Journal of Experimental Psychology: Animal Learning and Cognition
2017, Vol. 43, No. 3, 219–230

2329-8456/17/$12.00 http://dx.doi.org/10.1037/xan0000139

Do Salient Features Overshadow Learning of Other Features in Category Learning?

Gregory L. Murphy
New York University

Joseph E. Dunsmoor
The University of Texas at Austin

Hundreds of associative learning experiments have examined how animals learn to predict an aversive outcome, such as a shock, loud sound, or puff of air in the eye. In this study, we reversed this pattern and examined the role of an aversive stimulus, shock, as a feature of a complex stimulus composed of several features, rather than as an outcome. In particular, we used a category learning paradigm in which multiple features predicted category membership and asked whether a salient, aversive feature would reduce learning of other category features through cue competition. Three experiments compared a condition in which 1 category had among its 6 features a painful "sting" (shock) and the other category a distinctive sound (the critical features) to a control condition in which the sting and sound were represented by much less salient (and not aversive) visual depictions. Subjects learned the categories and then were tested on their knowledge of all 6 features as predictors of the category label. Surprisingly, the experiments consistently found that the salient, aversive critical features did not reduce learning of other features relative to the control. Bayesian statistics gave positive evidence for this null result. Equally surprisingly, in a fourth experiment, a nonaversive salient feature (brightly colored patterns) *increased* learning of other features compared to the control. We explain the results in terms of attentional strategies that may apply in a category learning context.

Keywords: category learning, blocking, overshadowing, cue competition, aversive conditioning

Theories of category learning have often incorporated aspects of associative learning theory from research on classical conditioning. In the typical category learning experiment, the subject views a stimulus (e.g., a picture or verbal list of features) and then chooses a category name. Feedback is given. Although learning is reinforcement based, there is a clear parallel to the classical conditioning paradigm in which a neutral cue (the *conditioned stimulus, CS*) is paired with another stimulus or outcome (the *unconditioned stimulus, US*). An animal will soon learn this pairing and come to expect the US when the CS is presented, as shown by anticipatory behaviors such as salivation, freezing, blinking, and so forth. By the same token, after sufficient learning, a subject in the category-learning study will learn that upon seeing a stimulus item that it is in one or another category. In real life, the classification may be followed by appropriate behavior, such as approaching the ice cream truck or avoiding the pit bull.

To adapt classical conditioning theory to the category learning paradigm, it is assumed that the properties of the stimulus are analogous to the CS, and the category name is analogous to the US. That is, the properties of the object are cues that predict category membership as an outcome. Because category exemplars are variable, a given feature is not present in every category exemplar (e.g., not all birds fly, not all chairs have arms), and some features are more predictive than others (Rosch & Mervis, 1975). Because associative learning is often competitive, this can result in uneven learning in which some features are learned as being predictive of category membership, and others are not. Indeed, learning some properties could inhibit learning of others. We review this literature below.

The present study further explores this analogy by considering the issue of aversive conditioning. In the aversive conditioning paradigm, a noxious stimulus like a shock is used as the US. When preceded by a CS, the animal will come to expect the noxious stimulus as shown by freezing or attempting to avoid it. However, from the perspective of categorization, noxious stimuli are often properties of a category rather than a separate outcome. That is, the pain from an injection is not a category itself—rather, the pain is part of the event of receiving an injection. The sting of a yellow jacket is not itself a category—the sting is one of the yellow jacket's behaviors, like flying and crawling into soda cans. When people learn about injections or yellow jackets, they learn these aversive properties as part of their category knowledge. If one were asked, "Tell me about yellow jackets," one would hardly omit their stinging any more than one would omit their yellow and black stripes.

Of course, in real life, stimuli do not always follow the fixed order found in a psychology experiment, either in conditioning or category learning. One might taste an ingredient in a stew before

This article was published Online First May 4, 2017.

Gregory L. Murphy, Department of Psychology, New York University; Joseph E. Dunsmoor, Department of Psychiatry, Dell Medical School, The University of Texas at Austin.

Partial support for this research was provided by NIMH Grant K99MH106719 to JED. Thanks to Todd Gureckis and the Concats Research Group for helpful comments on this work. Tianqi Chen provided valuable help in running the experiments.

Correspondence concerning this article should be addressed to Gregory L. Murphy, Department of Psychology, New York University, 6 Washington Place, 8th floor, New York, NY 10003. E-mail: gregory.murphy@nyu.edu

seeing it, rather than seeing it first and then ingesting it, as in Pavlov's original experiment. One might first be told, "There's a yellow jacket on your arm" and then see and be stung by it—in this case category membership preceding exposure to its properties. We focus on the paradigmatic case in which an object's features are experienced prior to category membership. As a result, it is important to remember that the aversive stimulus is a cue (CS) in our paradigm, predicting classification, unlike in aversive conditioning in which the noxious stimulus is the US, predicted by a neutral stimulus.

The goal of the present research was to investigate the role of aversive features in category learning and in particular the situation in which the aversive stimulus is a cue to category membership. The conditioning literature has found that typical unconditioned stimuli can serve as predictors of other stimuli. For example, Goddard and Jenkins (1988) showed that if food appears at a regular schedule, one feeding event is learned to predict a subsequent event. Furthermore, this learning blocked the learning of another cue's prediction of feeding. However, so far as we know, no comparable situation has ever been tested in category learning. For example, what is the effect is of the painful sting on learning about yellow jackets as a category? Imagine that yellow jackets did not sting but merely took a stroll on your arm, tickling you. How would this affect the learning of yellow jackets' other properties? What would be the difference in learning about a stinging yellow jacket versus a harmless one?

Two opposing answers immediately come to mind. The first is that the sting could dominate learning. Because this is such a salient aspect of yellow jackets, it could draw virtually all one's attention, making it harder to learn other aspects of yellow jackets that are less attended. The second answer is that the shock would raise the overall attention to yellow jackets and increase learning of the category's features. After a few stings, the yellow and black pattern of the insect would be highly salient. Furthermore, one would pay attention to the fact that yellow jackets often nest in the ground, even if one's interest in the nesting habits of insects in general were minimal. People might become quite good at distinguishing yellow jackets from other similar insects by learning their precise appearance and behavior.

Both of these answers have a degree of plausibility. The second one seems more functionally beneficial, as it is obviously useful to learn the other properties of a noxious category so that it can be avoided, or at least predicted prior to delivering its painful stimulus.[1] If people cannot identify monarch butterflies or bumblebees from their appearance alone, there is little harm done, because they are unlikely to make pests of themselves. Thus, it would be wise to learn other properties of categories that have noxious elements, like wasps, skunks, and poisonous snakes, as well as events or kinds of people with negative aspects. One can certainly make an evolutionary argument that animals should be particularly good at learning about these categories.

In Pavlovian conditioning, meaningful and surprising events increase attention to cues associated with the event, provoking new learning (e.g., Pearce & Hall, 1980). Pavlovian learning can then spread to cues indirectly associated with the event through the process of stimulus generalization, acquired equivalence, or sensory preconditioning. Thus, Pavlovian learning could link a number of seemingly neutral properties of yellow jackets with their sting, presumably increasing their association to the category of yellow jackets. Episodic memory is also generally better for emotional events than unemotional events (e.g., Joëls, Pu, Wiegert, Oitzl, & Krugers, 2006; LaBar & Cabeza, 2006), including for neutral stimuli conceptually related to fear-conditioned stimuli (Dunsmoor, Murty, Davachi, & Phelps, 2015).

