Behavioural Processes 123 (2016) 63–73

Contents lists available at ScienceDirect

Behavioural Processes

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

Attentional shifts in categorization learning: Perseveration but not learned irrelevance

Leyre Castro*, Edward A. Wasserman

Department of Psychological and Brain Sciences, The University of Iowa, United States

ARTICLE INFO

Article history:
Received 13 July 2015
Received in revised form 2 November 2015
Accepted 2 November 2015
Available online 5 November 2015

Keywords:
Categorization
Shift cost
Learned irrelevance
Attentional perseveration
Animal cognition

ABSTRACT

Once a categorization task has been mastered, if features that once were relevant become irrelevant and features that once were irrelevant become relevant, a decrement in performance—a shift cost—is typically observed. This shift cost may reflect the involvement of two distinguishable factors: the inability to release attention from a previously relevant feature (i.e., attentional perseveration) and/or the inability to re-engage attention to a previously irrelevant feature (i.e., learned irrelevance). Here, we examined the nature of this shift cost in pigeons. We gave four groups of pigeons a categorization task in which we monitored their choice accuracy; at the same time, we tracked the location of their pecks to the relevant and irrelevant attributes of the stimuli to determine to which attributes the birds were attending during the course of learning. After identical training in Phase 1, the roles of the relevant/irrelevant features were changed in Phase 2, so that one group could show only learned irrelevance, a second group could show only attentional perseverance, a third group could show both, and a fourth control group could show neither of these effects. Results disclosed evidence of attentional perseverance, but no evidence of learned irrelevance, either in accuracy or in relevant feature tracking. In addition, we determined that pigeons' allocation of attention to the relevant features followed rather than preceded an increase in choice accuracy. Overall, our findings are best explained by theories which propose that attention is learned and deployed to those features that prove to be reliable predictors of the correct categorization response (e.g., George and Pearce, 2012; Kruschke, 2001; Mackintosh, 1975).

This dog is a retriever, but that one is a husky. You can correctly classify these different breeds of dogs when you pay particular attention to their type of fur, the shape of their ears, and the color of their eyes. Observing that both dogs have four legs, two eyes, and a tail does not allow you to distinguish between the different breeds. The type of fur, the shape of ears, and eye color are, in this case, relevant features for classifying the dogs, whereas the number of legs and eyes, and having a tail are irrelevant features. Thus, to be able to categorize objects and events, organisms must focus on those features of the stimuli that are relevant to correctly perform the task and to disregard those features that are not.

Lawrence (1949) first set forth the idea that those stimuli that are relevant to the solution of a discrimination problem will be paid more attention than those that are irrelevant. Several subsequent theorists expanded on that notion; they suggested that attention to both the relevant and irrelevant features of the stimuli changes during the course of learning. For example, Mackintosh (1965, 1975),) proposed that, during discrimination learning, relevant stimuli will become more accurate and reliable predictors of the outcome of a trial than will irrelevant stimuli. Consequently, attention to the relevant stimuli should increase, whereas attention to the irrelevant stimuli should decrease. Indeed, both humans (e.g., Best et al., 2013; Blair et al., 2009a,b; Rehder and Hoffman, 2005a) and animals (e.g., Castro and Wasserman, 2014; Dittrich et al., 2010; Dopson et al., 2011; George and Pearce, 1999) do appear to attend more to those stimulus attributes that are relevant than to those stimulus attributes that are irrelevant to solving a discrimination problem.

The challenging reality is that attention cannot be directly measured, but must instead be inferred from an organism's behavior, thereby making it difficult to provide clear and compelling evidence of momentary increases and decreases in attention. Ideally, evidence of attention should be obtained while learning is taking place; yet, most studies of attention rely on behavioral measures obtained after learning has taken place (e.g., Dopson et al., 2011; Kruschke, 1996; Mackintosh and Little, 1969; Pearce et al., 2008; Roberts et al., 1988).

* Corresponding author at: Department of Psychological and Brain Sciences, E11 Seashore Hall, The University of Iowa, Iowa City, IA 52242, United States.
E-mail address: leyre-castroruiz@uiowa.edu (L. Castro).

http://dx.doi.org/10.1016/j.beproc.2015.11.001
0376-6357/

Because there seems to be a close connection between the direction of eye gaze and changes in attention and because there is empirical evidence suggesting that eye movements and attention operate simultaneously (e.g., Deubel and Schneider, 1996), eye-tracking technology has been applied to study the deployment of attention while learning is taking place. In human category learning, for example, Rehder and Hoffman (2005a) tracked participants' eye movements while they were solving a categorization task in which some elements were relevant to correctly classifying the category exemplars, whereas other elements were not. Rehder and Hoffman found that, as learning progressed, participants' allocation of attention gradually shifted toward the relevant elements of the stimuli. Participants' eye gaze direction revealed that they indeed allocated their attention in a way that optimized their classification performance.

