see Figure 2) [19]. Subjects first learned to fear typical mammals (e.g., a horse, a rabbit, and a bear) and were then tested for generalization to atypical mammals (e.g., an armadillo, an aardvark, and an otter), or vice versa (conditioned to atypical exemplars and tested on typical exemplars). Learning was similar across groups, but generalization was asymmetrical: subjects generalized their conditioned response (as exhibited in SCRs) to atypical members after being trained on typical members but did not generalize their fear to typical members after being trained on atypical members. Thus, fear generalization was stronger when the 'premise' involved typical CSs. This finding suggests that conceptual structures and processes may be involved in fear generalization, rather than perceptual similarity alone, which was identical in the two groups.

Demonstrations that fear conditioning is sensitive to a relatively complex attribute like typicality show that conditioning with real-world objects goes beyond the physical properties of the CS and might extend to the concept underlying it. Consequently, the degree to which real-world fear experiences generalize may be determined by whether the details surrounding the event are regarded as specific to that event or instead activate a more widespread representation of associated stimuli or situations. For example, a near-fatal car accident on the way to work in fine weather (a typical event) could generate a widespread fear of driving in all conditions, whereas a car accident that occurred late at night while driving in a blizzard (an atypical event) could lead to a selective fear of driving in similarly hazardous conditions. The causal attribution of the accident may play a role in what aspects of the event lead to later fear.

If concepts are important to fear generalization, the question arises as to which concept is the most relevant. Natural objects are often in multiple categories, including those that form a taxonomy: for example, French poodle, poodle, dog, mammal, and animal. If you are bitten by a French poodle and become afraid of it, which categories will become part of that fear reaction: Only French poodles?

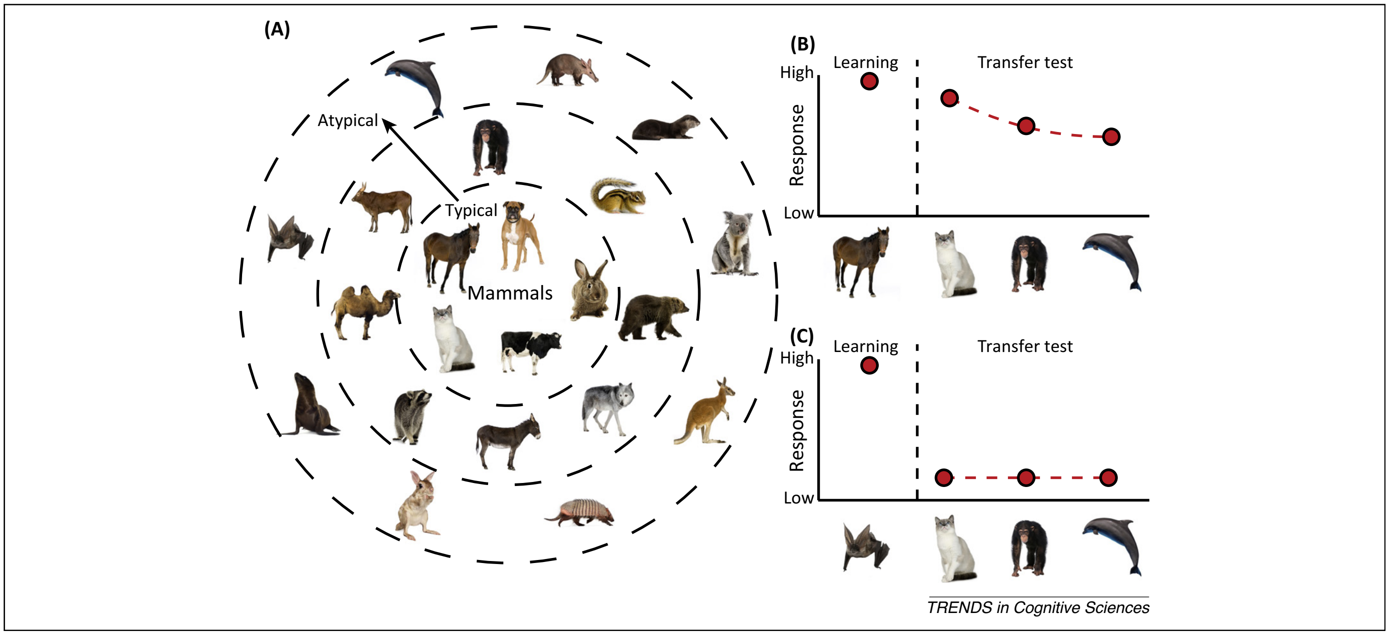


Figure 2. **(A)** Natural categories contain graded structures in which certain exemplars are considered more representative (or typical) of the category and other exemplars are less representative (or atypical) of the category. In this example, animals like dogs, cats, and horses are highly representative of the category domain 'mammals', whereas animals like bats, dolphins, and armadillos are not as readily identified as belonging to the same category. **(B)** Typicality can affect the generalization of learning in ways described in the category-induction literature [18,34], but that have not been fully appreciated in the domain of conditioned learning. For example, certain information that is learned about a typical exemplar (e.g., a horse) can lead to the conclusion that the same information applies to other members from the same category. **(C)** By contrast, information learned about an atypical exemplar (a bat, in this example) is often considered specific to that instance or others similar to it. This difference in generalization, depicted here between (B) and (C), is known as typicality asymmetry. Higher-order reasoning processes like typicality asymmetry may explain why some aversive experiences lead to widespread fear and avoidance of disparate stimuli or situations, while other aversive experiences do not generalize beyond the precise details of the experience.