However, the first possibility—that a salient feature would dominate learning at the expense of other features—seems more in keeping with many strands of learning theory (though see below for caveats). Since the ground-breaking work of Kamin (1969), it has been observed that a strong cue that predicts an outcome can reduce learning of other cues, sometimes to zero. For example, in the *overshadowing* paradigm, two CSs are presented as predictors of a US, but one cue is much more salient than the other (e.g., a loud tone and dim light). Typically, animals will respond only to the salient predictor, even though the less salient one statistically predicted the US just as well. (If the salient predictor is absent, animals do respond to the less salient predictor.) Similarly, in the *blocking* paradigm, one cue is initially trained to predict a US. Then when it is presented in conjunction with a new cue, both now predicting the US, the second cue is not learned. That is, when one cue has been learned to predict an outcome, additional redundant cues are "not necessary" to predict the outcome and are not learned as such. This is usually explained by either error-driven learning mechanisms (Rescorla & Wagner, 1972) or an attentional mechanism that reduces attention to cues after learning is complete (Pearce & Hall, 1980). Analogous results are found in other paradigms, such as nonmetric cue probability learning (see Kruschke & Johansen, 1999), in which learning one strong cue reduces learning of other cues. Note that these results fly in the face of the intuition that an animal should want to learn all it can about when a shock/bite/threat might occur and so would learn both cues. Of course, not every yellow jacket stings you, and other features could be learned on those occasions. The real world is not an exact parallel to a typical conditioning experiment.

Theories of category learning have drawn on the competitive nature of cue learning as embodied in these theories to various degrees. For example, Gluck and Bower (1988) used the Rescorla-Wagner learning rule to explain a surprising phenomenon in which category frequency did not influence feature learning as it should have. Kruschke (2001), drawing on Mackintosh's (1975) theory of conditioning, proposed a number of sophisticated models of category learning that used error-driven attentional shifts to explain blocking-like effects.

In many category-learning models, the amount of attention weight across stimulus dimensions is fixed (e.g., so that weights across dimensions must sum to 1.0; Kruschke, 1992; Nosofsky, 1984). This effectively creates a tradeoff, in which allocating attention to one dimension requires reducing it to others, again limiting the amount that can be learned about a category's features. Hoffman and Murphy (2006) discussed the implications of these assumptions for category learning. In general, they imply that as

[1] Although properties are used to classify the object, once its category membership is known, other properties can be inferred. If you can identify a yellow jacket by its appearance, you can predict that it could sting you. Thus, in real life, features act as both predictive and predicted stimuli; in the classic category-learning task used here, they are restricted to a predictive role (equivalent to a CS).

one or a few dimensions are learned, this impedes learning of other dimensions.

Discussions of how associative learning principles might help explain category learning seem to assume that cue competition effects like blocking are the rule. However, within the conditioning literature, there is considerable variability as to when those effects appear. One recent article reported 15 failed attempts to observe blocking in mice and rats (Maes et al., 2016). Overshadowing can occur, but it may depend on specific variables such as the time between the CS and US. When there is a long gap, there may even be a reverse effect, in which a weak cue improves when it is paired with another cue (Urcelay & Miller, 2009). Furthermore, there is some question as to whether the observed competition effects are always due to failure to form associations or instead reflect some kind of performance suppression (Arcediano, Escobar, & Miller, 2004). For example, the cue which was apparently not learned may then show spontaneous recovery after the overshadowing or blocking cue is extinguished (Kaufman & Bolles, 1981; Matzel, Schachtman, & Miller, 1985). We do not attempt to review this literature here. However, the fact that competitive effects in learning do not always occur makes it more important to ascertain empirically whether they occur in the category learning task, as is assumed in most discussions in the categorization literature.

Indeed, competition effects are not always found in category learning. Hoffman and Murphy (2006) found that people learned more properties of categories than they should have given cue competition. Adding more stimulus dimensions to a category resulted in more dimensions being learned, even though they did not improve category predictability (reduce error). When stimulus dimensions can be integrated into prior knowledge, that extra learning is even greater (Hoffman, Harris, & Murphy, 2008). Furthermore, phenomena such as blocking do not appear to occur when people are instructed to learn categories, as opposed to learning to predict an outcome such as high or low tone (Bott, Hoffman, & Murphy, 2007). On the other hand, cue competition effects in nonmetric cue learning seem well established (as reviewed by Kruschke & Johansen, 1999). Aversive stimuli are inherently attention-grabbing and so might well be expected to reduce learning of other category features, but given the inconsistent results of competition effects in category learning, an empirical test seems called for.

One potential reason that competition effects may not occur in category learning is that human subjects may strategically allocate more attention to category learning than they do to other learning tasks. If you think your job is to learn about swallows, for example, you might not stop learning when you have identified their distinctive tail, because you take as your goal to learn whatever you can about swallows (within reason). Thus, even when you can classify swallows with good accuracy based on their tails, you might continue to attend to other properties such as their coloration, swooping flight paths, feeding habits, and nesting sites. This seems inconsistent with the assumption of fixed attention weights in category-learning models, but such weights are *decision weights* rather than attention in a broader sense (see discussion in Hoffman & Murphy, 2006). Clearly, people can allocate more or fewer resources to a task and can to some degree allocate resources to particular stimuli that they are interested in (Kahneman, 1973). Blair, Watson, and Meier (2009) found that attentional allocation to stimulus dimensions continues to improve after people have stopped making errors and even in the absence of feedback. They explain this as an effect of an executive decision to allocate attention to features in order to continue to learn.

In this light, it may be that being stung by a yellow jacket could increase learning as a whole, by causing the unfortunate learner both to attend more carefully to yellow jackets and to continue to learn their properties even after classification is accurate.

So far as we know, no study has investigated whether category learning is affected by the noxiousness of one of its stimulus dimensions. There is a general recognition that more salient dimensions have greater weight in classification decisions (Medin & Schaffer, 1978), but we do not know of any study that compares category learning when a highly salient dimension is present or absent. Furthermore, there is a possibility that a painful stimulus will evoke acute stress or emotional arousal, with complex effects on learning. Classically, such effects have been argued to result in a narrowing of what is learned (Easterbrook, 1959), though the effects of stress and arousal on learning are quite complex (e.g., Joëls et al., 2006; Rodrigues, LeDoux, & Sapolsky, 2009). We postpone discussion of these possibilities until after discovering whether there are indeed any effects of a noxious stimulus. Although our main interest is in the effect of noxious features like stings or other painful stimuli, we also consider the case of a highly salient feature that is not at all noxious, in Experiment 3.

Experiment 1

We used the classic category-learning task with family resemblance categories (Rosch & Mervis, 1975) to investigate the effect of an aversive stimulus dimension. Although we report learning data, the main dependent measure was how many features people learned, in order to evaluate whether the noxious dimension aided or interfered with acquisition of other category-relevant knowledge. (In this context, *learning a feature* means learning which category the feature predicts.) All the stimulus dimensions were statistically equivalent, in that a given value was associated with one category most of the time but occurred in another category in one stimulus, using the *one-away* design popular in category-learning studies (see Table 1). Thus, each feature is probabilistically associated with one category five-sixths of the time.

