The Limits of Learning: Exploration, Generalization, and the Development of Learning Traps

Alexander S. Rich and Todd M. Gureckis
New York University

Learning usually improves the accuracy of beliefs through the accumulation of experience. But are there limits to learning that prevent us from accurately understanding our world? In this article we investigate the concept of a "learning trap"—the formation of a stable false belief even with extensive experience. Our review highlights how these traps develop through the interaction of learning and decision making in unknown environments. We further document a particularly pernicious learning trap driven by selective attention, a mechanism often assumed to facilitate learning in complex environments. Using computer simulation, we demonstrate the key attributes of the agent and environment that lead to this new type of learning trap. Then, in a series of experiments we present evidence that people robustly fall into this trap, even in the presence of various interventions predicted to meliorate it. These results highlight a fundamental limit to learning and adaptive behavior that impacts individuals, organizations, animals, and machines.

Keywords: learning, exploration, decision making, selective attention, biases

Supplemental materials: http://dx.doi.org/10.1037/xge0000466.supp

Conventional wisdom holds that learning from experience brings us closer to understanding the world around us. For instance, with enough practice we can learn even complex skills like playing an instrument or speaking a new language. However, in many situations learning new things trades off against the goal of maximizing reward. Our desire to try every restaurant in a city has to be balanced against our desire to have a good meal tonight. This trade-off, known as the exploration–exploitation dilemma, has been extensively studied in computer science (Sutton & Barto, 1998), organizational behavior (March, 1991), and psychology (Mehlhorn et al., 2015). The central premise of this article is that in these situations even adaptive learning can lead people (as well as other agents such as organizations, animals, and machines) to form robust false beliefs about the world. For instance, consider a child who is exploring different foods, and finds that she dislikes spinach and cucumbers. As a result of this experience, she may falsely believe she dislikes all vegetables and avoid vegetables in the future, never learning that she really likes broccoli even as she continues to explore other foods.

Inspired by the organizational-learning theorists Levinthal and March (1993), we call the development of these types of false or incomplete beliefs "learning traps" (see also, Erev, 2014; Teodo-rescu & Erev, 2014). We use this term because the failure of beliefs to move toward truth in these situations is caused not by ignorance, but instead by the interaction of exploratory choice with what is already believed. Learning traps represent a very general class of phenomena that reoccur in many situations for many types of learners.

In this article, we aim to provide a comprehensive summary of known learning traps and the factors that generate them. We begin by describing a variety of disparate findings that can be interpreted as reflecting learning traps, in fields from organizational theory (Denrell, 2007) to social psychology (Fazio, Eiser, & Shook, 2004) and domains from foraging behavior (Niv, Joel, Meilijson, & Ruppin, 2002) to stereotype formation (Denrell, 2005) and nepotism (Liu, Eubanks, & Chater, 2015). Our synthesis shows that a crucial ingredient for the formation of learning traps is generalization from past experience that "traps" people into premature exploitative behaviors and limits further learning and belief revision. Within this review, we progress from traps caused by simple forms of generalization to traps caused by more complex forms of generalization, and provide a unifying framing in which learning traps represent intrinsic limits of learning and adaptation.

In the empirical portion of the article we present a computational model and a series of simulations and studies with which we explored one particular learning trap. This learning trap can form when people selectively attend to a subset of relevant stimulus attributes while learning, which is interesting because selective attention is most often considered a facilitative mechanism for learning (e.g., improving speed of acquisition and performance by filtering out irrelevant information). As a test of the robustness of

this attentional learning trap, we explored a variety of manipulations aimed at reducing the prevalence of this trap for both simulated and human learners. To foreshadow, our results confirm the robustness and prevalence of learning traps. We conclude by discussing additional routes to the prevention of learning traps and highlighting other unstudied settings in which they are likely to emerge.

Trapped by Stochasticity (i.e., Noise)

The simplest learning traps arise in situations where a person (or other agent) repeatedly encounters a prospect and must decide whether or not to interact with it and experience a consequential outcome (e.g., approach–avoidance). A prospect might, for example, be another person, with the outcome of interaction being a pleasant or unpleasant conversation. A reasonable assumption in these cases is that the prospect will be stable over time, and that past experiences with the prospect will generalize well to future experiences. When this assumption holds, then correct knowledge about the whether the prospect is good or bad can be gained from the first encounter. If instead there is stochasticity in the outcomes, then there are times when generalization from early encounters will be misleading, and learning traps can develop. Note that for the purposes of this article stochasticity means random, unsystematic noise over time or across instances. Stochasticity that causes systematic changes over time, as described, for example, by Behrens, Woolrich, Walton, and Rushworth (2007), could be the source of additional learning traps but will not be directly addressed.

Perhaps the paradigmatic example of this type of learning trap is the "hot stove effect," introduced by Denrell and March (2001). If a prospect is positive on average, but stochastic, early experiences may make the prospect appear negative. If a person learns from and generalizes from these early experiences, they will begin to avoid the prospect. Once this has occurred, learning halts because no more information is gained about the prospect, and the person is left with a false negative belief that is difficult to overcome. Avoiding the option appears to be the safest strategy, but it is exactly the strategy that prevents learning the true structure of the environment. Notably, this trap does not occur when a prospect is usually negative but mistakenly believed to be positive. In this case, the false belief causes continued interaction, which eventually allows the belief to be corrected.

The hot stove effect is intrinsic to experiential, reward-driven learning in a variable environment, and is exhibited even by an optimal agent (Le Mens & Denrell, 2011), though it can be mitigated through a variety of factors including more persistent exploration (Rich & Gureckis, 2017). In addition, the effect can explain underexploration and apparent risk aversion across a broad range of experience-based decision-making domains, ranging from the decisions of firms to the foraging behavior of bees (Denrell & March, 2001; Niv et al., 2002). It also provides one explanation for why people tend to prefer novel prospects over previously experienced ones (Le Mens, Kareev, & Avrahami, 2016), as well as for why people prefer ingroup-members (with whom they must interact) over outgroup-members (with whom they repeatedly choose whether or not to interact; Denrell, 2005; Liu et al., 2015).

A second learning trap that develops from the expectation of stable outcomes is the underweighting of rare events. This learning trap was identified by researchers in the decisions from experience literature (Hertwig, Barron, Weber, & Erev, 2004), using a sampling paradigm in which participants repeatedly tested several prospects, paying a cost of time and sometimes money for each sample (Juni, Gureckis, & Maloney, 2016), before selecting one for a consequential choice. As participants explore each alternative they learn the common outcomes well and become confident enough to make a final, exploitative choice, but their samples often fail to include rare outcomes. Thus, people's beliefs about prospect values when making their final decisions systematically underweight these rare outcomes. This underweighting of rare events may also occur in settings where all choices are consequential, both in the lab (Teodorescu & Erev, 2014) and for real-world rare events ranging from driving accidents (Fuller, 1991) to nuclear disasters and scientific discoveries (Levinthal & March, 1993). Due to its interaction with the hot stove effect, the underweighing of rare events is likely more persistent when the rare event is positive, and the prospect may thus be mistakenly avoided. When the rare event is negative, the decision maker is likely to continue sampling the prospect and discover it in the long run, though the result of this may be disastrous. For example, failure to experience the rare event of discovery may drive someone out of science permanently, while failure to experience the rare event of a car accident may lead someone to continue driving without a seat belt until the event is experienced.

Trapped by Similarity

In environments with not just one prospect but several, new learning traps become possible as patterns of generalization become more complex. Rather than just determining how much to generalize from the past to the future, an agent must determine how much to generalize from experience with one prospect to experience with another.

One answer for how to generalize across prospects is based on their similarity. If two prospects are highly similar, for example sharing many perceptual features, then it may be safe to assume that the outcomes of interacting with them are similar as well (Shepard, 1987; Tenenbaum & Griffiths, 2001; Tversky, 1977). This notion of similarity is a core aspect of many theories of human category learning and categorization. Studies of natural and artificial categories show that a novel item is judged to be a typical category member to the extent that it is similar to other members of the category and dissimilar to members of different categories (Rosch & Mervis, 1975). In addition, many formal models of categorization determine the membership of an item through similarity-based comparison to past category members or to a more abstract category representations (Love, Medin, & Gureckis, 2004; Nosofsky, 1986).

The expectation that similar prospects will yield similar outcomes, like the expectation that prospects are stable over time, is a reasonable one. However, in value-based decision making with choice-contingent feedback, similarity-based generalization may cause the value of prospects near the boundary between good and bad outcomes to be permanently misestimated. This is particularly likely to be the case for positive prospects that are similar to negative ones. When a novel prospect is similar to other prospects with which an agent has had negative experiences, the decision maker must choose to either approach it to learn about its value, or

to avoid it, exploiting what it has learned from past experience. If the prospect is avoided, this creates a learning trap in which the decision maker never learns the true value of positive prospects that fall "in the shadow" of similar negative prospects that were experienced earlier.

Fazio, Eiser, and Shook (2004) confirmed the existence of this trap in a series of experiments investigating how people learn to approach or avoid in an environment where multiple distinct prospects varied along two continuous dimensions. (Prospects were "beans" that varied in shape and number of speckles.) They found a pronounced learning asymmetry, with participants correctly avoiding negative prospects, but often incorrectly avoiding positive prospects as well. When participants were given full (foregone) information, and told the value of prospects they avoided, the asymmetry disappeared, supporting the idea that they had entered a learning trap in which their behavior prevented further experience and learning. A related simulation study found that the same effect emerged in a connectionist network that generalized prospect values based on similarity (Eiser, Fazio, Stafford, & Prescott, 2003).

While this learning trap was first analyzed as a way of understanding the development of social attitudes, it is likely to emerge in many of the situations in which the hot stove effect has been observed. To take one example, Niv, Joel, Meilijson, and Ruppin (2002) found that simulated bees came to avoid flowers that produced a variable amount of nectar, even if they produced more nectar on average than stable flowers. One would predict that bees would also avoid flowers of a species that looks similar to a low-nectar species, even if the species itself produces a large, stable amount of nectar.

Trapped by Selective Attention

For the bulk of this article we will focus on a relatively unexplored learning trap that develops due to selective attention, one of the most basic forms of rapid generalization that extend beyond similarity. A classic finding in category learning is that the ease of learning a category structure depends not only on the similarity of the exemplars within and between categories, but on the number of dimensions required to distinguish categories (Nosofsky, Gluck, Palmeri, McKinley, & Glauthier, 1994; Shepard, Hovland, & Jenkins, 1961). Subsequent theories of categorization have posited that people adapt their allocation of attention to optimize performance, selectively attending those dimensions that discriminate categories while ignoring those that are irrelevant (Kruschke, 1992; Nosofsky, 1986). Thus, upon encountering an item, an agent with selective attention will weigh past experiences with items that are similar on attended dimensions strongly, while weighing experiences with items that are similar on nonattended dimensions weakly or not at all. Recent work in reinforcement learning has shown that selective attention plays a role in value based decision making as well (Niv et al., 2015).

While selective attention often aids learning, it creates a bias toward attention to fewer dimensions. That is, a learner with selective attention will initially expect there to be few dimensions of a prospect relevant to its outcome, and will only be induced to believe there are additional relevant dimension with more data. This bias means there are cases where selective attention can inhibit later learning, particularly when a person comes to ignore

a dimension that is later useful. This can occur in blocking and backward blocking, phenomena in which associating a single cue with an outcome prevents learning about other, concurrently presented cues (Kruschke & Blair, 2000; Mackintosh, 1975). It also happens in cases where a person learns first one category structure and then a second in which a previously irrelevant dimension must be attended (Hoffman & Rehder, 2010; Kruschke, 1996).