All mammals? Research on concept learning and use suggests that a mid-level concept, the basic level concept – in this case, dogs – is the most accessible one [20]. Studies of word learning use a formally similar task in which a name is paired with a single object, and then children are tested on how they generalize the name to new examples. In general, children generalize names to the basic-level category (absent other manipulations) [21]. It is simply not known whether fear generalization would work in the same way. Although research has now shown that people can be conditioned to fear a broad (superordinate) category like animals or tools [15,16,19], we do not know whether fear will spontaneously generalize to such broad categories based on experience with a single exemplar. An intriguing possibility is that the degree of conceptual generalization following a fear learning experience is determined by the emotional intensity of the event, much the same way that footshock intensity determines how broadly conditioned fear is generalized in rodent studies [22]. This could explain, in part, why in disorders like PTSD such seemingly disparate cues or situations can involuntarily trigger re-experiencing of a trauma event [23].

There are also important theoretical questions that need to be addressed when considering generalization from natural stimuli. Shepard's [14] Universal Law of Generalization proposes that organisms generalize responses based on an exponentially decreasing function of similarity to the learning stimulus. He emphasized that this function was based on psychological similarity, and not just physical similarity. For example, notes an octave apart might be perceived as more similar than notes closer together in frequency. Perhaps the effect of concepts is primarily to help establish one form of psychological similarity, for example, all dogs are somewhat similar by virtue of being dogs, so a bad experience with a French poodle would tend to spread to other dogs.

However, we suspect that some effects cannot be explained simply through psychological similarity. Shepard's Law does not seem able to explain the typicality asymmetry in fear generalization we described above. Furthermore, being attacked by the French poodle may cause generalization to occur to stimuli that are associated with that object even though they are not similar to it (e.g., leashes or doghouses). Perhaps generalization involves reasoning about concepts' underlying structure, which has been shown to influence induction [24]. In that case, people's reasoning about what it is in the French poodle that caused the attack could influence generalization. For example, beliefs in a category essence [25] might indicate that only animals sharing that essence would be dangerous. Much human reasoning is done by analogy, in which properties are attributed to one object based on relational similarity to a possibly very different object [26]. That would also go beyond the similarity-based approach to generalization.

If we hope to explain and ameliorate disorders of fear and anxiety in humans, who have detailed knowledge about the causal structure of the world and who have a variety of ways of connecting disparate stimuli, the field will need to explore such intriguing possibilities. The extensive literature on human causal reasoning [27] provides a strong starting point for such an exploration.

Box 2. Questions for future research

- A clinically important issue is to discover techniques that facilitate the generalization of fear extinction. How can we use research on generalization of human knowledge to tailor more effective forms of safety learning that persist across environments and across stimuli that trigger fear and anxiety?
- If higher-order learning and inference and the organization of conceptual knowledge are involved in fear generalization, will we have to expand the neural circuits believed to be involved in fear learning?
- How well do mathematical models of learning and induction account for conceptually based fear generalizations, and do Bayesian approaches offer advantages over traditional similarity-based models of stimulus generalization when it comes to explaining human fear behaviors [35,36]?

Concluding remarks

There is a renewed interest in using traditional Pavlovian conditioning and stimulus generalization paradigms to examine fear generalization in humans. Much of the recent human fear generalization research has focused on its perceptual basis, in keeping with the historical approach developed over the last century in animal models of learning and generalization. However, higher-order cognitive processes such as inductive reasoning and conceptual knowledge are involved when humans generalize their experiences. We propose that fear generalization also involves such higher-level processes. Researchers now need to go beyond the perceptual generalization model to discover what role these processes play in fear generalization and to discover whether or how higher-order conceptual processes integrate with evolutionarily conserved systems that mediate conditioned fear learning and expression (Box 2).

Acknowledgments

We thank Brian H. Ross and Abby Fyer for helpful comments on this manuscript. The writing of this article was supported by National Institute of Mental Health Training Award in Systems and Integrative Neuroscience grant T32 MH019524 and by National Science Foundation grant BCS 1128769.