Figure 1 shows the prototypes of the two categories. They were hand-drawn imaginary animals, with stimulus dimensions of head shape, ears/antlers, back hair/fin, number of legs, and tail type. In addition to these, there was a dimension described as a behavior that the animal performs when threatened: making a sound, indicated by a balloon saying "Brrupt!," or a sting, indicated by a lightning bolt next to the animal's head. We call these the *critical features* for each category. During learning, the visual stimuli and their categorizations were identical for the two groups. The only difference was that the *shock group* heard a sound over headphones and received a shock to the wrist when the sound and sting were presented. For the *no-shock group,* the critical features were presented only as the lightning bolt or cartoon balloon. (Although we call the groups *shock* and *no-shock* for simplicity, note that the shock group also heard the sound, and the no-shock group was also "no-sound.") We did not ask this group to wear headphones or be attached to the electrophysiological equipment, as this might make some of them worry that they would in fact be shocked or hear something, which could have a similar effect as actually being

Table 1
Categorical Structure

Category	Defense	Ears	Back	Feet	Head	Tail
Dax	**1**	1	1	1	1	0
	1	1	1	1	0	1
	1	1	1	0	1	1
	1	1	0	1	1	1
	1	0	1	1	1	1
	0	1	1	1	1	1
Kez	**0**	0	0	0	0	1
	0	0	0	0	1	0
	0	0	0	1	0	0
	0	0	1	0	0	0
	0	1	0	0	0	0
	1	0	0	0	0	0

Note. Each row represents a category exemplar. The defensive action dimension (shown in bold) was the critical dimension of shock/sound. In Experiment 3, that dimension was replaced by texture. In Experiment 4, the "exception features" were eliminated in the defensive action.

shocked, or perhaps create a stress reaction, due to uncertainty. The sound was loud enough to be very salient, but not painful. We decided against using two noxious stimuli (e.g., a truly painful loud sound in addition to the shock), both out of concern for the subjects and because the most relevant real-life situation seems to be in distinguishing dangerous from nondangerous things, rather than predicting whether you will be stung or bitten.

All subjects underwent the same amount of training. Although it is common in category-learning studies to continue the learning phase until a given criterion is reached (sometimes for hundreds of trials), doing so can make the results of later tests ambiguous (Murphy & Allopenna, 1994). For example, if the shock group took four blocks to reach criterion and the no-shock group took six blocks, better learning of the categories' features in the latter group could be explained via the difference in learning trials rather than cue competition.

The test phase contained animals with only one feature, and subjects were asked to classify the animal based on that feature. The main dependent measure was accuracy in this test phase, which measured knowledge of feature-category associations. It seems obvious that the shock and sound will be better learned by the shock group than by a group that only saw visual representations of these things; that difference is not of great interest. The important question is whether the presence of the shock and salient sound reduces or increases learning of the other stimulus dimensions. If the parallel with overshadowing is correct, then the shock will reduce learning of other features; if the notion that attention as a whole will increase with a noxious stimulus, then learning should be actually greater in that case.

Method

Participants. There was no past literature to draw on to estimate the size of any effect of a shock in category learning. Past studies in our lab testing learning of individual features have used as few as 16 subjects per group (e.g., Hoffman et al., 2008). However, there is a further consideration in the present case, namely that we were reluctant to expose more people than necessary to repeated unpleasant shocks. Therefore, we decided to fix the number of subjects at 20 per group in each experiment. Subjects were members of the NYU community who were paid for their participation. They were randomly assigned in equal numbers to conditions. The procedures were approved by the New York University Institutional Review Board.

Materials. The stimuli were pictures of fictional animals described as members of the categories Dax and Kez. As shown in Figure 1, the animals differed in six dimensions: head, ears/antlers, back hair/hump, tail, feet, and defensive action. The last was represented as a yellow lightning bolt to indicate a sting or a cartoon bubble representing a sound. Each category contained six objects in the one-away design shown in Table 1. In this way, every feature (including the critical ones) was primarily associated to only one category but was neither necessary nor sufficient for classification.

Test items were constructed by pairing one of the visual features with the oval body of the animal. The body provided a scaffolding on which legs, tails, and so forth could be added. Since that shape was present in all stimuli, it provided no information about category membership. This resulted in 12 different test items.

Procedure. Subjects were solicited through an advertisement indicating that they would receive shocks as part of the experiment. All participants gave informed consent to the procedure. The no-shock subjects learned that they would not actually receive shocks only after the experiment had begun. Those in the shock condition were connected to a Grass Technologies (Warwick, RI) stimulator via electrodes connected to the left wrist, as in past research (Dunsmoor, Martin, & LaBar, 2012; Dunsmoor & Murphy, 2014). A calibration procedure was performed for each subject to identify a level that was considered "highly annoying but

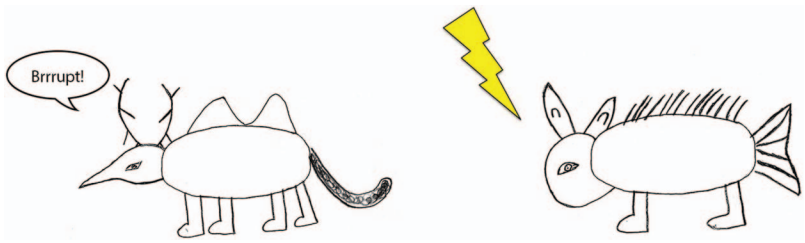

Figure 1. Prototypes of the two categories used in Experiments 1, 2, and 4. The lightning bolt indicating shock was colored yellow. Actual stimuli usually differed by the exchange of a feature (e.g., the animal on the left would appear with the tail of the animal on the right), as shown in Table 1. See the online article for the color version of this figure.

not painful." The shocks were 200 ms long. The animal sound (actually a tree frog croak) was delivered over headphones.

Subjects were told that they were going to learn about two kinds of animals, Daxes and Kezes, by seeing examples of each kind. They would guess which kind of animal each one was and receive feedback. Over time, they should become able to identify the two types of animals. The instructions mentioned all six stimulus dimensions, including the visual representations of the sting and sound, described as defensive behaviors. Those in the shock group were further told that they would feel the sting and hear the sound.

On each trial, the picture was shown at the beginning for 1250 ms during which no response was allowed. The sound or shock, if present, occurred at the same time as initial picture presentation. Underneath the picture appeared the instructions, "Press 1 if it is a Dax. Press 2 if it is a Kez." Subjects responded by pressing the 1 or 2 keys on the numeric keypad with their right hand (as the shock group's left wrist had the electrodes on it). After response, a 3,000-ms feedback message indicated whether the response was correct or incorrect as well as which category the animal in the picture belonged to. The picture remained on screen during feedback. The intertrial interval was 1,000 ms. Each block consisted of 12 randomly ordered trials, and subjects were permitted to rest in between blocks. Training ended after four blocks, which we predicted (correctly) would result in learning of some noncritical features while avoiding a ceiling effect. Thus, we would be able to detect either an increase or decrease in learning as a function of shock and sound.

In the test phase, subjects read that they would be seeing only one feature at a time but should choose the category that the feature occurred with most often. They were told that accuracy was most important but that they should make their response as soon as they had made up their mind and that their RTs were being recorded. Each feature appeared twice, except for critical features which appeared four times (given that there were only two of them). The actual shock and sound were not presented during testing, in order to make the test stimuli equivalent for the two groups. (The experimenter informed participants that there would be no shock or sound before the test phase.) Although this changed the precise presentation of the critical dimensions for the shock group, the experiment's main question was how learning of the *non*critical features would differ based on the presence of the shock and sound. We did not want those trials to be influenced by anticipation or aftereffects of an aversive stimulus.

Results

The two groups learned the categories about equally well. In the final learning block, the shock and no-shock groups had mean accuracies of .76 and .77, respectively. However, this does not entail that the two groups learned the same amount, because performance on whole items does not tell us how many properties of those items subjects learned. For that, we need to examine the test phase, where individual features were classified.