As Castro and Wasserman (2014) suggested, a possible analog to eye gaze direction in humans may be peck location in birds. In that study, pigeons had to learn to classify exemplars from two different artificial categories; each of the categories was associated with one particular response button. The categories contained some relevant features and some irrelevant features. When a category exemplar was presented on the screen, the pigeons had to peck it several times. However, they did not have to peck at any specific feature of the category exemplar; pecks to both the relevant and irrelevant features equally satisfied the observing response requirement. We found that, as categorization accuracy increased, pecks at the relevant features of the category exemplars progressively increased as well (conversely, pecks at the irrelevant features progressively fell as categorization accuracy increased). This relationship suggests that, as pigeons were learning to categorize the stimuli, they were also paying increasing attention to their relevant attributes. So, pigeons' attention to both relevant and irrelevant stimulus attributes can effectively be monitored during learning of a categorization task.

Once attention is focused on the relevant attributes of the stimuli, how might attention be affected if what had been relevant or irrelevant suddenly changes? Under these circumstances, attention must shift to the new relevant cues to be able to solve the new task; but, until this new learning takes place, a decrement in performance—a shift cost—will necessarily be observed. When those cues that had been relevant become irrelevant and those cues that had been irrelevant become relevant, impaired performance may result from two distinguishable mechanisms: the inability to release attention from a previously relevant feature (i.e., attentional perseveration) and/or the inability to re-engage attention to a previously irrelevant feature (i.e., learned irrelevance).

Prior research on human and animal associative learning has typically studied attentional shifts by comparing performance when new stimuli from a prior relevant dimension become the relevant stimuli (intradimensional shift) to performance when new stimuli from a prior irrelevant dimension become the relevant stimuli (extradimensional shift). Generally, the intradimensional shift yields faster learning than the extradimensional shift (Birrel and Brown, 2000; George and Pearce, 1999; Kruschke, 1996; Mackintosh and Little, 1969; Roberts et al., 1988; Shepp and Eimas, 1984). The explanation for this advantage relies on attentional mechanisms. Attention during original learning is presumably deployed to the relevant dimension; when a shift occurs, attention to that relevant dimension perseverates after the shift. Therefore, according to attentional theories (e.g., George and Pearce, 2012; Kruschke, 2001; Mackintosh, 1975), the intradimensional shift is easier because attention is already focused on the dimension that is relevant to solving the new task.

Other studies have shown that when a cue is uncorrelated with reinforcement or turns out to be irrelevant to the solution of a discrimination, subsequent learning about that cue is often retarded (e.g., Galbraith, 1973; Hall, 1976; Le Pelley and McLaren, 2003; Mackintosh, 1973; Winefield, 1978). This phenomenon has been called learned irrelevance to emphasize that a decrease in attention had taken place because the organism learned that the cue was irrelevant to the task at hand. Attentional theories can, of course, also explain learned irrelevance. Because irrelevant cues do not predict the outcome of a trial or the correct response in a discrimination task, attention to them decreases; hence, when prior irrelevant cues become relevant, it takes longer to focus attention on them and, consequently, learning is delayed.

Despite the richness of this prior research, there has been little effort in the associative learning literature to analyze the interrelation between attentional perseveration and learned irrelevance. For example, when intradimensional and extradimensional shifts are compared, one cannot decisively determine if the intradimensional advantage is due to an increase in attention to relevant cues that persists in subsequent phases—attentional perseveration—or to a decrease in attention to irrelevant cues from which is difficult to recover in subsequent phases—learned irrelevance; both effects may be taking place.

There have nonetheless been studies on human executive function that have tried to assess the separate contributions of attentional perseveration and learned irrelevance. For example, Maes et al. (2004) presented college students with different stimuli that could vary in three possible dimensions—color, shape, or number—and that could be relevant, irrelevant, or that remained constant. Once the participants reached the learning criterion, the role of the dimensions was changed without any warning. In one group, the former relevant dimension became irrelevant and the former irrelevant dimension became relevant, so that both attentional perseveration and learned irrelevance errors were possible. In a second group, the former relevant dimension became irrelevant and the former constant dimension became relevant; thus, participants could only make errors based on attentional perseveration (continuing to attend to the former relevant dimension). Finally, in the third group, the former irrelevant dimension became relevant and the previous constant dimension became irrelevant; thus, participants could only make errors based on learned irrelevance (failing to attend to the former irrelevant dimension). Maes et al. found that learned irrelevance errors exceeded attentional perseveration errors; moreover, there was no difference between the group that could only make learned irrelevance errors and the group that could make both types of errors. Thus, participants' performance was greatly affected by learned irrelevance, but not by attentional perseveration.

Here, we examined the nature of the shift cost in a category learning task in pigeons; we were particularly interested in determining whether learned irrelevance (LI), attentional perseveration (P), or both (P + LI) take place when relevant and irrelevant features change their roles. In Phase 1 of training—which was identical for the four groups of pigeons—the birds had to learn to categorize two different artificial categories: Categories A and B. Two relevant features defined Category A, two other relevant features defined Category B, and four irrelevant features were common to Categories A and B.

For all four groups, the relevant and irrelevant features in Phase 2 were different from those in Phase 1, so all pigeons had to learn a new category discrimination. In group P + LI, the prior irrelevant features became relevant and the prior relevant features became irrelevant in Phase 2, so a shift cost could be due to both attentional perseverance and learned irrelevance (as indicated by the initials P and LI in its name). Group P received new relevant features and the prior relevant features became irrelevant in Phase 2, so a shift cost could only be due to attentional perseverance (as indicated by the initial P in its name). For group LI, the prior irrelevant features became relevant in Phase 2 and the irrelevant features were new,