References

1 Ehlers, A. *et al.* (2004) Intrusive re-experiencing in post-traumatic stress disorder: phenomenology, theory, and therapy. *Memory* 12, 403–415
2 LeDoux, J.E. (2000) Emotion circuits in the brain. *Annu. Rev. Neurosci.* 23, 155–184
3 Pape, H.C. and Paré, D. (2010) Plastic synaptic networks of the amygdala for the acquisition, expression, and extinction of conditioned fear. *Physiol. Rev.* 90, 419–463
4 Davis, M. (1992) The role of the amygdala in fear and anxiety. *Annu. Rev. Neurosci.* 15, 353–375
5 Pavlov, I.P. (1927) *Conditioned Reflexes*, Oxford University Press
6 Guttman, N. and Kalish, H.I. (1956) Discriminability and stimulus-generalization. *J. Exp. Psychol.* 51, 79–88
7 Lissek, S. *et al.* (2008) Generalization of conditioned fear-potentiated startle in humans: experimental validation and clinical relevance. *Behav. Res. Ther.* 46, 678–687
8 Vervliet, B. *et al.* (2004) Generalization of extinguished skin conductance responding in human fear conditioning. *Learn. Mem.* 11, 555–558
9 Dunsmoor, J.E. *et al.* (2009) Generalization of conditioned fear along a dimension of increasing fear intensity. *Learn. Mem.* 16, 460–469
10 Dymond, S. *et al.* (2014) Fear generalization in humans: systematic review and implications for anxiety disorder research. *Behav. Ther.* Published online October 1, 2014. http://dx.doi.org/10.1016/j.beth.2014.10.001
11 Lissek, S. (2012) Toward an account of clinical anxiety predicated on basic, neurally mapped mechanisms of Pavlovian fear-learning: the case for conditioned overgeneralization. *Depress. Anxiety* 29, 257–263
12 Hull, C.L. (1943) *Principles of Behavior*, Appleton-Century-Crofts
13 Spence, K.W. (1937) The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.* 44, 430–444
14 Shepard, R.N. (1987) Toward a universal law of generalization for psychological science. *Science* 237, 1317–1323
15 Dunsmoor, J.E. *et al.* (2012) Role of conceptual knowledge in learning and retention of conditioned fear. *Biol. Psychol.* 89, 300–305
16 Dunsmoor, J.E. *et al.* (2014) Aversive learning modulates cortical representations of object categories. *Cereb. Cortex* 24, 2859–2872
17 Rosch, E. (1978) Principles of categorization. In *Cognition and Categorization* (Rosch, E. and Lloyd, B.B., eds), pp. 27–48, Erlbaum
18 Osherson, D.N. *et al.* (1990) Category-based induction. *Psychol. Rev.* 97, 185–200
19 Dunsmoor, J.E. and Murphy, G.L. (2014) Stimulus typicality determines how broadly fear is generalized. *Psychol. Sci.* 25, 1816–1821
20 Murphy, G.L. (2002) *The Big Book of Concepts*, MIT Press
21 Bloom, P. (2000) *How Children Learn the Meanings of Words*, MIT Press
22 Baldi, E. *et al.* (2004) Footshock intensity and generalization in contextual and auditory-cued fear conditioning in the rat. *Neurobiol. Learn. Mem.* 81, 162–166
23 Ehlers, A. and Clark, D.M. (2000) A cognitive model of posttraumatic stress disorder. *Behav. Res. Ther.* 38, 319–345
24 Rehder, B. (2009) Causal-based property generalization. *Cogn. Sci.* 33, 301–344
25 Gelman, S.A. (2003) *The Essential Child: Origins of Essentialism in Everyday Thought*, Oxford University Press
26 Gentner, D. *et al.* (2001) *The Analogical Mind: Perspectives from Cognitive Science*, MIT Press
27 Sloman, S. (2005) *Causal Models: How People Think About the World and its Alternatives*, Oxford University Press
28 Rescorla, R.A. (1988) Pavlovian conditioning – it's not what you think it is. *Am. Psychol.* 43, 151–160
29 Olsson, A. and Phelps, E.A. (2007) Social learning of fear. *Nat. Neurosci.* 10, 1095–1102
30 Öhman, A. and Mineka, S. (2001) Fears, phobias, and preparedness: toward an evolved module of fear and fear learning. *Psychol. Rev.* 108, 483–522
31 Hofmann, S.G. (2008) Cognitive processes during fear acquisition and extinction in animals and humans: implications for exposure therapy of anxiety disorders. *Clin. Psychol. Rev.* 28, 199–210
32 Mitchell, C.J. *et al.* (2009) The propositional nature of human associative learning. *Behav. Brain Sci.* 32, 183–246
33 Lovibond, P.F. and Shanks, D.R. (2002) The role of awareness in Pavlovian conditioning: empirical evidence and theoretical implications. *J. Exp. Psychol. Anim. Behav. Process.* 28, 3–26
34 Rips, L.J. (1975) Inductive judgments about natural categories. *J. Verbal Learn. Verbal Behav.* 14, 665–681
35 Tenenbaum, J.B. *et al.* (2006) Theory-based Bayesian models of inductive learning and reasoning. *Trends Cogn. Sci.* 10, 309–318
36 Tenenbaum, J.B. and Griffiths, T.L. (2001) Generalization, similarity, and Bayesian inference. *Behav. Brain Sci.* 24, 629–640