The bias toward categories with few relevant dimensions can also lead to a distinct learning trap. For example, consider the schematic diagram in Figure 1. Here, a hypothetical agent repeatedly encounters prospects that vary on two binary dimensions. These dimensions jointly determine the prospect quality in a somewhat complex manner, with prospects that have a value of 1 on both dimensions yielding negative outcomes and all other prospects yielding positive outcomes.[1] Prospects that are approached yield a deterministic positive or negative outcome, while those that are avoided provide no information. The agent must explore the environment by approaching uncertain prospects to learn about them, but also must eventually exploit what it has learned and avoid prospects it expects to be negative if it wishes to maximize positive outcomes.

In this setting, an agent with selective attention will attempt to learn not just which encountered prospects are positive and negative, but also which dimensions are useful and worth attending when deciding what to approach and what to avoid. As the agent gains experience, it will adjust its allocation of attention to optimize its performance based on its observations.

The true structure of the environment incentivizes attention to both dimensions. Despite this, variation in the prospects the agent happens to explore early on may cause one dimension to be attended more than the other and believed to have a stronger effect on prospect outcome. If this tendency is strong enough, the agent may begin to exploit its perceived knowledge and act based on that dimension alone, approaching prospects that have a value of 0 on that dimension, and avoiding those with a value of 1.

Once learned attention influences behavior in this way, the agent has entered a learning trap. All prospects with a value of 1 on the attended dimension are avoided, including those that have a value of 0 on the unattended dimension, with which the agent has had no negative experiences. The bias away from these prospects is persistent, since the agent avoids all prospects that would provide evidence of the importance of the second dimension. The agent may avoid a positive region of the environment indefinitely, and may also consequently hold false beliefs about how the environment is divided into meaningful categories.

Trapped by Simplicity: A Unifying View

In the following section, we will use a version of ALCOVE (Kruschke, 1992), a well-known model of human category learning with a selective attention mechanism, to explore the conditions under which an attentional learning trap can develop. However, we hypothesize that other models of categorization, including models based on very different principles than ALCOVE and lacking

---

[1] While we assume an interactive, deterministic relationship between the dimensions and the binary outcome, a similar structure can occur if two dimensions have independent, logistic relationships to the outcome (e.g., $p(\text{negative}) = 1 / (1 + \exp(-3(2d_1 + 2d_2 - 3))))$.
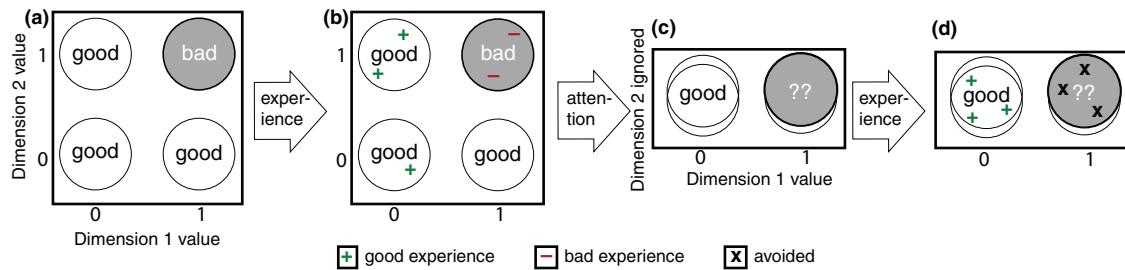
Figure 1. (a) A deterministic environment containing prospects with two binary features, where only prospects possessing both features are negative. By attending to both features, a decision maker can avoid all negative prospects while exploiting all positive prospects. (b) Early experience happens to highlight the relevance of Feature 1. (c) The agent begins to attend only Feature 1, and to ignore Feature 2, making items with and without Feature 2 appear equivalent. Items without Feature 1 are positive, while the value of items with Feature 1 appear stochastic. (d) The agent now avoids items with Feature 1, because some are negative. This prevents the agent from gaining feedback that would cause it to change its behavior. See the online article for the color version of this figure.

explicit selective attention mechanisms, would likely fall into a similar trap for slightly different computational reasons.

For example, we expect that rule-based models of categorization, such as the rule-plus-exception model (Nosofsky, Palmeri, & McKinley, 1994) and the rational rules model (Goodman, Tenenbaum, Feldman, & Griffiths, 2008), would fall into the "attentional" learning trap quickly and easily, because they prefer simple, one-dimensional rules. In the category structure described in Figure 1, either model would be likely to form a rule to respond based on Dimension 1 or Dimension 2 exclusively. While both models will adopt more complex rules if experience warrants, the use of a unidimensional rule would prevent the proper feedback from being obtained if feedback is choice-dependent. It has long been appreciated that selective attention and rule-based categorization are related phenomena (Kruschke, 1992).

Clustering models such as SUSTAIN (Love et al., 2004) and the rational model (Anderson, 1991) could also fall into the trap. Both models tend toward simple category representations with few clusters, and could become trapped if they attempted to form a two-cluster representation of the environment in Figure 1 and assigned a region of positive items to the negative cluster. Again, this mistake would not be corrected because prospects in the negative cluster would be avoided, preventing the model from learning that its representation was incorrect and that another cluster should be recruited. (In the case of SUSTAIN, the trap would be exacerbated by an attentional tuning mechanism similar to ALCOVE's).

The common thread joining these disparate models and making them susceptible to learning traps is their preference for simplicity. Thus, what we call an "attentional" learning trap could also be considered more generally as a "simplicity" learning trap; an agent comes to believe the environment has a simpler structure than it truly does, which prevents it from exploring further to come to a more complex, more correct belief. Other learning traps, including those based on stochasticity and similarity, could be derived from a simplicity-based bias as well. An environment that is stable is simpler than one that changes, and an environment where similar prospects produce similar outcomes is simpler than one where they do not. Indeed, many aspects of cognition seem to have an inductive bias toward simple inferences and generalizations (Chater & Vitányi, 2003; Feldman, 2003).

As researchers in machine learning have shown, some degree of inductive bias is necessary for learning to proceed—without it, a huge number of generalizations will be equally plausible from any finite set of experiences (Geman, Bienenstock, & Doursat, 1992; Mitchell, 1980). Thus, the tendency to believe the world is simple may be a necessary component for effective learning, and one that leads to better performance in many environments even as it inevitably leads to worse performance in others (Wolpert, 1996). This differentiates learning traps from other phenomena such as confirmation bias that prevent belief revision (Nickerson, 1998). Confirmation bias is generally (though not always, see Navarro and Perfors (2011)) thought of as reflecting a flaw in inference and reasoning, but learning traps, in their most basic form, are not a true suboptimality. Rather, they are an unavoidable byproduct of learning—whether by people, machines, animals, or organizations—in situations where the need for additional information must be traded off with the need for reward maximization.

Computational Explorations of the Attentional Learning Trap

Our discussion so far has highlighted intuitive reasons why learning traps form as well as cited empirical evidence that humans are susceptible to some these traps (e.g., the hot stove effect). However, it is helpful to explore this issue in slightly more detail through computer simulation. Computer simulations allow us to conduct counterfactual analyses that help isolate necessary and sufficient aspects of the learner–environment interaction that provoke learning traps. They also help us to set the stage for our later empirical studies. We particularly focus our work on the novel attentional learning trap introduced above.

ALCOVE-RL: A Computational Model of Learning and Generalization From Experience

While the attentional learning trap could be fruitfully analyzed from multiple perspectives, we approach it from the angle of selective attention, using a modified version of the ALCOVE (Kruschke, 1992) model of categorization, for several reasons. ALCOVE is a successful and widely used model of human category learning (see, e.g., Nosofsky et al., 1994), and can be con-

veniently modified for use in a reinforcement-learning context (see Jones & Cañas, 2010). Additionally, ALCOVE shares deep formal structure with the simple adaptive learning models used to study stochasticity-based learning traps (Denrell & March, 2001) and the connectionist models used to study similarity-based learning traps (Eiser et al., 2003). Like these models, ALCOVE learns in an incremental, mechanistic manner to reduce error (or increase reward). This makes ALCOVE an interesting model in which to demonstrate a learning trap, because unlike models that directly aim for simple representations, its biases are implicit. Just as the update rules of adaptive learning models imply a "built in" assumption that the environment is fairly stable, the attentional learning rules of ALCOVE may imply a built-in assumption that there are few relevant dimensions, which will cause a learning trap when this assumption is incorrect and feedback is choice-dependent.

ALCOVE was originally designed for learning in supervised category-learning tasks with corrective feedback. To use it to study learning traps, we created a modified model, ALCOVE-RL, for use in a reinforcement-learning task where feedback is continuous and action-dependent, following the work of Jones & Cañas (2010).

As in the original ALCOVE model, when a new prospect is presented to our model it first activates the set of input nodes a^{in} based on its value on each dimension. This activation then spreads to a set of hidden nodes a^{hid} representing the space of exemplars (i.e., possible prospects). Each hidden node a_j^{hid} has a position h_j in the psychological space defined by the input dimensions, and is activated based on its similarity to the input node values.

$$a_j^{hid} = \exp\left[-c\left(\sum_i \alpha_i |h_{ji} - a_i^{in}|\right)\right]$$

The similarity function is parameterized by a specificity constant c, as well as learnable attention parameters a determining the breadth of generalization along each dimension. Essentially, internal representations of past prospects are activated based on their similarity to the current input, with similarity counting more on dimensions with high attention settings than on those with low ones.

The activation of the hidden nodes then spreads to output nodes a^{out}, which are each associated with an action. As in ALCOVE the output activation of a_k^{out} is determined by first taking the sum of the a^{hid} activations, weighted by the learnable parameters w_k. The w_k parameters represent the model's association of each hidden node with a reward value. The summed outputs are then divided by the sum of a^{hid} activations to produce a weighted average. This normalization is not necessary in ALCOVE for category learning, where the only goal is for the correct output node to have the highest activation. But in reinforcement learning, normalization ensures that output activation reflects predicted reward and does not overshoot or undershoot this value due to overall activation strength.

$$a_k^{out} = \sum_j w_{kj} a_j^{hid} / \sum_j a_j^{hid}$$

Following ALCOVE and classic models of decision making (Luce, 1959), the model's choice of action then follows a probabilistic choice rule, with its degree of determinism controlled by the parameter ϕ.

$$Pr(K) = \exp(\phi a_K^{out}) / \sum_k \exp(\phi a_k^{out})$$

The feedback mechanism in ALCOVE-RL again diverges from the original ALCOVE implementation. In ALCOVE, feedback is received for all outputs, and there is no penalty for outputs with greater magnitude than the correct values. In our model, upon receiving a reward r, the feedback t_k to the model is

$$t_k = \begin{cases} r & \text{if } k \text{ is the chosen action} \\ a_k^{out} & \text{otherwise} \end{cases}$$

In other words the model is told that the predicted reward for the chosen action should move toward r, and that no change is needed for the predictions on other actions since no information about their outcomes was gained.