A two-way ANOVA of classification of the critical versus noncritical features for the two groups did not reveal any significant effects (main effect of feature type: $F(1, 38) = .76$, partial eta-squared = .02; main effect of shock: $F(1, 38) = .01$, partial eta-squared = 0; or interaction: $F(1, 38) = 1.08$, partial eta-squared = .03). Five subjects had below-chance accuracy in the final block of learning. If they are eliminated from the test analyses, the results are virtually unchanged, other than a slight increase in accuracy.

At test, the shock group unsurprisingly was numerically more accurate than the no-shock group in the critical features (.73 vs. .68 correct, respectively), as shown in Table 2. However, the difference was small, and the overall level of accuracy was surprisingly low (chance = .50). The critical question in the experiment was whether there was any difference in learning of the noncritical features. There was not, with the shock group having a proportion correct of .73 ($SD = .16$) and the no-shock group of .77 ($SD = .16$), $t(38) = .87$, $p < .40$, $d = .28$.

Given that accuracy was not very high, reaction times (RTs) are not of great interest. (With accuracy rates of noncritical features at 75%, there was much missing data, and a significant proportion of correct responses were guesses.) As a result, we do not formally report RT results. However, we examined the results for every experiment to ascertain that there was no speed–accuracy trade-off that would undermine the accuracy results, and none appeared.

Discussion

Overall, we did not find evidence of a difference in learning of the bulk of the categories' features depending on whether the critical stimulus involved an actual sound and shock or not. Although both stimuli should be extremely attention-grabbing and one of them aversive, this did not reduce the amount learned about the other properties of the category. But by the same token, it did not seem to generate additional attentional resources to improve learning of the other features. (Given that the results are null, we performed a Bayes Factor analysis, described after Experiment 2.)

One surprising result of the experiment is that the shock group did not learn the critical features very well. Surely if you were being shocked by items of a given category you would learn this pairing and pretty quickly. Overall accuracy was only.73 for these

Table 2
Mean Accuracy Rates (and SDs) for Classifying Features in the Test Phase, Experiments 1–4

	Group	
	Shock	No Shock
Experiment 1		
Critical	.73 (.29)	.68 (.33)
Noncritical	.73 (.16)	.77 (.16)
	Shock	No Shock
Experiment 2		
Critical	.85 (.24)	.81 (.33)
Noncritical	.66 (.15)	.70 (.18)
	Color	B&W
Experiment 3		
Critical	.98 (.06)	.80 (.28)
Noncritical	.82 (.15)	.69 (.16)
	Shock	No Shock
Experiment 4		
Critical	.98 (.11)	.95 (.13)
Non-critical	.70 (.18)	.70 (.18)

features (shock and sound). Of course, the critical dimension was probabilistically associated to the category, as all the stimulus dimensions were. In each block, any given feature was associated five times to one category and once to the other category. That seems a pretty strong association that should allow for confident learning (across four blocks) of the most salient dimension, but that did not always happen. It is possible that the removal of the actual shock and sound at test reduced the classification accuracy for the shock group, although it would be surprising if they had learned that an actual shock predicted that an animal was a Kez but could not retrieve this from the shock symbol presented along with it.

One possible explanation for this relatively low accuracy is probability matching. Subjects could be responding with the correct category roughly the same proportion as each feature occurred with that category (.83 of the time). This is possible, but it should be noted that the task is not reinforcement learning or prediction, but rather a question with a single correct answer: Which category did the feature appear with most often? No feedback was given, so the motivation to change answers in order to predict the infrequent outcome is not present. We will argue below that later results make this interpretation of the accuracy data unlikely.

This quirk in the results leads to a theoretical concern. According to a competition-based analysis of learning, it is learning one predictive pairing that leads to reduced learning of other features. But if that pairing is not actually learned, it would not necessarily cause cue competition. Thus, although it seems clear that the prediction of a noxious stimulus harming learning of other dimensions did not come to pass, this may not be because of any failure of error-driven learning but rather to the failure of learning that noxious property. (However, the failure to learn about the noxious stimulus becomes a notable result in its own right.)

We attempted to address this theoretical concern in Experiment 2 by boosting the learning of the critical dimension so that it could provide cue competition. We did this by a modified version of the blocking paradigm (Kamin, 1969), in which people were first trained on a block of items containing only the critical features, shock and sound (or just their visual counterparts). Like the usual learning blocks, a given stimulus was associated with one category five times and the other category once. This should permit easy learning of this dimension that could then help or interfere with learning of the other stimulus dimensions. Then the usual learning trials and test followed.

Experiment 2

Method

This experiment was essentially identical to the previous one, with the exception that an initial block of pretraining contained only the sound or shock property (with the oval body and visual depicture of each, as in the test items). When that was completed, subjects performed four blocks of learning, as in Experiment 1. The test was also the same. Forty new subjects from the same population served in this experiment, half receiving actual sounds and shocks and the other half only the visual representations.

Results

The learning results were about the same as those found in Experiment 1. The shock group had a mean accuracy of .77 in the final block, whereas the no-shock group had a mean of .80. These were not significantly different, $t(38) = .69$, $p = .50$, $d = .28$.

The test results were improved on those of Experiment 1 in the sense that accuracy was now higher for the critical features, as shown in Table 2. A two-way ANOVA showed that learning was reliably higher for critical than noncritical features, $F(1, 38) = 11.08$, $p < .01$, partial eta-squared = .23, about a .15 difference in proportion correct. Accuracy was not at ceiling, but it was strong enough that it might influence learning of the other dimensions. However, the ANOVA revealed no other significant effects: shock versus no shock, $F(1, 38) = .004$, partial eta-squared $< .01$; interaction, $F(1, 38) = .55$, partial eta-squared = .01. In particular, there was no difference between the learning of noncritical features in the two groups: .665 versus 695 for the shock and no-shock groups, $t(38) = .57$, $p > .50$, $d = .18$.

Discussion

The rather surprising null effects of Experiment 1 were repeated in Experiment 2. Although there was now a clear advantage for the critical stimulus dimension, presumably due to the initial block of training, there were no differences between the shock and no-shock group. Because the effects were null, we computed the Bayes Factor (*BF*) for the noncritical feature accuracy data of both Experiments 1 and 2 (Rouder, Speckman, Sun, Morey, & Iverson, 2009, describe how to calculate BFs from *t* values). In both cases, the BF found greater evidence for the null hypothesis: 5.6 for Experiment 1 and 6.9 in Experiment 2. (A ratio of 1 would indicate equal evidence for the null and alternative hypotheses; the observed BFs indicate that the odds are about 6 to 1 that the null is correct.) These BFs are considered "substantial" evidence for the null hypothesis, allowing some confidence in believing that the shock had little effect on learning other properties. If the test data of the noncritical dimensions of the two experiments are combined in a 2-way ANOVA, with experiment and group as factors, there is no effect of shock versus no-shock, $F(1, 76) = 1.03$, partial eta-squared = .01, and the Bayes Factor is again in favor of the null, BF = 6.8.

It is surprising that actually receiving a shock and hearing a loud sound does not draw more attention to the stimulus dimension. The authors can confirm that being shocked on half the trials is not the same experience as viewing a picture of a lightning bolt. One of them attracts your attention much more than the other does. So, if the shock and sound are causing the subjects to pay greater attention to that dimension, this does not seem to detract from learning the other properties of the category.

