Can losses help attenuate learning traps?

Amy X. Li[1] (amy.x.li@outlook.com), Todd Gureckis[2] (todd.gureckis@nyu.edu),
Brett K. Hayes[1] (b.hayes@unsw.edu.au)
[1]School of Psychology, UNSW Sydney, Sydney, NSW 2052, Australia
[2]Department of Psychology, New York University, 6 Washington Place, New York, NY 10003, USA

Abstract

Recent work has demonstrated robust *learning traps* during learning from experience – decision-making biases that persist due to the choice-contingent nature of outcome feedback. In two experiments, we investigate the effect of outcome valence on learning trap development. Participants chose to approach or avoid category exemplars associated with rewards or losses, and, to maximize reward, must learn a categorization rule based on two stimulus dimensions. We replicate previous findings showing that when outcome feedback was contingent upon approaching exemplars, people frequently fell into the trap of using an incomplete categorization rule based on only a single dimension, which was suboptimal for long-term reward. Notably, learning trap development was attenuated in an environment with frequent loss outcomes, even when participants received explicit information about the base rates of gains and losses. The implications of these findings for theoretical models and future research are discussed.

Keywords: categorization; learning traps; decision-making; valence; approach-avoid behavior

People are often required to learn about the world from interacting with their environment, rather than from passive observation alone. The ability to learn from one's experiences is critical, for example, for a young child to learn which foods they dislike, or about whom they can rely upon to provide security. However, in some situations, mechanisms that operate in experiential learning can lead to the development of persistently suboptimal patterns of behavior. The process whereby such suboptimal behaviors are developed and sustained has been termed a *learning trap* (Erev, 2014).

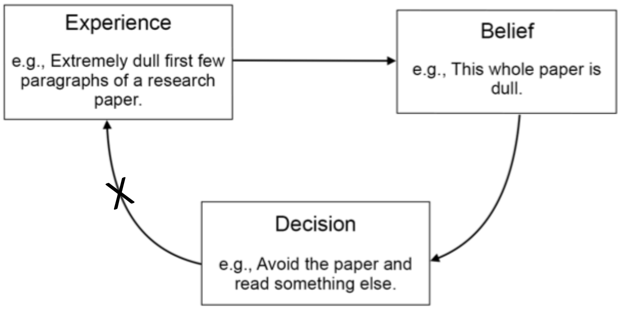
Experience
e.g., Extremely dull first few paragraphs of a research paper.
Belief
e.g., This whole paper is dull.
Decision
e.g., Avoid the paper and read something else.
X

Figure 1: The interaction between experiences, beliefs, and decisions in learning trap formation. Since experiences are decision-contingent, erroneous initial beliefs cannot be corrected when the belief itself leads to the avoidance of corrective experience (represented by the "X").

To illustrate how suboptimal behavior can arise from everyday experiential learning, consider the example illustrated in Figure 1. Here, early experience (e.g., reading the first page of a research paper) leads to a more general negative belief (e.g., "this is a dull paper"). This in turn leads to subsequent selective avoidance (avoiding reading the rest of the paper current paper being read) which limits further learning about the "true" value of the stimulus). In doing so, the learner is protected from further negative experiences. However, this behavior pattern also limits the opportunity for revising inaccurate beliefs formed early in the learning process (e.g., interesting results that appear later in the paper). Thus, an interaction between a person's experiences, beliefs, and decisions "traps" the learner into a pattern of behavior that, in the long run, is maladaptive for reward and learning.

Learning traps thus form an especially stubborn and pernicious class of errors in experiential learning, and can be observed in maladaptive behavior across a number of domains (e.g., Denrell & March, 2001; Fazio, Eiser, & Shook, 2004). For example, the hot stove effect (Denrell & March, 2001) describes risk aversion as a product of adaptive sampling; even if a risky option is positive on average, inherent uncertainty in its outcomes means that there will be times when recent experience will be negative, often leading to the future avoidance of risky options. A similar process is also thought to play a role in prejudice formation and in-group bias (Denrell, 2005; Liu, Eubanks, & Chater, 2015); negative biases about outgroups are likely to persist as this belief leads to the avoidance of corrective experience.

The Role of Categories in Experiential Learning

Everyday learning often involves representing a multidimensional environment in a way that enables learning and decision making to our benefit. To aid adaptive functioning, we must therefore learn appropriate categorization rules that will allow us to predict relationships between combinations of object features and outcomes.

Using a modified category learning paradigm with an "approach-avoid" component, Rich and Gureckis (2018) recently demonstrated how mechanisms in category learning, combined with the choice-contingent nature of experiential learning, can "trap" us into patterns of suboptimal behavior. In this task, participants approached or avoided stimuli that belonged to one of two categories – "friendly bees", yielding a 1-point reward if approached, or "dangerous bees", yielding a 3-points loss if approached. Bee exemplars varied on four binary dimensions, two of which were relevant for predicting approach outcomes; a specific combination of features on the two relevant dimensions predicted “dangerous” bees, while the remaining stimuli were “friendly”. A two-dimensional rule was therefore needed to correctly classify all stimuli, and to develop the correct strategy for approach-avoid decisions.

Crucially, when outcomes associated with stimuli were only available upon approach, in a *contingent feedback* condition – resembling the typical case when learning from experience in real life – less than 25% of participants successfully learned the two-dimensional rule, despite ample opportunity to learn. Instead, the majority learned a simpler, suboptimal rule involving approaching bees based on their feature on a single relevant dimension. In contrast, most of those in a *full-information feedback* condition, who received outcome feedback regardless of approach, successfully learned a two-dimensional rule.