Given this action-dependent feedback, the model then updates its exemplar-action weights and attention weights through the same mechanisms at ALCOVE in order to decrease future error.

$$\Delta w_{kj}^{out} = \lambda_w (t_k - a_k^{out}) a_j^{hid}$$

$$\Delta \alpha_i = -\lambda_{\dot{a}} \sum_j \left[\sum_k (t_k - a_k^{out}) w_{kj}\right] a_j^{hid} c |h_{ji} - a_i^{in}|$$

Exploring the Determinants of the Attentional Learning Trap Through Simulation

The computational ALCOVE-RL model makes it possible to test through simulation whether the attentional learning trap described above can emerge through learning with choice-contingent feedback, and whether ALCOVE's selective attention mechanism is in fact crucial for the trap to develop. Successful production of a learning trap through simulation will set the stage for studies with human participants.

To explore the behavior of ALCOVE-RL across a range of conditions, we devised a task similar to Figure 1. In particular, we presented the model with a four-feature category learning problem, where approaching prospects (i.e., exemplars) with both Features 1 and 2 yielded a payoff of -4 and approaching any other prospect yielded a payoff of 1. This environment matches exactly the structure depicted in Figure 1, but with two added irrelevant features. Each run of the model lasted for a 20-block learning phase of 16 trials each. Within each block all prospects were observed in a pseudorandom order, with the condition that each subblock of eight trials included two negative prospects and six positive prospects. The learning phase was followed by a final test block in which ALCOVE-RL made choices but underwent no learning updates.

We simulated five distinct conditions in order to understand what properties of model and of the environment were sufficient to produce an attentional learning trap. The <contingent, att> condition instantiated the kind of agent and environment described in our introduction to the learning trap. In this condition, the model's attentional learning capability was active, and the model only received feedback on the value of a prospect when it approached. In the <full-info, att> condition, we tested whether partial feedback was critical to the learning trap by modifying the environment so that the model received feedback on the value of all prospects regardless of its choice (i.e., foregone rewards, see Love & Otto, 2010). The <contingent, no-att> and <full-info, no-att>

conditions tested the role of selective attention in the learning trap. These conditions mirrored the first two, but with the attention to all dimensions yoked to remain equal. Finally, in the *<random-info, att>* condition we tested whether a learning trap might be produced not because an agent receives choice-contingent feedback but simply because it receives *less* feedback. In this condition the model received feedback on 50% of trials, but the feedback trials were randomly selected and independent of the model's choices.

Each condition was simulated for 1,000 model runs. All simulations were run with a specificity constant $c = 6$, a temperature parameter $\phi = 15$, an output-weight learning rate $\lambda_w = 0.1$, and an attention learning rate $\lambda_a = 0.1$. These parameters were selected to create an agent with a high propensity to fall into learning traps. In particular, the high λ_w and λ_a cause the agent to form beliefs quickly, while the high ϕ causes the agent to act fairly deterministically based on those beliefs, reducing the amount of corrective feedback it can receive. Thus, the results of the simulation should not be interpreted as a claim about the degree to which ALCOVE-RL will fall into an attentional learning trap under all possible parameter settings, but rather as an exploration of how one particular model with one parameter setting behaves as we alter the feedback from the environment and its ability to deploy selective attention.

Results

Figure 2 shows a breakdown of ALCOVE-RL's behavior over 20 blocks of experiential learning and the final test block. Models in each condition were classified based on whether or not they perfectly followed the correct two-dimensional rule or one of the two one-dimensional rules indicative of an attentional learning trap.

Over a quarter of agents in the *<contingent, att>* condition fell into the learning trap, with their behaviors influenced by only one of the relevant dimensions. In all other conditions, there was little tendency to learn an incorrect, unidimensional rule, and the correct rule tended to be learned quickly even in cases where an incorrect pattern of behavior was temporarily adopted. Providing full feedback in the *<full-info, att>* condition allowed for very swift learning, while maintaining contingent feedback but removing selective attention in the *<contingent, no-att>* condition lead to accurate learning as well. Interestingly, when given full feedback the model learned more quickly with selective attention than without it, showing that selective attention is only detrimental to learning this category structure when feedback is contingent.

It is also not the case that the bias of the *<contingent, att>* condition was caused by a general lack of information. In the *<random-info, att>* condition, where the model received feedback on half of the trials, randomly chosen, learning proceeded quickly and without bias. Thus, the attentional learning trap occurred due not to an overall poverty of information but to a specific pattern of behavior that prevented information from being gained about prospects that could correct the model's misallocated attention.

Finally, it is worth noting that in our simulations thus far we have assumed that nothing is encoded in the absence of feedback. As we have shown, learning traps can develop simply through the interplay of generalization and exploratory choice. However, data from recent experiments—among the few to study category learning with choice-contingent feedback—have suggested that when people do not experience feedback from a prospect they employ *constructivist* coding. In constructivist coding, rather than learning nothing upon avoiding an exemplar, the learning agent updates its beliefs as though it had approached the prospect and received the expected or predicted (likely negative) outcome, reinforcing existing beliefs with additional learning (Elwin, Juslin, Olsson, & Enkvist, 2007; Henriksson, Elwin, & Juslin, 2010). With this coding scheme, we would expect learning traps to become more intense, as the negative belief that lead a prospect to be avoided would be strengthened with each repeated avoidance. To test this, we simulated a version of ALCOVE-RL in which upon avoiding a prospect the model updated its beliefs as though it had approached and received a small negative outcome of −0.5. The results of this *<constructivist, att>* condition, plotted against the *<contingent, att>* condition in Figure 3, confirm that even a small amount
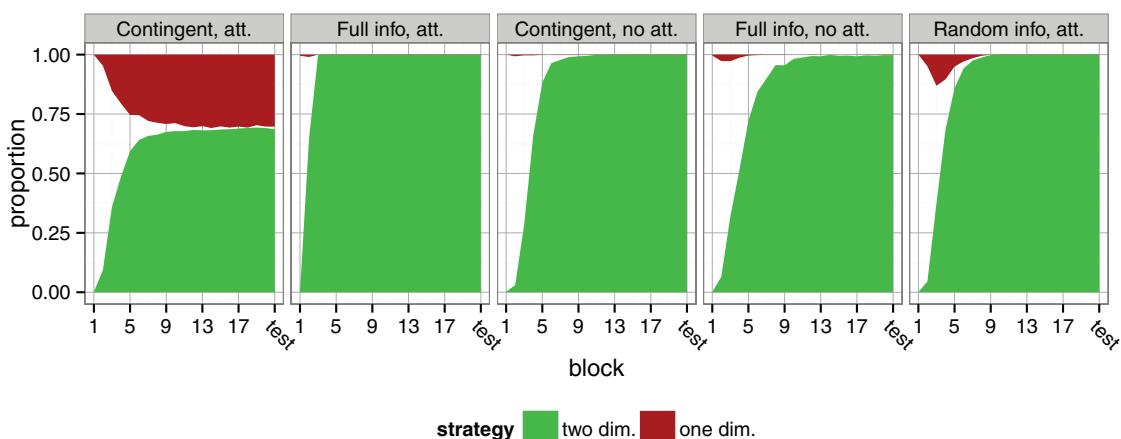

Figure 2. ALCOVE-RL model simulations of approach–avoid decision making in five attentional/informational conditions. Panels show the proportion of model runs adopting the correct two-dimensional strategy or one of the one-dimensional learning traps in each of the 20 learning blocks and the test block. See the online article for the color version of this figure.
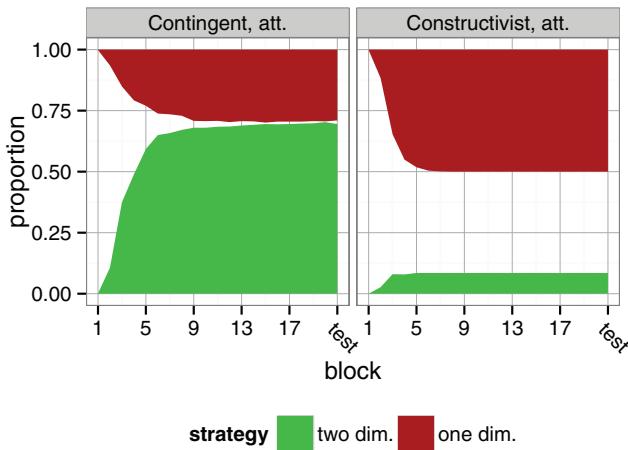
Figure 3. ALCOVE-RL model simulations with constructivist coding in the absence of feedback, compared with the standard contingent-information condition. Panels show the proportion of model runs adopting the correct two-dimensional strategy or one of the one-dimensional learning traps in each of the 20 learning blocks and the test block. Compared with the standard model, constructivist coding causes the model to fall into the attentional learning trap far more often. See the online article for the color version of this figure.

of constructivist coding leads to a more pronounced learning trap. (A stronger constructivist coding scheme in which avoided items are assumed to have produced the full -4 negative outcome causes the model to avoid all prospects entirely.) To preview our experimental results, the more severe learning trap caused by constructivist coding is in line with the behavior we observed among our human participants.

Experiment 1

Our simulations verify the pernicious nature of learning traps and the necessary aspects of the learner–environment interaction that encourage their development. To test the degree to which people are susceptible to the attentional learning trap, we performed a simple experiment similar to the category-learning task described above.[2] Given the broad support for ALCOVE in the category learning literature we expected that a similar learning trap would affect human learners. However, it is possible that human learners have more sophisticated learning strategies that help them avoid these types of costly errors when adapting their behavior to a task environment.

Method

Participants. One hundred one participants (48 female; 53 male) were recruited via Amazon Mechanical Turk. Preliminary power calculations indicated that 50 participants per condition would allow us to consistently detect a difference of 30% or greater of participants falling into the learning trap. Participants received $1.25 for participation and received a performance-based bonus that ranged up to $1.80. Two participants indicated that they used an external memory device (e.g., pen and paper) during the task, and four required more than two attempts to pass a postin-

structions quiz. These participants were excluded from further analyses.

Stimuli. Stimuli were computer-generated cartoon bees that varied on four binary dimensions; they had two or six legs, a striped or spotted body, single or double wings, and antennae or no antennae, for a total of 16 unique stimuli. Example stimuli are shown in Figure 4. Two of the four dimensions were chosen as relevant, counterbalanced across participants. Of the four possible combinations of values on these two dimensions, one was chosen at random; stimuli with this combination of values were "dangerous," and the remaining stimuli were "friendly."

Procedure. The experiment resembled a standard category-learning paradigm (e.g., Nosofsky et al., 1994), but with an added component of approach–avoid decision making. Participants played the role of a beekeeper collecting honey from several beehives. They were told that each hive contained a single variety of bees, and that while most hives contained friendly bees that would give them honey, some hives had been invaded by dangerous bees that would sting them if they tried to harvest.

On each trial, participants visited a new beehive, and were shown one of the bees in the hive. Based on the bee's appearance, they then had to choose either to attempt to harvest honey from the bee variety in that hive or to avoid the hive. When participants chose to harvest, they received honey and added $0.02 to their bonus if the bee variety was friendly, but were stung and lost $0.10 from their bonus if it was dangerous. When participants chose to avoid a hive, they gained $0.00. Participants started the game with a bonus of $0.40.