One possibility is that both of the proposed mechanisms are at work. Perhaps the shock and sound are in fact "stealing" attention from the other dimensions, but at the same time are increasing the pool of resources devoted to the task. The result may be that the two effects roughly balance one another. That explanation can be tested by using a salient critical dimension that does not involve any sort of noxiousness. Such a dimension would likely be learned first, and more strongly, but would not serve to increase the total amount of attention allocated to the task. In that case, we should

see interference in learning the other dimensions compared to a control that lacks the salient dimension. Experiment 3 uses such a design.

Experiment 3

In Experiment 3, we removed the shock and sound and their visual counterparts from the stimuli. We replaced them with texture patterns that occurred in the middle of the animal body—stripes or dots. In the *color* condition, these patterns filled up the entire body and were bright red or green. In the context of otherwise monochromatic stimuli, this dimension was by far the most salient and would be expected to attract attention. In the *black and white* (b&w) condition, the patterns were smaller, filling only an invisible circle in the middle of the body, and were monochromatic, like the rest of the figure. The stripes and dots were also slightly smaller than in the other condition, so that they would fit into this reduced space. Informally, the impression was that this stimulus dimension was no more salient than that of the antlers or back element, for example.

Obviously, color patterns are not noxious, nor should they evoke any anxiety or emotional reaction. However, these patterns should draw attention and therefore be quickly learned. If the analogy to overshadowing and blocking is correct in the context of category learning, then this should lead to worse learning of the other dimensions. We decided to "hit subjects over the head" with the manipulation by adding a pretraining block for the color stimuli but not the black-and-white stimuli. This should guarantee that the colored texture is learned prior to the other dimensions, thereby creating the situation in which it could interfere with learning other stimuli.

Kruschke and Johansen (1999) reported extensive tests of varying stimulus dimensional salience on nonmetric cue probability learning, which is similar to a category-learning task, although different in some important respects (which we discuss in the General Discussion). They consistently found that people used salient dimensions more than less salient ones as cues for classification. However, they apparently did not test the precise situation that would be most relevant to the present case, in which both dimensions actually predict classification, and the salience of one dimension is varied while keeping the other constant. The closest example is their Experiment 4, in which one dimension was predictive, and the other dimension was irrelevant. They varied the salience of the irrelevant dimension and discovered that the utilization of the relevant dimension decreased when the irrelevant one was salient. This would suggest that in our task, now without any emotional effects of shock, we should see less learning of the rest of the dimensions in the color condition.[2]

Kruschke and Johansen (1999) noted that an opposite effect has also been found, as Busemeyer, Myung, and McDaniel (1993) reported unpublished research in which they found a *cue cooperation effect* rather than the expected cue competition when instructions did not inform subjects that one cue was more effective than another. It is unclear why the basic learning processes that presumably underlie this task should be sensitive to such high-level information. Kruschke and Johansen never found cue cooperation, however, and category-learning models generally predict competition. Given the important differences between the cue-learning experiments—which have only two dimensions and sometimes only four stimuli in the entire experiment—and family resemblance category learning, it seems important to test stimulus salience with a standard category structure and learning procedure.

Method

The overall method was very similar to that of Experiments 1 and 2. The primary difference was the change in the stimuli described above: removal of the shock and sound, along with their visual representations in the stimuli. Instead, the stimuli had a large, brightly colored pattern—dots or stripes—taking up much of their body (color condition) or else a smaller, black and white pattern (b&w condition). The pattern is therefore considered the critical stimulus dimension in this design. This resulted in a stimulus structure identical to that of the previous experiments (as in Table 1, but with pattern replacing the defensive action). The color condition had a pretraining block, like that of Experiment 2, in which the texture dimension (on the body) was presented alone. That was followed by the usual training phase. The b&w condition had only the learning phase (no pretraining), as in Experiment 1. Thus, the two groups had identical experience with the noncritical dimensions, namely, the four learning blocks.

Forty subjects were again randomly assigned to the two conditions; however, the computer crashed during the running of the final subject at the end of the semester, so the final tally was 20 subjects in the color group and 19 in the b&w group.

Results

During learning, there was a slight advantage for the color group, $M = .91$, compared to .88 for the b&w group in the final block. This was not reliable, $t(37) = 1.08$, $p > .25$, $d = .35$, though any differences might be masked by ceiling effects. Therefore, the test trials are again essential for measuring how well individual features were learned.

The manipulation of salience plus pretraining was clearly very successful as revealed in the test trials. Subjects were virtually perfect in classifying the critical feature (texture) when it was large and colored: .98 accuracy, as opposed to .80 in the b&w condition (see Table 2), $t(37) = 2.79$, $p < .01$, $d = .89$. Surprisingly, however, rather than causing reduced learning of the noncritical features, subjects were more accurate in the color condition: .82 versus .69. The two-way ANOVA confirmed that the critical features were learned better, $F(1, 37) = 13.98$, $p < .001$, partial eta-squared = .27, and that the color group learned more in general, $F(1, 37) = 11.78$, $p < .001$, partial eta-squared = .24, with no interaction, $F < 1$, partial eta-squared = .02. The difference between the two groups in learning the noncritical features in

[2] A related situation is found in their Experiment 3 in which one dimension was irrelevant and the other, predictive dimension varied in salience across groups. Although they did not statistically compare the effect of salience on the irrelevant dimension, an examination of their results (see Figure 15) suggests that subjects ignored the irrelevant dimension more when the relevant one was salient. Their model RASHNL seems to make a prediction in the same direction (also Figure 15). However, these effects, if they truly exist, are small, probably because the dimensions were statistically irrelevant and therefore tended toward zero effects regardless of the other dimension. Thus, we likely should not generalize these results to a family resemblance structure in which all dimensions are predictive.

particular was reliable, $t(37) = 2.46$, $p < .02$, $d = .79$. It should be emphasized that the two groups had identical experience with these noncritical dimensions.

Discussion

The experiment's results were most unexpected. First, the test results did not resemble the shock conditions of the earlier experiments, where there were no group differences. Here, people in the color condition learned more in general, in both the critical and noncritical features. Second, there was no sign of cue competition, in that the color group was virtually perfect in learning the critical features, but they did not suffer any decrement in learning the other features—quite the opposite. Thus, the results lead to two obvious questions. First, why was there no cue competition such that a highly learned dimension leads to less learning of the other dimensions? Indeed, how can learning of a salient dimension actually improve learning of the other dimensions? Second, why do noxious and non-noxious salient stimulus dimensions lead to different learning patterns? We postpone addressing these questions until after the final experiment.

One minor surprising result is that the critical features were more accurate than the noncritical features for the b&w group, even though they were not colored or pretrained. We believe that this reflects the fact that the texture dimension is in the center of the animal and therefore more likely to be fixated and encoded than some of the other features, which were more peripheral. That fact did not vitiate our manipulation, as the critical features were learned significantly better—essentially perfectly—in the color group, which learned both kinds of features significantly better.

Finally, the near-perfect accuracy of classifying the colored texture (.98) in this experiment argues against the interpretation of poor performance in previous experiments as reflecting probability matching. The probabilities of the textures' associations to categories were identical to those of the shock and sound to their categories in previous experiments, yet probability matching was not seen here.

Experiment 4

The difference between the results with the colored salient dimension and the shock-sound dimension is striking. Even in Experiment 2's blocking condition, which was similar to Experiment 3's procedure, we did not find any difference in learning of the noncritical features. Why should the shock and sound have no effect when colored patterns do have an effect? The reverse pattern of results would be easy to explain, but this one is very puzzling.