In the learning phase, participants encountered each of the 16 bee varieties four times, for a total of 64 trials. They were informed of the number of trials, and the number of remaining trials was displayed throughout learning. While trials were not overtly separated into blocks, each of the 16 bee varieties was encountered in each block of 16 trials. Within a block, stimuli were randomized with the condition that every eight stimuli contained two dangerous and six friendly bee varieties.

Participants were split into two conditions, which differed in the feedback received upon avoiding a beehive in the learning phase. In the contingent condition, no feedback was provided when a participant avoided a hive. In the full-information condition, participants were informed of whether the bee variety was friendly or dangerous and of what their payoff *would have been* had they harvested the hive.

The learning phase was followed by a 32-trial surprise test phase. During the test phase, participants encountered each variety twice and chose to harvest or avoid hives as before, but received no feedback about the outcomes of their actions and were not able to see changes to their bonus. Stimuli were ordered using the same randomization procedure as the learning phase. This phase provided a comparison of learning under equivalent conditions.

After the test phase, participants were informed of their total bonus, and were asked two final questions: "About what percent-

---

[2] Experiment and analysis code for all experiments is available at https://github.com/NYUCCL/LearningTrap. Data from all experiments is available at https://osf.io/hrb3u/?view_only=d374ae90615d4e5993ee1c502493673d. All experiments were approved by the NYU Institutional Review Board (IRB-FY2016-231 - Active Learning in Dynamic Task Environments).
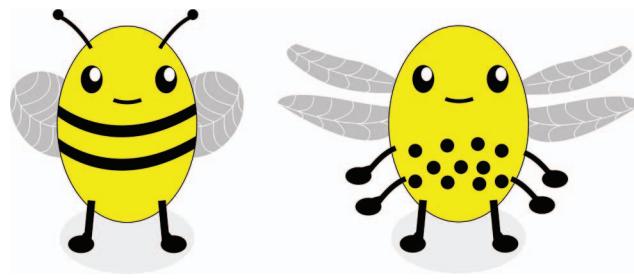
Figure 4. Examples of the stimuli used in Experiment 1, with opposite values on all four binary dimensions. See the online article for the color version of this figure.

age of beehives do you think contained dangerous bees?" and "Which features do you think were useful in deciding whether a bee variety was friendly or dangerous?" For the first question, participants entered a percentage between 0 and 100, and for the second question participants could choose any combination of the four features using check boxes.

Results

After exclusions, there were 48 participants in the continent-information and 47 participants in the full-information condition. Figure 5 shows participants' behavior over the four blocks of learning and during the test phase. While all statistical analyses use continuous measures of performance, for the purpose of visualization we categorized participants in each block based on whether they followed the correct two-dimensional rule or a one-dimensional rule indicative of a learning trap. Unlike in our model simulations, where we required perfect adherence to a rule for a model to be given a classification, we classified participants as following a given rule if 15 of 16 choices for a block adhered to the rule (30 of 32 choices for the test phase). Alternative cutoff thresholds yield qualitatively similar results.

We statistically compared the conditions using a Bayesian parameter estimation approach (A. Gelman et al., 2013). For Experiment 1, we used Bayesian equivalents of standard two-sample t and z tests. For continuous measures we modeled the data from the two conditions as being drawn from independent normal distributions with unknown mean and standard deviation. We gave the condition means weakly informative priors of *Normal* (.5, 1), and the condition standard deviations priors of *Normal* (.0, 1) truncated at zero. For binary measures we adopted a beta-binomial model in which each condition's proportion was drawn independently from a *Beta* (2, 2) prior, and the data was modeled as Bernoulli trials given the condition's proportion.

All models were implemented and fit using the Stan modeling language (Carpenter et al., 2017), which performs Bayesian inference using Hamiltonian Monte Carlo sampling. For each model we ran four independent chains of Monte Carlo sampling for 10,000 samples, the first 5,000 samples of which we discarded as "burn-in." Model convergence was confirmed using Stan's built-in $\hat{R}$ statistic.

For comparisons between conditions we report 95% posterior credible intervals (*CI*) for the difference between the conditions. Credible intervals that exclude zero can be interpreted as indicating high confidence that two conditions differ.

Early in the task, participants had little information and had to accrue experience to behave effectively later. In the first block of learning, participants in the contingent-information condition approached prospects on 75% of trials, while those the full-information condition approached only 58% of the time, CI [.07, .22]. This suggests participants valued the information that was gained by approaching, in line with other recent findings that people are information-seeking in simple decision-making tasks (Rich & Gureckis, 2017; Speekenbrink & Konstantinidis, 2015; Wilson, Geana, White, Ludvig, & Cohen, 2014).

As Figure 5 shows, participants increasingly adopted either the correct strategy or an inferior one-dimensional strategy as learning progressed. In the contingent-information condition, where no feedback was received about avoided prospects, participants tended toward one-dimensional strategies. In the full-information condition, they tended toward the correct two-dimensional strategy. Interestingly, the degree of dissimilarity between full-information and contingent-information behavior, and degree to which people fall into the learning trap, are both greater than in our simulations of the ALCOVE-RL model. While our explorations of ALCOVE-RL were intended to refine our understanding of the learning trap, and not to quantitatively fit human behavior, this hints that some aspects of people's behavior in the task, are not captured by the standard ALCOVE-RL model. Intriguingly, the simulations of ALCOVE-RL with constructivist coding appear qualitatively closer to people's behavior in both the current experiment and Experiment 2.

To create a numerical measure of this divergence of behavior, we calculated two behavioral "scores." The two-dimensional (2D) score denoted the proportion of a participant's choices that were consistent with the true, 2D task structure, and was equivalent to proportion correct choices. The one-dimensional (1D)
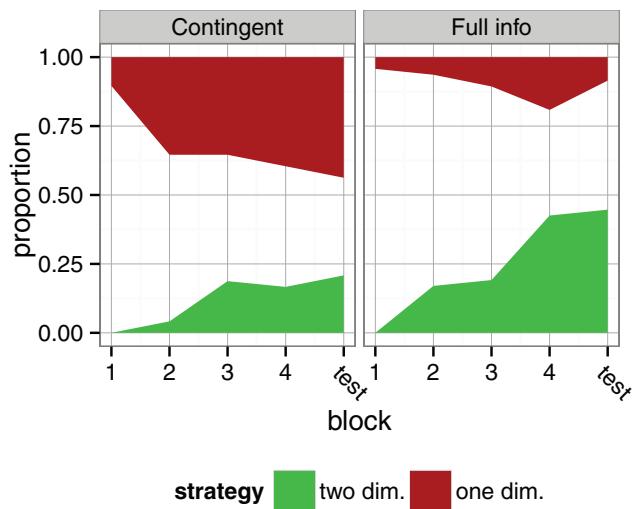

Figure 5. Participant behavior in the Experiment 1 beehive decision-making game. Panels show proportion of participants adopting the correct two-dimensional strategy or one of the one-dimensional learning traps in each of the four learning blocks and in the test phase. Participants were coded as using a two-dimensional or one-dimensional strategy if at least 15/16 of their choices in a block were consistent with that strategy. See the online article for the color version of this figure.

score denoted the proportion of participant's choices that were consistent with an attentional learning trap. This score was calculated by finding the proportion of responses that were consistent with each of the two 1D rules that participants might form on the relevant task dimensions, and then taking the maximum over these two proportions. A value of 1 on the 2D or 1D score indicates that participants followed the respective rule perfectly, and a value of .5 is expected if participants behaved randomly. If participants followed a 2D rule perfectly they would receive a 1D score of .75, and vice versa.

The Bayesian model posteriors plotted in the upper panels of Figure 6 confirm the qualitative results in Figure 5. In the test phase, participants in the contingent-information condition had an average 1D score of .83, while those in the full-information condition had an average score of .75, $CI = [.03, .14]$. In contrast, participants in the contingent-information condition had an average 2D score of .72, while those in the full-information condition had a higher score of .82, $CI = [.02, .18]$.

This tendency to fall into a learning trap in the contingent-information but not full-information condition extended, albeit less clearly, to the explicit posttask questions, as seen in the lower panels of Figure 6. Participants in the contingent condition responded on average that 37.6% of prospects were bad (one par-

ticipant was excluded for providing a negative response), while participants in the full-info condition responded that only 28.2% were bad, $CI = [.02, .16]$. The true proportion was 25%. This supports the conjecture that action-dependent feedback can affect a person's beliefs about the environment, and is consistent with the findings of Fazio et al. (2004) that approach–avoid learning leads to the belief that the environment is more negative than reality. In addition, only 22.9% of participants in the contingent-information condition identified the right combination of relevant features, while 40.4% of participants in the full-info condition did so, though the true difference is plausibly zero, $CI = [−.02, .33]$. Contingent-information participants identified only one of the relevant dimensions (and no irrelevant ones) in 37.5% of cases, while full-information participants did so only 25.5% of the time, although this difference also did not lie outside the 95% credible interval, $CI = [.07, .29]$.

In summary, Experiment 1 provided preliminary evidence that people are susceptible to an attentional learning trap that emerges in cases of choice-contingent feedback. In the following sections we test the robustness and generality of the learning trap, while also exploring ways to prevent people from entering the trap.
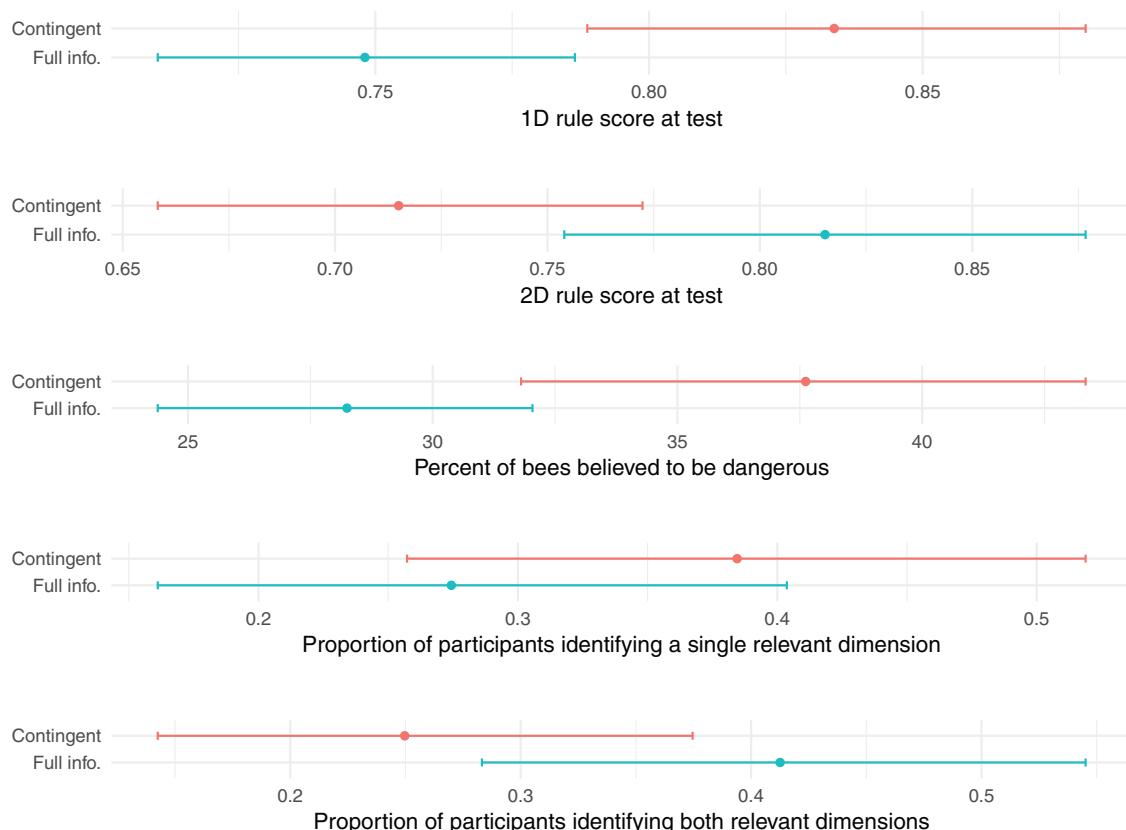

Figure 6. Comparisons of several measures of behavior between the contingent-information and full-information conditions of Experiment 1. Points indicate posterior population mean from Bayesian inference, and error bars indicate 95% credible intervals. All measures support the conclusion that participants with contingent information fell into the attentional learning trap more readily than those with full information. See the online article for the color version of this figure.

Testing the Robustness of the Attentional Learning Trap

In Experiment 2, we introduced three interventions that we hypothesized might affect the severity of the attentional learning trap. Our goals in testing these interventions were twofold: first, to test whether the attentional learning trap is robust over a variety of stimulus conditions, and second, to investigate whether any of these interventions may prove effective at preventing the learning trap, and perhaps serve as a prototype for interventions in applied settings.

As described in the introduction, learning traps can have harmful consequences in a remarkably wide range of domains, making it important to study methods of diminishing them. To provide one example where the attentional learning trap could have harmful effects, and to preview the cover story for Experiment 2, consider a company hiring new employees. If there are multiple attributes, each of which is sufficient to make an applicant suitable for the job, the fact that learning an applicant's true value is contingent on their being hired may mean that only one or a subset of these factors is ever attended to and learned. This can obviously have a negative effect on the company's success. Just as importantly, if the company learns to attend excessively to a feature like college attendance that covaries with socioeconomic variables, it can have negative long-run societal effects as well.

Because the role of exploration in the development of learning traps is relatively well established, while the role of generalization has rarely been studied, we focused our three interventions on slowing generalization from past experience to novel prospects, rather than on directly increasing exploration. While generalization is a vital aspect of intelligent behavior, and reducing generalization is certainly not always beneficial, we hypothesized that decreasing the speed of generalization would cause people to sample prospects more exhaustively before beginning to avoid those prospects they believed to be negative, thus reducing the likelihood of falling into the attentional learning trap.

In the following sections, we first describe the intuitions behind each intervention, explore them computationally using ALCOVE-RL, and then test them experimentally with people.

Individuating Prospects

One clear way to decrease generalization and potentially limit the attentional learning trap is to make stimuli increasingly distinct and idiosyncratic. When stimuli are more distinctive, people tend to treat them more as individuals and show increased ability to learn identification compared with categorization. While identification learning is more difficult than categorization with generic artificial stimuli (Love et al., 2004; Shepard et al., 1961), Medin, Dewey, and Murphy (1983) found that people were more easily able to pair unique first names than categorical last names with photographs of faces. Love et al. (2004) argued that this phenomenon could be accounted for with the SUSTAIN model of categorization by assuming that the faces had many distinctive features beyond those manipulated by the experimenters, which decreased the similarity among stimuli and thus increased the odds of representing each stimulus individually.

In an approach–avoid decision-making task, increased individuation of stimuli should make a person less likely to generalize information gained from experience with one prospect to decisions about another. Attention paid to idiosyncratic features will slow the biasing of attention toward a single dimension, giving the person more opportunity to explore a variety of stimuli and learn the true structure of the environment. Essentially, increased individuation of prospects shifts the task away from category-learning, and toward learning about whether to approach individual prospects.

Occluding Feature Information

A second approach to decreasing the attentional learning trap may be to restrict information by randomly occluding some features of a prospect such that the decision maker cannot observe their values. While this intervention could of course impair a person's decision-making ability, it could actually improve performance in the long run by causing a greater spread of attention. E. G. Taylor and Ross (2009) found that participants learned more about nondiagnostic features in a category-learning task when features were randomly occluded, and hypothesized that feature occlusion discourages rapid narrowing of selective attention and promotes a broader attentional strategy. In the context of approach decisions, if a person is attending strongly to a dimension that is occluded, he or she may be forced to use other features, which may lead to the discovery that they are relevant. Even when the favored features is not occluded, the possibility of their future absence may cause people to be less quick to rely solely on one feature.

Increasing Noise

Stochasticity in prospect outcomes is the driving force behind the hot stove effect and underweighting of rare events; without it, experience is never misleading, and there is no possibility for an incorrect belief about a prospect to develop. When negative beliefs formed about one prospect can generalize to another prospect, rather than simply to the same prospect at a later time, stochasticity is no longer required for biased behavior to result. Instead, it is plausible that a small degree of noise might aid the learning process in the long term. Noise is used in optimization algorithms such as simulated annealing (Kirkpatrick, Gelatt, & Vecchi, 1983) to overcome learning-trap-like local minima. Gureckis and Love (2009) found that noise improved human performance in a dynamic decision-making task that required people to discover a nonobvious solution to a problem. A noisy outcome, while potentially triggering false belief in the region the experienced prospect, might also cause a reallocation of attention that could have globally beneficial consequences. Such an experience might cause a nonattended but useful dimension to attract attention, setting the agent on a new trajectory of behavior and learning and pulling it out of a learning trap.

Modeling Debiasing Interventions

To test the potential effect of these interventions on generalization and behavior, we performed model simulations comparing them to the basic contingent-information condition. To modify the model for the individuation condition, we added an extra dimension with 16 nominal values representing idiosyncratic features of each stimulus (following Love et al., 2004). For the occluded-

dimension condition, on 25% of trials a randomly chosen dimension was masked such that it did not contribute to the model's network activation and its attention weight was not updated. For the noisy condition, on 5% of trials the potential outcome was flipped from the positive outcome to the negative outcome or vice versa.

The models were simulated for a learning phase of 20 blocks, followed by a one-block test phase with no learning updates, where no dimensions were masked in the occluded-dimension condition and the individuating dimension was masked in the individuation condition. Results are plotted in Figure 7.

In the test phase, all three interventions lead to a greater proportion of the models adopting the correct two-dimensional rule. Interestingly, in early blocks the occlusion and increased noise interventions inhibit learning; while the model is less likely to enter the learning trap, it is also slower to learn the correct strategy. This indicates that these interventions represent a trade-off, allowing (potentially) superior long-term performance but at the cost of worse short-term performance.

Experiment 2

In Experiment 2, we tested whether individuating prospects, occluding feature information, or adding noisy outcomes would affect the degree to which participants fell into the attentional learning trap. In addition, to further test the generality of the learning trap we introduced a more life-like "job application" cover story, lengthened the training phase, and reduced the relative penalty for approaching negative prospects.

Prior to conducting Experiment 2, we conducted two large pilot experiments investigating different ways of implementing the interventions (see Intervention Pilot Experiments in the supplemental material available online). These pilot experiments showed the learning trap to be robust and did not show the interventions to be effective, foreshadowing the results of Experiment 2. Experiment 2 thus represents a final, large-n effort to replicate the learning trap and document the degree of effectiveness of the interventions.

Method

Participants. Four hundred participants (176 female; 220 male) were recruited via Amazon Mechanical Turk. A power analysis showed that a sample size of about 80 participants per condition would allow us to reliably detect a 25% difference in percentage of people falling into the learning trap, or on continuous measures an effect size of .45. Participants received $2.00 for participation and received a performance-based bonus that ranged up to $1.68. Forty participants were excluded for requiring more than two attempts to pass a postinstructions quiz.

Stimuli. Stimuli were fake job applications that varied on four binary dimensions. Applicants had a "degree" in "business" or "economics", a "past employer" of either "Hudson Inc." or "Nile Co.;" a "skill" in either "computer programming" or "graphics editing", and a "past position" of either "product development" or "market research," for a total of 16 unique stimuli. Example stimuli are shown in Figure 8. Two of the four dimensions were chosen as relevant, counterbalanced across participants. Of the four possible combinations of values on these two dimensions, one was chosen at random; stimuli with this combination of values were "unsuitable" applicants, while the remaining stimuli were "suitable."

Procedure. Experiment 2's procedure is similar to that of Experiment 1. In this experiment, participants played the role of a recruiter considering a series of job applications. They were instructed that their goal was to generate revenue for their company.

In the learning phase, participants encountered each of the 16 unique stimuli eight times, for a total of 128 trials. The number of applications (i.e., trials) remaining was displayed throughout the learning phase. As in Experiment 1, stimuli were ordered such that each block of 16 contained all 16 stimuli, and each subblock of eight contained two negative and six positive stimuli, with stimulus order otherwise randomized.

On each trial, participants were presented with an application. The application started out blank, and participants had to press the space bar four times to reveal each of the four dimensions in a
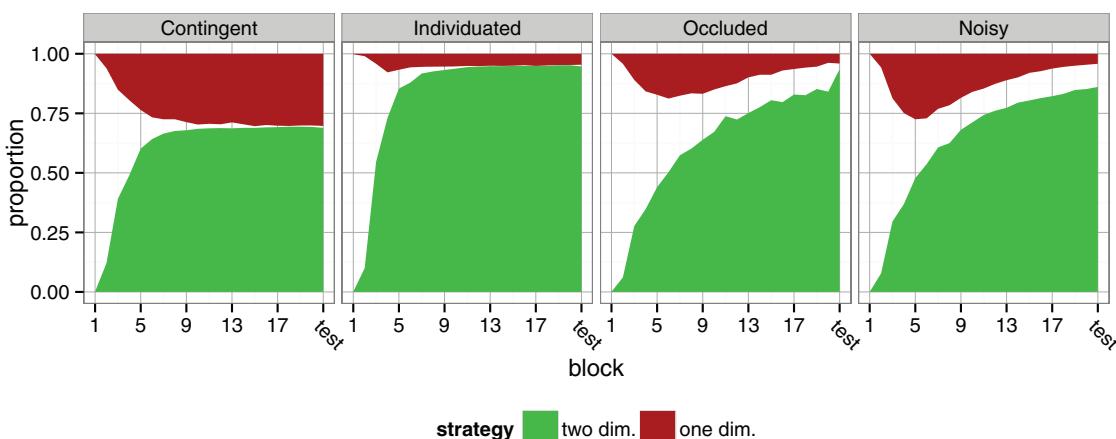

Figure 7. ALCOVE-RL model simulations of approach–avoid decision making in a basic contingent-information setting, and with three interventions. Panels show the proportion of model runs adopting the correct two-dimensional strategy or one of the one-dimensional learning traps in each of the 20 learning blocks and the test block. See the online article for the color version of this figure.

Figure 8. Examples of the stimuli used in Experiment 2. From left: an example stimulus from the contingent-information or full-information conditions, two example stimuli from the individuated condition, and an example stimulus from the occluded condition. See the online article for the color version of this figure.

random order. This was done to reduce any bias toward attending to dimensions near the top of the application. The participant then had to choose whether to accept or reject the application. Accepting a suitable applicant generated revenue of $1,000 for the company, while accepting an unsuitable applicant caused a loss of $3,000. Rejecting an applicant caused no change in revenue. Revenue began at $50,000, and was converted to a cash bonus at the rate of $0.01 per $1,000.