One possibility is that the probabilistic nature of the features (cues) is interacting with the noxious nature of the critical stimuli. Although people have mostly learned which category has the shock and which has the sound, the fact that the cues are to some degree inconsistent may cause problems for learning the noxious stimuli. That is, if the green dots were only associated with Kezes five out of six times, that is still fairly predictive. However, if the shock occurs five out of six times in Kezes, this may cause anxiety or at least impair learning, given that one of the Daxes is also going to shock you. The result may be that the association between shock/sound and their respective categories may be interfered with, thereby preventing the occurrence of the effect found with colored patterns. For example, rather than thinking "Kezes usually shock you, and Daxes usually don't," subjects may be conceptualizing the relationship as "Kezes shock you, and Daxes also do some of the time." As a result, the critical feature could be interfering with features in both categories.

One piece of evidence for this hypothesis is the relatively low level of learning of the critical dimensions we noted in Experiments 1 and 2. In the first experiment, performance was only at .73 accuracy, which seems remarkably poor performance. Learning of the critical dimensions improved to .85 in Experiment 2, but that also seems lower than expected, given that one entire block of learning was devoted to the critical stimuli, followed by four further blocks with whole exemplars. In contrast, when the color patterns were presented with a prior block of learning, performance was essentially perfect, .98. Thus, it may be that shock reduces the learning of imperfect relationships of the sort found in family resemblance structures.

To investigate this possibility, Experiment 4 essentially repeated Experiment 2, only changing one item in each category so that shocks and sounds were now perfectly associated with one or the other category. So, in Table 1, the final Dax had values 1 1 1 1 1 1 instead of 0 1 1 1 1 1 1. Similarly, the final Kez now had values 0 0 0 0 0 0. Except for these changes to two exemplars, the rest of the stimuli remained as shown in the table. This perfect predictability for shock and sound was also the case during the blocking manipulation: The initial block of learning perfectly correlated shock and sound, presented alone, with category membership. Perhaps with the imperfect association now removed, the noxious dimension will have the same effect as the colored patterns did in the previous experiment. The noncritical dimensions had the same structure and were presented exactly as often as in the previous experiments, so these changes did not alter their statistical relations to the category.

Method

The method was nearly identical to that of Experiment 2. Two groups received identical learning and test trials, except that one was actually shocked and heard sounds, whereas the other was not. For both groups, an initial block of training on the critical dimension preceded four blocks of learning on whole exemplars. The only change from Experiment 2 was in the category structure, as now shock and sound were perfectly associated to different categories in both the initial block and four learning blocks, as described above. The rest of the category structure was unchanged: Each noncritical feature occurred five times with one category and once with the other in each block. Forty new subjects completed the experiment.

Results

Category learning was very high in both groups, probably because each category now had a perfectly predictive feature. Accuracy in the final block was .98 for the no-shock group and .996 for the shock group. However, as we have seen, this does not mean that subjects learned all the features. Results of the test trials are reported in Table 2. Not surprisingly, the two groups were at ceiling with the critical features, which were trained in an initial block and were perfectly predictive. Thus, this is the classic

blocking stimulus, which in this case was perfectly learned by most subjects. However, there was no differential effect on the noncritical stimuli, as both groups averaged .70 in classification accuracy. There may have been a blocking effect in the learning of these stimuli, but the presence of an actual shock and sound did not alter it.

Although cross-experimental comparisons must be speculative, it is interesting that the overall level of noncritical feature accuracy was no worse in this experiment than in Experiment 2 (see Table 2), even though the critical features were (unsurprisingly) learned much better in the present experiment. Cue competition would suggest that performance on noncritical features should be worse in this experiment.

Because of the null finding, we repeated the Bayes Factor analysis we performed after Experiment 2, now with the full set of experiments that used the shock and sound stimuli (Experiments 1, 2, and 4). There was no effect of shock versus no-shock on classification of the noncritical stimuli, $F(1, 114) = .60$, $p = .44$, partial eta-squared $< .01$. This translates into a Bayes Factor of 3.91 in favor of the null hypothesis. The overall level of accuracy for noncritical stimuli across experiments were .70 and .72 for the shock condition and no-shock conditions. If there is an effect, it must be extremely small.

General Discussion

In none of the experiments did we find the sort of cue competition effects that one might have expected from a salient stimulus dimension that makes up part of a category. In Experiments 1, 2, and 4, there was virtually no effect of a shock or loud sound compared to the visual representation of the shock and sound. In Experiment 3, a salient and pretrained dimension did influence the learning of other features, but the effect was opposite to that of cue competition: The salient cue served to increase learning of other features of the category. We first address the latter effect, and then discuss the issue of noxious stimuli.

Cue Competition (or Absence Thereof)

In classic associative learning, once an outcome can be predicted, further learning about other cues may not occur or is weaker than when the outcome cannot yet be predicted (Pearce & Bouton, 2001). Assuming that a category name is an "outcome," and the features of its exemplars are cues, one would expect the same result in category learning. However, Bott et al. (2007) found that the traditional blocking manipulation of pretraining on a single dimension led to reduced learning only when subjects were told that they were predicting whether the computer would emit a high or low tone. When they were told that they were learning two named categories, there was no such reduction. Furthermore, Hoffman and Murphy (2006) discovered that people learned more features than were necessary to correctly categorize, seemingly contradicting the prediction that once learning is accurate, more cues will not be learned. Adding more dimensions to the stimuli resulted in more being learned, even though the additional dimensions didn't have to be learned to improve performance. Blair et al. (2009) found that learning continued after performance became perfect. Thus, the present results comport with past findings that cue competition or error-driven learning does not always characterize human category learning.

As we suggested in the Introduction, one reason for such results in category learning is that people may strategically allocate attention to learn as many features as possible, even when they are already responding accurately. That is, they may have the belief that all of a category's features are potentially relevant and should be learned, because knowledge of properties is ultimately useful for category use, even if it is not needed for classification (Markman & Ross, 2003). In motivated learners, such a reallocation could overcome competition effects, which have often been given an attentional explanation (Kruschke, 2001; Mackintosh, 1975).

That assumption might explain why there is no cue competition, but it is difficult for it to account for the opposite result, namely that learning a salient dimension actually facilitates learning of other properties, as found in Experiment 3 (and nonsignificantly occurred in Bott et al., 2007, Experiment 3). To help explain this, we refer to an argument made in earlier work on category formation (Kaplan & Murphy, 2000), that learning one or more features can provide a "hook" that aids learning of other features. For example, when an animal with bright red stripes appears, the subject may tentatively identify it as a Dax. Then, before responding, the subject may attend to the head and antlers and think, "Daxes seem to have the round head and those antlers." When the classification is confirmed via feedback, the association between those features and the correct category is reinforced, as Blair et al. (2009) argue.

In contrast, when there is no such salient, pretrained feature, subjects are more uncertain during the classification trial and therefore cannot associate the presented features to the expected category as efficiently. That is, when they see the animal with black stripes, they aren't confident about which category it is in, and so they do not use their initial category guess to encode other category-feature associations.

This hypothesis may be related to Urcelay and Miller's (2009) account of why overshadowing did not occur—indeed, instead they found potentiation—when the CS preceded the US by 20 s. Simplifying their account somewhat, they attributed this potentiation to the use of a configural memory of the two CSs. The animal remembered the configuration of the two stimuli (the target stimulus and the more salient overshadowing stimulus), and the target stimulus was sufficient to evoke that configural memory under those conditions at test. Since the overshadowing stimulus was so salient, this configural stimulus was strongly associated to the US, more than the target stimulus alone would have been. That seems analogous to our claim that highly salient stimuli (or those related to prior knowledge connected to the category) provide a hook for the other stimuli to be learned. There are no doubt important differences between conditioning experiments in rats and category-learning experiments in humans, but perhaps there are important similarities as well.