Participants were split into five conditions. In the full-information condition, participants who rejected an applicant were informed of whether the applicant *would have been* suitable or unsuitable, and how the company's revenue would have changed. In the contingent condition, participants were given no feedback upon rejecting an applicant.

In the three intervention conditions, feedback during the learning phase was contingent as in the contingent condition. However, as shown in Figure 8, the stimuli were modified in ways hypothesized to reduce the learning trap. In the individuation condition, participants were instructed that the application system had a feature that assigned each of the 16 unique dimension value combinations a random unique icon to help them keep track of what they had observed. These icons were small pictures of animals that on some trials were displayed below the four dimensions, as shown in Figure 8. The icons were shown on 90% of trials in the first block of learning, which gradually decreased to 0% by the last block, so that participants could not rely solely on the icons.

In the occluded condition, on some trials one of the four dimensions was chosen at random and covered with a black bar (see Figure 8). If the participant hired the applicant, the hidden dimension was then revealed. This intervention was applied to 50% of trials in the first five blocks of training, and then was removed for the last three blocks.

In the noisy condition, applicant outcomes were changed from suitable to unsuitable or unsuitable to suitable on some randomly selected trials. Participants were informed that participants who appear suitable might occasionally be unsuitable, and vice versa. This intervention was applied to 10% of trials in the first five blocks of training, and then was removed for the last three blocks.

The exact design of these interventions was informed by the data collected in our pilot experiments (described in the supplement). However, determining the exact difference in effect between subtly different interventions would require prohibitively large sample sizes, and the interventions were thus also guided by our own intuitions.

In all conditions, the learning phase was followed by a surprise 32-trial test phase, using the same randomization procedure as the learning phase. Participants chose to accept or reject as before, but received no feedback about the outcomes of their actions and were not able to see changes to revenue. Both interventions were also removed during the test phase so that it was equivalent across all five conditions.

After the test phase, participants were informed of their total bonus. As in Experiment 1, they were asked "About what percentage of applicants do you think were unsuitable?" and "Which fields do you think were useful in deciding whether an applicant was suitable or unsuitable?" We also added a third posttask question, which asked participants whether they believed they had learned completely how to use the applicant features to determine which applicants were suitable. Participants chose from a drop-down list either "I think I learned completely how the features determined suitability," "I think there may have been aspects of applicant suitability that I did not learn," or "I think there were definitely aspects of applicant suitability that I did not learn."

Results

After exclusions, there were 73 participants in the contingent-information condition, 66 in the full-information condition, 72 in the individuation condition, 70 in the occluded condition and 79 in the noisy condition. None of the experiment's results are qualitatively affected when excluded participants are included. Figure 9 shows participants' behavior over the eight blocks of learning and the test phase, using the same threshold for classifying participants as following a 2D or 1D strategy described in Experiment 1. We report first on the replication of the contingent-information and full-information conditions, followed by the results of the interventions.

Experiment 2 included five conditions, increasing the risk of what from a frequentist perspective would be considered Type I errors due to multiple comparisons. To reduce this risk and improve the quality of our estimates, we adopted a Bayesian multi-level modeling approach to our analyses of Experiment 2, which assumes that the group means for each condition are drawn from the same overarching population distribution. This causes the group mean posteriors to be drawn toward each other during inference to a degree determined by the variability of the data, resulting in better estimates and fewer "false positives" (A. Gelman, Hill, & Yajima, 2012).
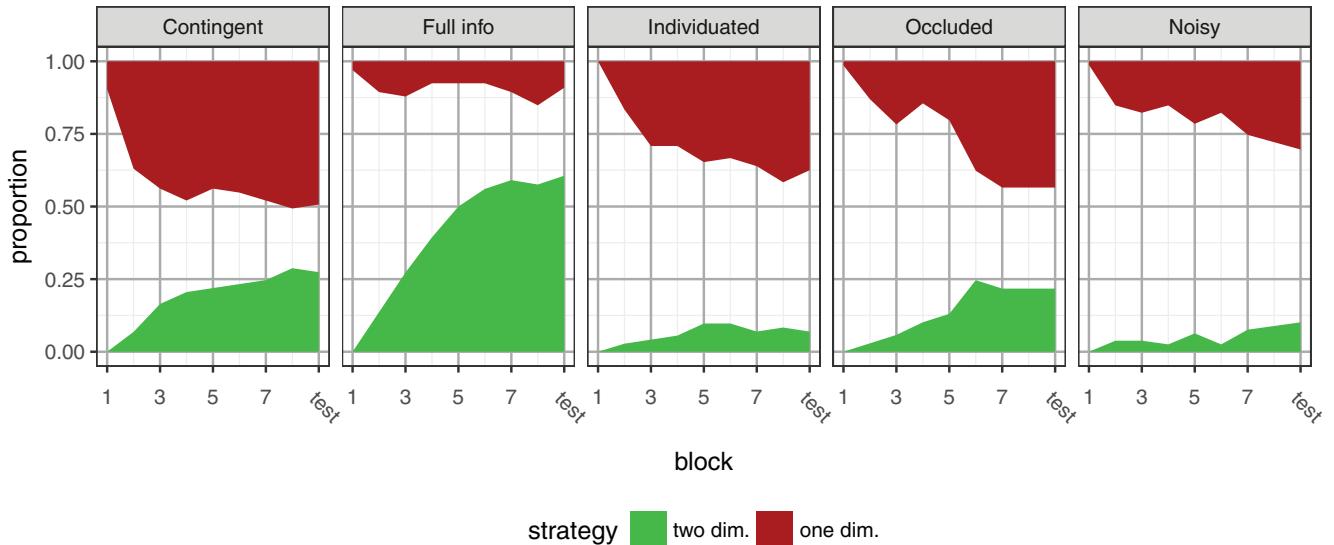
Figure 9. Participant behavior in the Experiment 2 job application decision-making game. Panels show proportion of participants adopting the correct two-dimensional strategy or one of the one-dimensional learning traps in each of the eight learning blocks and in the test phase. Participants were coded as using a two-dimensional or one-dimensional strategy if at least 15/16 of their choices in a block were consistent with that strategy. See the online article for the color version of this figure.

For continuous measures, rather than modeling the data as being drawn from *independent* normal distributions, we assumed that the condition means were themselves drawn from a population-level normal distribution with unknown mean and variance. We gave the population distribution mean a *Normal*(0.5, 1) prior, and the population distribution standard deviation a *Normal*(0, 0.5) distribution truncated at zero.

Similarly, for binary measures we posited that rather than the condition proportions being drawn independently, they were drawn from the same population-level beta distribution with unknown parameters. We reparameterized the standard $Beta(\alpha, \beta)$ distribution in terms of mean μ and precision κ, where $\alpha = \mu\kappa$ and $\beta = (1-\mu)\kappa$, and specified a $Beta(2, 2)$ prior on μ and a $Gamma(1, .1)$ prior on κ.

We again estimated all models using Stan, with four independent chains of 10,000 samples.

As in Experiment 1, participants with contingent information were highly likely to accept candidates in the first block of learning, accepting 71% of applicants. Participants with full information were again less likely to do so, accepting 64%, CI [.02, .12], and were less likely to accept candidates than participants in the intervention conditions as well, with all CI's excluding zero.

Measures of the degree to which participants fell into the learning trap across conditions are plotted in Figure 10. Comparing the contingent and full-information conditions, the results fully replicate Experiment 1. In the test phase, participants in the two conditions diverged in their tendencies to adopt an incorrect one-dimensional strategy or the correct 2D strategy. Full-information participants had an average 1D score of .75, while contingent-information participants averaged .86, CI [.04, .15]. The pattern for the 2D score was reversed; full-information participants reached an average 2D score of .87, while those in the contingent-information condition reached a lower score of .78, CI [.03, .14].

While 25% of applicants were in fact unsuitable, contingent-information participants on average estimated this percentage to be 36.5%. Full-information participants estimated the percentage to be 24.4%, much closer to the true value and lower than the contingent-information condition, CI [.07, .17]. Full-information participants were also more likely to correctly identify the two relevant dimensions, CI [.16, .46], while contingent-information were more likely to identify a single relevant dimension, CI [.27, .55].

Only around 10% of participants in the contingent and full-information conditions reported that there were "definitely" aspects of applicant suitability they did not learn, so we pooled these participants with those who reported that there "may" have been aspects they did not learn. In the contingent-information condition 45.2% of participants believed they "learned completely" how features determined suitability, while in the full-information condition 60.6% of participants believed they had learned completely. Full information may cause slightly more confident learning, but the credible interval for the difference contingent information does not exclude zero, CI [−.02, .29]. Of the contingent-information participants who thought they had fully learned the task, 57.6% (19 out of 33) had in fact fallen into the learning trap. This suggests that in many cases participants entered a learning trap not just because it yielded "good-enough" performance, even while suspecting it to be incomplete, but because they believed it to represent the true structure of the environment.

Over all, as seen in Figure 10, the three interventions did not prevent the learning trap, or "prevented" the learning trap only insofar as they prevented effective learning altogether.