Thus, a major conclusion of this work is that it is questionable whether cue competition is a significant factor in learning family resemblance categories, at least of moderately plausible objects. We do not question the reliability of cue competition that has been found in many forms of classical conditioning and nonmetric cue learning. However, learning family resemblance categories may evoke additional attentional strategies that result in greater learning of all features, related to long-standing explanations of why we learn categories at all (Markman & Ross, 2003; Murphy, 2010; Rosch, 1978).

Why, then, did Kruschke and Johansen (1999) find so many cases of cue competition in their experiments using nonmetric cue learning? There are many differences between their task and ours, and it is not possible to know for certain which ones might account for this difference. One important difference is the small number of dimensions they used—two dimensions, each with two values. As a result, there were only four possible items in the entire experiment. Category membership was highly probabilistic, both with respect to any cue and to any stimulus. That is, the identical stimulus appeared in the two different categories rather than being assigned to only one. In our family resemblance structure, a given stimulus value is associated five-sixths of the time to one category and one-sixth of the time to the other. Given that each item has six dimensions, five of which have typical features, each classification was deterministic, even though the individual properties were probabilistic. Finally, we tested each dimension by presenting individual features on their own, which subjects classified. Therefore, our measure of feature learning was independent for each feature.

We speculate that it is the highly probabilistic nature of the stimuli that may be responsible for Kruschke and Johansen's (1999) finding of competition. In our experiment, people can learn rules such as, "When the animal has antlers, it also tends to have a fish tail and be a Dax." Although such rules will not be true all the time, they are usually true, and with enough features entering into such rules, the one feature in an item that does not match the others can be ignored. Accuracy can reach 100% if only three features are learned. In contrast, Kruschke and Johansen inform their subjects that if they learn well, they will reach 70–80% performance. In two of the cells with moderately predictive features (the .2, .2 condition in their Figure 3), the categories are equally likely. So, even if subjects had correctly learned the cue contingencies, on half the trials in that condition, they could not exceed chance.

Another, related possibility is the fact that all the cue combinations occur equally often. If we call the two dimensions X and Y, then there are four stimuli that appear equally often in the experiment: X_1Y_1, X_1Y_2, X_2Y_1, and X_2Y_2. In the family resemblance structure, the mere fact of co-occurrence helps to identify the category structure. That is, antlers plus fish tail occurs more often than antlers plus horse tail, thereby providing a clue that the two are associated to the same category. (This is possibly analogous to findings of configural learning in conditioning.) Because there is no such structure in the cue-learning example, associations among cues are not helpful.

It is more difficult to explain the results of Busemeyer et al. (1993), in part because they do not give a full report of their finding of cue cooperation. As noted above, when they did not tell subjects that one stimulus dimension was more important than the other, they found cooperation (reported in their footnote 5). But they did not report those data, and their paradigm was not a categorization task but function learning, in which people were given two pieces of information, levels of two hormones, and then had to make a numerical prediction about an outcome, plant growth. Thus, although their finding of cooperation is intriguingly similar to our Experiment 3, the paradigm is different enough from our category learning case to make us wary in drawing conclusions from it. However, it is useful in providing another example of cue cooperation effects in a human learning task.

Finally, as noted earlier, stimulus competition effects may not be as robust in animal learning as often assumed in the category-learning literature (Maes et al., 2016). Further tests can also reveal that what appears to be attenuated stimulus learning was instead a deficit in performance; for example, posttraining extinction of the overshadowing stimulus can reveal that the seemingly unlearned stimulus did acquire associative value (Kaufman & Bolles, 1981; Matzel et al., 1985). Thus, it is possible that the failure to see blocking in category learning may be in keeping with aspects of the associative learning literature that challenge the ubiquity of cue competition in animal learning.

Noxious Stimuli

Less easy to explain is the absence of any effect—competitive or cooperative—of shocking people and playing a loud sound, as compared to visual representations of the same properties. When these dimensions were presented as a normal part of each stimulus with no pretraining (Experiment 1), subjects did not even learn that dimension better than the others. In the shock condition, accuracy was identical for the critical and noncritical stimuli. Although the accuracy data for critical stimuli are noisy (since they are based on only two features, as opposed to 10 for the noncritical dimensions), we find it very surprising that everyone did not learn that the Kezes shock them (five out of six times) whereas the Daxes make a loud sound.[3] When the critical dimensions were pretrained (Experiment 2), accuracy increased to .85, which is still not perfect. (In Experiment 3, the pretrained colored dimensions were essentially perfect.) Only when we changed the category structure so that the noxious dimension was perfectly associated with categories did that dimension achieve near-perfect performance (Experiment 4). But even if the critical dimension was not always learned, one might expect the shock and sound to attract attention away from other dimensions. Yet there was no difference in learning as a function of noxious stimuli.

One possible explanation is that the visual representations were already so salient that both conditions were near the maximum levels of salience. The shock itself, therefore, did not actually increase attention to the "sting" feature. This is somewhat hard to believe, especially given the relatively poor performance (accuracy around .70) in classifying the critical features without a shock or sound. There was plenty of room for performance to improve. Since performance could and did improve with pretraining (Experiment 2), it doesn't seem correct to say that subjects were already focusing on the shock features (whether actual or only visual) the maximum amount.

One explanation for the failure of shock and sound to generate interference, or indeed any other effect, is the countervailing effects of increased attention and arousal. The sound and shock no doubt make subjects pay more attention to the task as a whole. This is true both because of the stimulus salience and the more elaborate preparations for that condition, involving the attachment of electrodes and calibration procedure, which probably made subjects

[3] If a standard guessing correction is applied, the observed accuracy of .73 for tested critical features results in an estimate of .46 actual accuracy. That is, the .73 reflects .46 known correct answers plus .27 correct guesses (and .27 incorrect guesses). Thus, only about half the subjects learned which categories the sound and shock occurred with.

feel that something interesting and novel was happening. However, the noxious nature of the shock (and, to a much lesser degree, the sound) may have had the effect of interfering with learning. Anticipatory anxiety over whether the next trial would contain a shock could be distracting during feedback, when much learning occurs (Blair et al., 2009), as could discomfort while the experiment continued. Indeed, this anxiety also seems to have prevented subjects from learning the critical features themselves. In Experiment 1, accuracy was low on the shock/sound dimension, presumably because the features were not perfect predictors. Learning probabilistic cues could be difficult under such conditions.

It does seem an unexpected coincidence that, on this account, the attentional effect of the shock is about the same size as the negative effect of stress, leading to no group differences in three experiments. Our proposal is clearly speculative, and we do not give great weight to it. This would be the traditional place in an article to suggest that further research into this topic is needed—or, indeed, a place to report other actually conducted experiments on shock's effect during learning. However, we must confess a reluctance to shock more subjects in the goal of determining why shock is not having an effect.

In the introduction, we described a theoretical and a descriptive goal of investigating noxious stimulus dimensions. Theoretically, our story is still incomplete. We have ruled out the notion that a noxious stimulus will be learned faster than a less noxious equivalent and then interfere with learning other features, as cue competition would predict. But why this doesn't happen is not yet clear. Furthermore, given the results of the colored textures, the correct question may actually be why shock and sound do not *increase* learning of other properties.

Descriptively, the picture is clearer. Our results suggest that when people learn about categories with a noxious component, like negative social events, yellow jackets, or uncomfortable medical treatments, they do not learn more or less about those categories than about similar ones lacking a noxious element. Instead, they learn about the same amount. Of course, people may choose to avoid exemplars of noxious categories, which then will have the effect of reducing their learning (Rich & Gureckis, 2015). If you run away every time a yellow jacket appears, you will never learn about its fascinating nesting habits, say. Furthermore, noxious objects may have particularly salient (non-noxious) properties, such as the distinctive yellow and black pattern of the yellow jacket, or red hourglass of the black widow spider. Those no doubt aid in detection and classification. However, when stimulus salience, amount of experience, and the other category properties are equated, as in our studies, there appears to be little difference in what else is learned in such categories.

References

Arcediano, F., Escobar, M., & Miller, R. R. (2004). Is stimulus competition an acquisition deficit or a performance deficit? *Psychonomic Bulletin & Review, 11,* 1105–1110. http://dx.doi.org/10.3758/BF03196744

Blair, M. R., Watson, M. R., & Meier, K. M. (2009). Errors, efficiency, and the interplay between attention and category learning. *Cognition, 112,* 330–336. http://dx.doi.org/10.1016/j.cognition.2009.04.008

Bott, L., Hoffman, A. B., & Murphy, G. L. (2007). Blocking in category learning. *Journal of Experimental Psychology: General, 136,* 685–699. http://dx.doi.org/10.1037/0096-3445.136.4.685

Busemeyer, J. R., Myung, I. J., & McDaniel, M. A. (1993). Cue competition effects: Empirical tests of adaptive network learning models. *Psychological Science, 4,* 190–195. http://dx.doi.org/10.1111/j.1467-9280.1993.tb00486.x

Dunsmoor, J. E., Martin, A., & LaBar, K. S. (2012). Role of conceptual knowledge in learning and retention of conditioned fear. *Biological Psychology, 89,* 300–305. http://dx.doi.org/10.1016/j.biopsycho.2011.11.002

Dunsmoor, J. E., & Murphy, G. L. (2014). Stimulus typicality determines how broadly fear is generalized. *Psychological Science, 25,* 1816–1821. http://dx.doi.org/10.1177/0956797614535401

Dunsmoor, J. E., Murty, V. P., Davachi, L., & Phelps, E. A. (2015). Emotional learning selectively and retroactively strengthens memories for related events. *Nature, 520,* 345–348. http://dx.doi.org/10.1038/nature14106

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review, 66,* 183–201. http://dx.doi.org/10.1037/h0047707

Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General, 117,* 227–247. http://dx.doi.org/10.1037/0096-3445.117.3.227

Goddard, M. J., & Jenkins, H. M. (1988). Blocking of a CS–US association by a US–US association. *Journal of Experimental Psychology: Animal Behavior Processes, 14,* 177–186. http://dx.doi.org/10.1037/0097-7403.14.2.177

Hoffman, A. B., Harris, H. D., & Murphy, G. L. (2008). Prior knowledge enhances the category dimensionality effect. *Memory & Cognition, 36,* 256–270. http://dx.doi.org/10.3758/MC.36.2.256

Hoffman, A. B., & Murphy, G. L. (2006). Category dimensionality and feature knowledge: When more features are learned as easily as fewer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 301–315. http://dx.doi.org/10.1037/0278-7393.32.3.301

Joëls, M., Pu, Z., Wiegert, O., Oitzl, M. S., & Krugers, H. J. (2006). Learning under stress: How does it work? *Trends in Cognitive Sciences, 10,* 152–158. http://dx.doi.org/10.1016/j.tics.2006.02.002

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.

Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In R. M. Church & B. A. Campbell (Eds.), *Punishment and aversive behavior* (pp. 279–296). New York, NY: Appleton-Century Crofts.

Kaplan, A. S., & Murphy, G. L. (2000). Category learning with minimal prior knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 829–846.

Kaufman, M. A., & Bolles, R. C. (1981). A nonassociative aspect of overshadowing. *Bulletin of the Psychonomic Society, 18,* 318–320. http://dx.doi.org/10.3758/BF03333639

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99,* 22–44. http://dx.doi.org/10.1037/0033-295X.99.1.22

Kruschke, J. K. (2001). Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology, 45,* 812–863. http://dx.doi.org/10.1006/jmps.2000.1354

Kruschke, J. K., & Johansen, M. K. (1999). A model of probabilistic category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1083–1119. http://dx.doi.org/10.1037/0278-7393.25.5.1083

LaBar, K. S., & Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience, 7,* 54–64. http://dx.doi.org/10.1038/nrn1825

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review, 82,* 276–298. http://dx.doi.org/10.1037/h0076778

Maes, E., Boddez, Y., Alfei, J. M., Krypotos, A.-M., D'Hooge, R., De Houwer, J., & Beckers, T. (2016). The elusive nature of the blocking effect: 15 failures to replicate. *Journal of Experimental Psychology: General, 145,* e49–e71. http://dx.doi.org/10.1037/xge0000200

Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin, 129,* 592–613. http://dx.doi.org/10.1037/0033-2909.129.4.592

Matzel, L. D., Schachtman, T. R., & Miller, R. R. (1985). Recovery of an overshadowed association achieved by extinction of the overshadowing stimulus. *Learning and Motivation, 16,* 398–412. http://dx.doi.org/10.1016/0023-9690(85)90023-2

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207–238. http://dx.doi.org/10.1037/0033-295X.85.3.207

Murphy, G. L. (2010). What are categories and concepts? In D. Mareschal, P. C. Quinn, & S. E. G. Lea (Eds.), *The making of human concepts* (pp. 11–28). Oxford, England: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199549221.003.02

Murphy, G. L., & Allopenna, P. D. (1994). The locus of knowledge effects in concept learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 904–919. http://dx.doi.org/10.1037/0278-7393.20.4.904

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 104–114. http://dx.doi.org/10.1037/0278-7393.10.1.104

Pearce, J. M., & Bouton, M. E. (2001). Theories of associative learning in animals. *Annual Review of Psychology, 52,* 111–139. http://dx.doi.org/10.1146/annurev.psych.52.1.111

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review, 87,* 532–552. http://dx.doi.org/10.1037/0033-295X.87.6.532

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning: II. Current research and theory* (pp. 64–99). New York, NY: Appleton-Century-Crofts.

Rich, A. S., & Gureckis, T. M. (2015). The attentional learning trap and how to avoid it. In R. Dale, C. Jennings, P. Maglio, T. Matlock, D. Noelle, A. Warlaumont, & J. Yoshimi (Eds.), In *Proceedings of the 37the annual conference of the Cognitive Science Society* (pp. 1973–1978). Austin, TX: Cognitive Science Society.

Rodrigues, S. M., LeDoux, J. E., & Sapolsky, R. M. (2009). The influence of stress hormones on fear circuitry. *Annual Review of Neuroscience, 32,* 289–313. http://dx.doi.org/10.1146/annurev.neuro.051508.135620

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Erlbaum.

Rosch, E., & Mervis, C. B. (1975). Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology, 7,* 573–605. http://dx.doi.org/10.1016/0010-0285(75)90024-9

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16,* 225–237. http://dx.doi.org/10.3758/PBR.16.2.225

Urcelay, G. P., & Miller, R. R. (2009). Potentiation and overshadowing in Pavlovian fear conditioning. *Journal of Experimental Psychology: Animal Behavior Processes, 35,* 340–356. http://dx.doi.org/10.1037/a0014350

Received September 12, 2016
Revision received March 27, 2017
Accepted March 28, 2017 ■