Specifically, the occlusion intervention appeared to have little effect, creating no difference outside the 95% credible interval from the contingent condition on any measure. The individuation and noise interventions did make people less likely to adopt a 1D
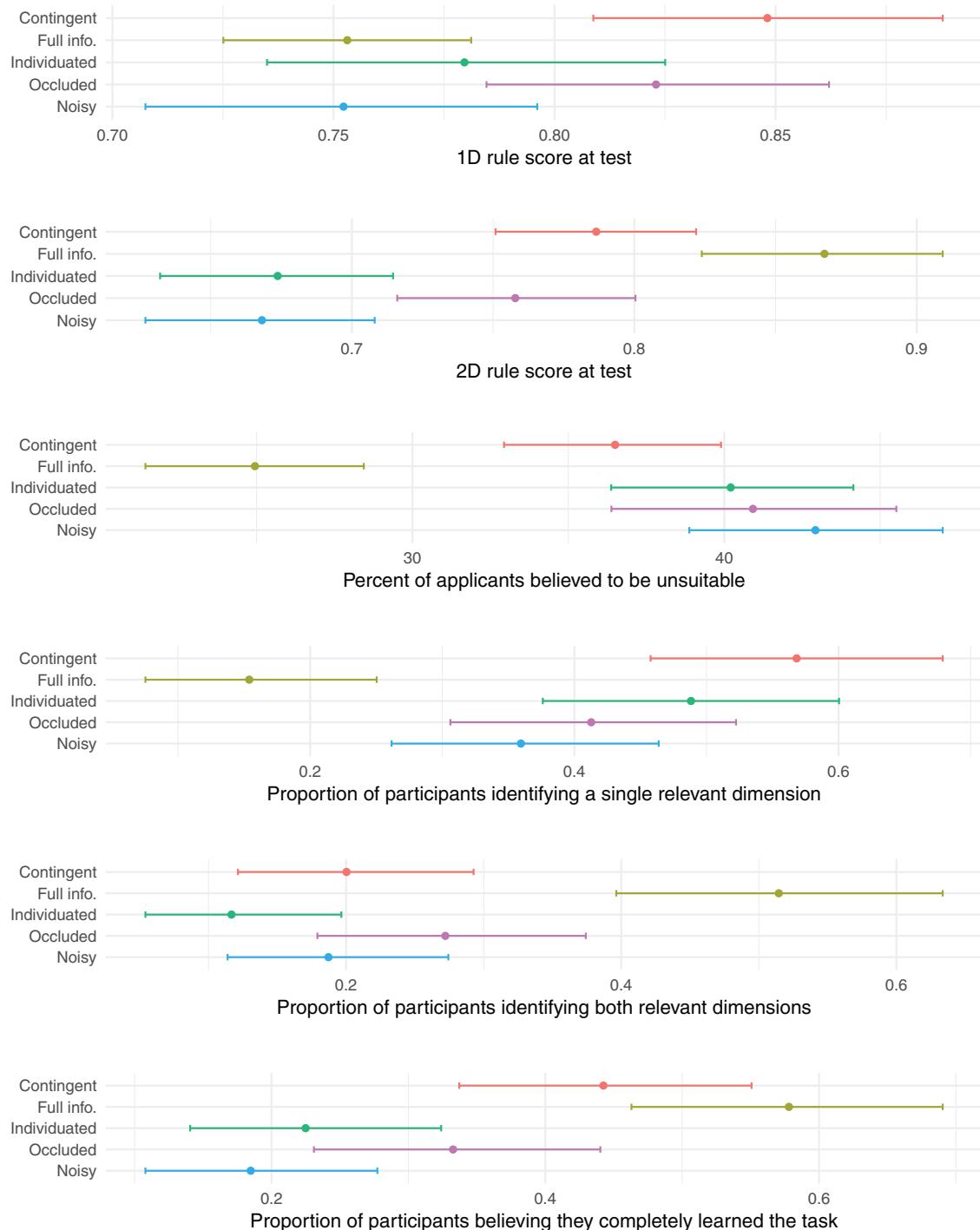
Figure 10. Comparisons of several measures of behavior across the five conditions of Experiment 2. Points indicate posterior population mean from Bayesian inference, and error bars indicate 95% credible intervals. All measures are support the conclusion that participants with contingent information fell into the attentional learning trap more readily than those with full information, but that the interventions did not lead to more robust learning compared with the contingent-information condition. See the online article for the color version of this figure.

rule, CI [.01, .13] and CI [.03, .16], respectively. Compared with contingent-information condition participant, people in these conditions had lower 1D scores during the test phase. However, this was a hollow victory in that participants in these conditions also both had lower 2D scores at test, CI [.05, .17] and CI [.06, .17]. In other words, participants who experienced these interventions had difficulty learning any stable structure of the task, correct or incorrect. This impression was reinforced when we asked participants whether they had learned the task completely. Participants in both the individuated and noisy conditions were less likely than those in the contingent-information condition to report that they had, CI [.07, .36] and CI [.12, .39].

Finally, compared with participants in the full-information condition, participants in all three intervention conditions had lower 2D scores during the test phase, were less likely to accurately identify the two relevant dimensions following the test, believed a greater proportion of applicants were unsuitable, and were less likely to report believing that they had learned completely how to determine applicant suitability, with the 95% credible intervals for all differences excluding zero. In other words, none of the interventions allowed people to obtain the accurate understanding of the environment possible through receiving full feedback following each choice.

General Discussion

Across two experiments and hundreds of participants (plus supplementary pilot experiments) we found evidence for a robust and novel learning trap. The attentional learning trap we describe is the joint product of the limits on exploration presented by choice-contingent feedback tasks and the inductive biases inherent in selective attention. Because the learning trap is related to the information restrictions of limited exploration, we found that it rarely occurred when people received full feedback regardless of their choices. Interestingly, the learning trap persisted when prospects were individuated, when prospect features were occasionally occluded, and when noise was added to prospect outcomes. Below, we address why these interventions did not prevent the learning trap and how learning traps might be diminished, as well as additional situations where we might expect learning traps to form.

The Potential for Reducing Learning Traps

The three interventions implemented in Experiments 2 did not, in any significant way, prevent participants from falling into the attentional learning trap. To the extent that they prevented rapid generalization across a single dimension, they also seemed to stop effective generalization entirely. That is, they prevented people from learning both the true, two-dimensional structure and the incomplete one-dimensional structure. This was evidenced both by participants' poor performance and their self-reports. While we believe future attempts at these sorts of interventions may prove more fruitful, our results do offer one possible lesson: decreasing generalization may be the wrong, or at least a risky, approach to reducing learning traps. As mentioned in the introduction, learning cannot occur without generalization (Mitchell, 1980). Thus, the challenge is to balance reduced generalization such that incorrect beliefs aren't quickly formed, but not so much that the correct beliefs are never formed.

An alternative approach is to directly increase exploration. Rather than slowing the process of learning, this approach seeks to increase people's willingness to continue enduring costs to gain information. The full-information conditions of both experiments showed that accurate learning is possible in our tasks given enough information, suggesting this approach could be effective.

Exploration might be increased in multiple ways. First, recent studies show that human exploration is sensitive to the future usefulness of gaining information (Rich & Gureckis, 2017; R. C. Wilson et al., 2014). Thus, increasing the salience or perceived number of future choices in an environment might increase exploration and decrease false beliefs. Second, exploration can be prompted by curiosity, which acts as an innate drive toward information-seeking (Berlyne, 1966; Kidd & Hayden, 2015; Loewenstein, 1994). This means that causing a greater intrinsic interest in a domain may be another tool for decreasing learning traps. Third, one might increase exploration by decreasing, at least temporarily, the drive for reward. Experiments with both animals (Tolman, 1948) and humans (Schwartz, 1982) have shown that excessive drive for external reward can prevent decision makers from learning the true structure of the world. If people can be given the opportunity to explore an environment in a disinterested manner, they will have a greater opportunity to build an accurate world model and will be less susceptible to learning traps.

Finally, keep in mind that the complete prevention of learning traps is often implausible and undesirable, because it requires excessive exploration at the cost of exploitation (Levinthal & March, 1993). If the primary goal is to maximize positive outcomes, rather than simply possess the most accurate beliefs about the world, it may make sense to allow the possibility of false beliefs (Le Mens & Denrell, 2011). Further research is needed to reduce biases in organizational and social behavior as much as possible, but some degree of bias may be an unavoidable byproduct of learning from experience.

The Possibilities of New Learning Traps

While the review in our article attempted to integrate many past phenomena that are interpretable as a "learning trap," one interesting question is if there are other situations where such phenomena may occur. A number of possibilities come to mind, all having to do with more complex forms of generalization than we considered so far. For instance, people represent some real-world categories in terms of taxonomic hierarchies. Within these hierarchies there is often a psychologically prioritized level of abstraction, the basic level, that tends to be most naturally used when naming an object (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Waxman (1990)). Dog, for example, is a basic level category for most people, while animal is a superordinate category and golden retriever is a subordinate category. Because people tend to generalize at the basic level, there is a risk that negative attributes from single experience will be learned for an entire basic-level category, even if it truly applied only to a subordinate category.

Categories are also often connected to naive theories and causal beliefs, which can shape a person's understanding of the category and its members (G. L. Murphy & Medin, 1985; Rehder & Hastie, 2001). One common kind of causal theory is that members of a category share a deep underlying cause, or *essence*, that gives rise to their other properties (S. A. Gelman, 2004; Medin & Ortony,

1989). Essentialist beliefs are observed for natural kinds and for social categories such as race and gender (Allport, 1954; Hirschfeld, 1995; M. G. Taylor, 1996; M. G. Taylor, Rhodes, & Gelman, 2009). Because the attributes of instances of essentialized categories are believed to be caused by a shared essence, people tend to generalize a property of one instance to other instances more for essentialized categories than for nonessentialized ones (S. A. Gelman, 1988; S. A. Gelman & Coley, 1990). This may make essentialized categories particularly susceptible to learning traps. In the domain of social categories, Denrell (2005) has described how the hot stove effect can lead to the development of negative perceptions of outgroups. Future work could examine whether essentialism plays a role in this type of social learning trap, and whether reductions in essentialist beliefs might help to prevent it.

Conclusion

Learning allows people to behave adaptively in a world that cannot be completely known a priori. But as we show in this report, learning processes cannot be relied on to converge steadily to true belief as a learner gains experience. When learning from experience influences reward-seeking choices, and reward-seeking choices produce the experience for further learning, the entire learning–choosing system can become stuck in patterns of poor decisions and false beliefs. In the words of Levinthal and March (1993), "learning has its own traps."

In this article, we have tried to clarify the link between choice processes and learning processes in the development of learning traps. We have shown how, in environments with explore–exploit trade-offs created by choice-contingent feedback, many apparent suboptimalities are natural consequences of people's inductive biases. These inductive biases, rather than being flaws in the learning system, are prerequisites for effective generalization (Mitchell, 1980).

Context of Research

This project emerged out of a broader family of projects examining how humans balance exploration and exploitation in complex environments. While some of our other studies have focused on how people use environmental cues to determine when to explore more or less (e.g., Rich & Gureckis, 2017), this project aimed to understand how patterns of exploration affect beliefs when the environment is more complex than a simple repeated choice task. In doing so, it also addressed a perceived gap in the categorization literature, which has rarely looked closely at the challenges of learning via choice-contingent feedback. We believe a more complete understanding of exploratory choice can be gained by unpacking the two-way interaction between exploration and beliefs about the environment, and hope to continue to investigate this relationship as outlined in the Discussion.

This project also intersects with active and self-directed learning, another research focus of our lab (T. M. Gureckis & Markant, 2012). While active control over the contents of learning often produces gains over passive learning, learning traps provide a counterpoint: If the learner can choose to avoid stimuli that are expected to be unpleasant, active learning may lead to worse learning outcomes than passive exposure.

References

Allport, G. W. (1954). *The nature of prejudice*. Garden City, NY: Double-day.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98,* 409–429. http://dx.doi.org/10.1037/0033-295X.98.3.409

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience, 10,* 1214–1221. http://dx.doi.org/10.1038/nn1954

Berlyne, D. E. (1966). Curiosity and exploration. *Science, 153,* 25–33. http://dx.doi.org/10.1126/science.153.3731.25

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., . . . Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*. Advance online publication. http://dx.doi.org/10.18637/jss.v076.i01

Chater, N., & Vitányi, P. (2003). Simplicity: A unifying principle in cognitive science? *Trends in Cognitive Sciences, 7,* 19–22. http://dx.doi.org/10.1016/S1364-6613(02)00005-0

Denrell, J. (2005). Why most people disapprove of me: Experience sampling in impression formation. *Psychological Review, 112,* 951–978. http://dx.doi.org/10.1037/0033-295X.112.4.951

Denrell, J. (2007). Adaptive learning and risk taking. *Psychological Review, 114,* 177–187. http://dx.doi.org/10.1037/0033-295X.114.1.177

Denrell, J., & March, J. G. (2001). Adaptation as information restriction: The hot stove effect. *Organization Science, 12,* 523–538. http://dx.doi.org/10.1287/orsc.12.5.523.10092

Eiser, J. R., Fazio, R. H., Stafford, T., & Prescott, T. J. (2003). Connectionist simulation of attitude learning: Asymmetries in the acquisition of positive and negative evaluations. *Personality and Social Psychology Bulletin, 29,* 1221–1235. http://dx.doi.org/10.1177/0146167203254605

Elwin, E., Juslin, P., Olsson, H., & Enkvist, T. (2007). Constructivist coding: Learning from selective feedback. *Psychological Science, 18,* 105–110. http://dx.doi.org/10.1111/j.1467–9280.2007.01856.x

Erev, I. (2014). Recommender systems and learning traps. In M. Ge & F. Ricci (Eds.), *Proceedings of the first international workshop on decision making and recommender systems* (pp. 38–41). Bolzano, Italy: Free University of Bozen-Bolzano.

Fazio, R. H., Eiser, J. R., & Shook, N. J. (2004). Attitude formation through exploration: Valence asymmetries. *Journal of Personality and Social Psychology, 87,* 293–311. http://dx.doi.org/10.1037/0022-3514.87.3.293

Feldman, J. (2003). The simplicity principle in human concept learning. *Current Directions in Psychological Science, 12,* 227–232. http://dx.doi.org/10.1046/j.0963-7214.2003.01267.x

Fuller, R. (1991). Behavior analysis and unsafe driving: Warning-learning trap ahead!. *Journal of Applied Behavior Analysis, 24,* 73–75. http://dx.doi.org/10.1901/jaba.1991.24-73

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. New York, NY: Taylor & Francis.

Gelman, A., Hill, J., & Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness, 5,* 189–211. http://dx.doi.org/10.1080/19345747.2011.618213

Gelman, S. A. (1988). The development of induction within natural kind and artifact categories. *Cognitive Psychology, 20,* 65–95. http://dx.doi.org/10.1016/0010-0285(88)90025-4

Gelman, S. A. (2004). Psychological essentialism in children. *Trends in Cognitive Sciences, 8,* 404–409. http://dx.doi.org/10.1016/j.tics.2004.07.001

Gelman, S. A., & Coley, J. D. (1990). The importance of knowing a dodo is a bird: Categories and inferences in 2-year-old children. *Developmental Psychology, 26,* 796–804. http://dx.doi.org/10.1037/0012-1649.26.5.796

Geman, S., Bienenstock, E., & Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation, 4,* 1–58.

Goodman, N. D., Tenenbaum, J. B., Feldman, J., & Griffiths, T. L. (2008). A rational analysis of rule-based concept learning. *Cognitive Science, 32,* 108–154. http://dx.doi.org/10.1080/03640210701802071

Gureckis, T. M., & Love, B. C. (2009). Learning in noise: Dynamic decision-making in a variable environment. *Journal of Mathematical Psychology, 53,* 180–193. http://dx.doi.org/10.1016/j.jmp.2009.02.004

Gureckis, T. M., & Markant, D. B. (2012). Self-directed learning: A cognitive and computational perspective. *Perspectives on Psychological Science, 7,* 464–481. http://dx.doi.org/10.1177/1745691612454304

Henriksson, M. P., Elwin, E., & Juslin, P. (2010). What is coded into memory in the absence of outcome feedback? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 1–16. http://dx.doi.org/10.1037/a0017893

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science, 15,* 534–539. http://dx.doi.org/10.1111/j.0956-7976.2004.00715.x

Hirschfeld, L. A. (1995). Do children have a theory of race? *Cognition, 54,* 209–252. http://dx.doi.org/10.1016/0010-0277(95)91425-R

Hoffman, A. B., & Rehder, B. (2010). The costs of supervised classification: The effect of learning task on conceptual flexibility. *Journal of Experimental Psychology: General, 139,* 319–340. http://dx.doi.org/10.1037/a0019042

Jones, M., & Cañas, F. (2010). Integrating reinforcement learning with models of representation learning. In S. Ohlsson & R. Catrambone (Eds.), In *Proceedings of the 35th annual meeting of the cognitive science society* (pp. 1258–1263). Austin, TX: Cognitive Science Society.

Juni, M. Z., Gureckis, T. M., & Maloney, L. T. (2016). Information sampling behavior with explicit sampling costs. *Decision, 3,* 147–168. http://dx.doi.org/10.1037/dec0000045

Kidd, C., & Hayden, B. Y. (2015). The psychology and neuroscience of curiosity. *Neuron, 88,* 449–460. http://dx.doi.org/10.1016/j.neuron.2015.09.010

Kirkpatrick, S., Gelatt, C. D., Jr., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science, 220,* 671–680. http://dx.doi.org/10.1126/science.220.4598.671

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. (1996). Dimensional relevance shifts in category learning. *Connection Science, 8,* 225–248. http://dx.doi.org/10.1080/095400996116893

Kruschke, J. K., & Blair, N. J. (2000). Blocking and backward blocking involve learned inattention. *Psychonomic Bulletin & Review, 7,* 636–645. http://dx.doi.org/10.3758/BF03213001

Le Mens, G., & Denrell, J. (2011). Rational learning and information sampling: On the "naivety" assumption in sampling explanations of judgment biases. *Psychological Review, 118,* 379–392. http://dx.doi.org/10.1037/a0023010

Le Mens, G., Kareev, Y., & Avrahami, J. (2016). The evaluative advantage of novel alternatives: An information-sampling account. *Psychological Science, 27,* 161–168. http://dx.doi.org/10.1177/0956797615615581

Levinthal, D. A., & March, J. G. (1993). The myopia of learning. *Strategic Management Journal, 14,* 95–112. http://dx.doi.org/10.1002/smj.4250141009

Liu, C., Eubanks, D. L., & Chater, N. (2015). The weakness of strong ties: Sampling bias, social ties, and nepotism in family business succession. *The Leadership Quarterly, 26,* 419–435. http://dx.doi.org/10.1016/j.leaqua.2015.02.007

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin, 116,* 75–98. http://dx.doi.org/10.1037/0033-2909.116.1.75

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review, 111,* 309–332. http://dx.doi.org/10.1037/0033-295X.111.2.309

Love, B. C., & Otto, A. R. (2010). You don't want to know what you're missing: When information about forgone rewards impedes dynamic decision making. *Judgment and Decision Making, 5,* 1–10.

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis.* New York, NY: Wiley and Sons.

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

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science, 2,* 71–87. http://dx.doi.org/10.1287/orsc.2.1.71

Medin, D. L., Dewey, G. I., & Murphy, T. D. (1983). Relationships between item and category learning: Evidence that abstraction is not automatic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 607–625. http://dx.doi.org/10.1037/0278-7393.9.4.607

Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179–196). Cambridge, MA: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511529863.009

Mehlhorn, K., Newell, B. R., Todd, P. M., Lee, M. D., Morgan, K., Braithwaite, V. A., . . . Fiedler, K. (2015). Unpacking the exploration—Exploitation tradeoff: A synthesis of human and animal literatures. *Decision, 2,* 191–215. http://dx.doi.org/10.1037/dec0000033

Mitchell, T. M. (1980). *The need for biases in learning generalizations.* Department of Computer Science, Laboratory for Computer Science Research, Rutgers University, Camden, NJ.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92,* 289–316. http://dx.doi.org/10.1037/0033-295X.92.3.289

Navarro, D. J., & Perfors, A. F. (2011). Hypothesis generation, sparse categories, and the positive test strategy. *Psychological Review, 118,* 120–134. http://dx.doi.org/10.1037/a0021110

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2,* 175–220. http://dx.doi.org/10.1037/1089-2680.2.2.175

Niv, Y., Daniel, R., Geana, A., Gershman, S. J., Leong, Y. C., Radulescu, A., & Wilson, R. C. (2015). Reinforcement learning in multidimensional environments relies on attention mechanisms. *The Journal of Neuroscience, 35,* 8145–8157. http://dx.doi.org/10.1523/JNEUROSCI.2978–14.2015

Niv, Y., Joel, D., Meilijson, I., & Ruppin, E. (2002). Evolution of reinforcement learning in uncertain environments: A simple explanation for complex foraging behaviors. *Adaptive Behavior, 10,* 5–24. http://dx.doi.org/10.1177/1059-712302-010001-01

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115,* 39–61. http://dx.doi.org/10.1037/0096–3445.115.1.39

Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C., & Glauthier, P. (1994). Comparing models of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961)). *Memory & Cognition, 22,* 352–369. http://dx.doi.org/10.3758/BF03200862

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review, 101,* 53–79. http://dx.doi.org/10.1037/0033-295X.101.1.53

Rehder, B., & Hastie, R. (2001). Causal knowledge and categories: The effects of causal beliefs on categorization, induction, and similarity. *Journal of Experimental Psychology, 130,* 323–360.

Rich, A. S., & Gureckis, T. M. (2017). Exploratory choice reflects the future value of information. *Decision*. Advance online publication. http://dx.doi.org/10.1037/dec0000074

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

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382–439. http://dx.doi.org/10.1016/0010-0285(76)90013-X

Schwartz, B. (1982). Reinforcement-induced behavioral stereotypy: How not to teach people to discover rules. *Journal of Experimental Psychology: General, 111,* 23–59. http://dx.doi.org/10.1037/0096-3445.111.1.23

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237,* 1317–1323. http://dx.doi.org/10.1126/science.3629243

Shepard, R. N., Hovland, C. L., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs, 75,* 1689–1699.

Speekenbrink, M., & Konstantinidis, E. (2015). Uncertainty and exploration in a restless bandit task. *Topics in Cognitive Science, 7,* 351–367. http://dx.doi.org/10.1111/tops.12145

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction.* Cambridge, UK: Cambridge University Press.

Taylor, E. G., & Ross, B. H. (2009). Classifying partial exemplars: Seeing less and learning more. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 1374–1380. http://dx.doi.org/10.1037/a0016568

Taylor, M. G. (1996). The development of children's beliefs about social and biological aspects of gender differences. *Child Development, 67,* 1555–1571. http://dx.doi.org/10.2307/1131718

Taylor, M. G., Rhodes, M., & Gelman, S. A. (2009). Boys will be boys; cows will be cows: Children's essentialist reasoning about gender categories and animal species. *Child Development, 80,* 461–481. http://dx.doi.org/10.1111/j.1467-8624.2009.01272.x

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences, 24,* 629–640. http://dx.doi.org/10.1017/S0140525X01000061

Teodorescu, K., & Erev, I. (2014). On the decision to explore new alternatives: The coexistence of under- and over-exploration. *Journal of Behavioral Decision Making, 27,* 109–123. http://dx.doi.org/10.1002/bdm

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review, 55,* 189–208. http://dx.doi.org/10.1037/h0061626

Tversky, A. (1977). Features of similarity. *Psychological Review, 84,* 327–352. Advance online publication. http://dx.doi.org/10.1037/0033-295X.84.4.327

Waxman, S. R. (1990). Linguistic biases and the establishment of conceptual hierarchies: Evidence from preschool children. *Cognitive Development, 5,* 123–150. http://dx.doi.org/10.1016/0885-2014(90)90023-M

Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A., & Cohen, J. D. (2014). Humans use directed and random exploration to solve the explore-exploit dilemma. *Journal of Experimental Psychology: General, 143,* 2074–2081. http://dx.doi.org/10.1037/a0038199

Wolpert, D. H. (1996). The lack of a priori distinctions between learning algorithms. *Neural Computation, 8,* 1341–1390. http://dx.doi.org/10.1162/neco.1996.8.7.1341

E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at https://my.apa.org/portal/alerts/ and you will be notified by e-mail when issues of interest to you become available!