Rich and Gureckis (2018) explained these results by suggesting that mechanisms in experiential learning (as outlined in Figure 1), coupled with selective attention mechanisms and a preference for learning simpler rules, led those in the contingent feedback condition into a persistent learning trap. Early in learning, participants in the contingent feedback condition may notice that certain features on one of the relevant dimensions sometimes lead to losses, and then prematurely develop the belief that this dimension is always associated with negative outcomes. Crucially, in subsequent learning, the participant avoids any instances with this feature – and hence does not receive the feedback that would correct the one-dimensional rule – forgoing many trials where they could have received a reward. Consequently, a persistent learning trap develops, leading to suboptimal performance in the long term and preventing learners from exploring the environment in a way that allows them to revise their erroneous belief.

Valence Asymmetries in Attention and Exploration

One factor that may affect a learner’s propensity to continue exploring learning environments, but has not yet been considered in the context of learning traps, is related to the valence of outcomes associated with learners’ choices. Although early work in decision-making highlighted the concept of loss aversion (e.g., Kahneman & Tversky, 1979; Tversky & Kahneman, 1992), recent findings have called into question its generalisability (e.g., Walasek & Stewart, 2015). These findings suggest that the risk of losses does not always lead to behavioral avoidance of particular choice options. A more robust finding is that the prospect of a loss increases *attention* to options more than the prospect of an equal gain – a phenomenon termed *loss attention* (Yechiam & Hochman, 2013; see Lejarraga et al., 2019 for a review).

Using simple two-choice tasks, studies have demonstrated the increased choice exploration of options involving losses relative to options that involve gains. For example, Yechiam and Hochman (2013; Experiment 1) presented participants with a safe alternative with a certain gain (e.g., +35 points with 100% probability) and a risky alternative. The risky alternative involved either an uncertain sure gain (+1 or +200 points, each with .5 probability), or a large gain or a small loss (-1 or +200 points, each with .5 probability). They found that the preference for the risky alternative was higher when it involved small losses compared to when it did not. Increased attentional exploration of losses relative to gains has also been demonstrated using process-tracing methods (e.g., Lejarraga et al., 2019; Pachur et al., 2018). Such studies suggest that negatively valenced options – that is, options that lead to losses – may increase both attention to item features and behavioral exploration of choice options, relative to positively valenced options that lead to gains.

Loss Attention and Learning Traps

These findings of valence asymmetries in attention and exploration suggest that the valence of choice options may be an important factor in the development of learning traps like those studied by Rich and Gureckis (2018). In a payoff structure such as the one used by Rich and Gureckis (2018), gains are frequently encountered (75% of items) while losses are less common (25% of items), and therefore levels of loss attention would be relatively modest. In contrast, studies of loss attention suggest that a payoff structure where losses are encountered more frequently than gains may *enhance attention to relevant category dimensions*, thereby resulting in learners being less likely to fall into a one-dimensional learning trap.

Given the converging evidence in the domains of attention and decision-making, the literature raises the possibility that negatively valenced category environments could attenuate learning trap development. This question, however, has not yet been examined, and constitutes the primary question of interest in the present studies.

The Current Studies

To address this aim, we examined learning in two contrasting payoff schedules; *frequent-gains*, where most exemplars predict positive outcomes (small gains), and *frequent-losses*, where most exemplars predict negative outcomes (small losses). Payoffs in the frequent-gains condition were identical to those used by Rich and Gureckis (2018), while the valence of these payoffs was reversed in the frequent-loss condition. Both payoff conditions required attention to two stimulus dimensions in order to optimize performance. Figure 2 shows a summary of this structure.

By comparing participants’ approach and avoidance responses in each category structure, we assessed whether the valence of payoff schedules could affect learning trap development. To create baseline versions of each type of task, we also presented corresponding *full feedback* conditions for each payoff condition, where feedback about the gain or loss outcome associated with each stimulus was provided regardless of whether the stimulus was approached.

In an environment with frequent-gains, we expected to replicate the main result from Rich and Gureckis (2018): when feedback about outcomes is choice-contingent, learners will likely fall into the trap of using a simpler, yet incomplete, one-dimensional decision rule. That is, we expect learners will choose to approach and avoid exemplars based on their values on a single feature dimension, resulting in an avoidance of some “friendly bee” exemplars.

In contrast, in the frequent-losses condition with contingent feedback, we expected that increased loss-induced attention will enhance attention to the relevant category dimensions. This should lead to less use of the suboptimal one-dimensional ("1D") rule compared to the corresponding frequent-gains condition; likewise, we expected that learners would be more likely to learn the correct two-dimensional ("2D") categorization rule. Since learning traps arise from choice-contingent feedback, we expected that the one-dimensional learning trap would be less evident in the full feedback conditions for both payoff conditions.

Experiment 1

Method

Participants 215 adults (M_{age} = 37.68 years, 142 male, 72 female, 1 other) from the United States were recruited, and participated, online through Amazon Mechanical Turk. Participants were paid a base amount, plus a bonus depending on points they earned during the task (*M* = $3.66 AUD).

Materials Stimuli were computer-generated images of bees that varied along four binary visual dimensions**.** Stimuli were constructed by selecting one of two possible feature values on each of four visual dimensions. The dimensions were *antennae* (two or none), *body* (spotted or striped), *legs* (two or six), and *wings* (round or long). As such, the four dimensions generated a total of 16 unique stimuli.

Two out of the four stimulus dimensions were randomly chosen to be relevant for the outcomes of approach-avoid decisions. For the two relevant dimensions, a combination of feature values was selected to form the *target combination* (see Figure 2 for an example).

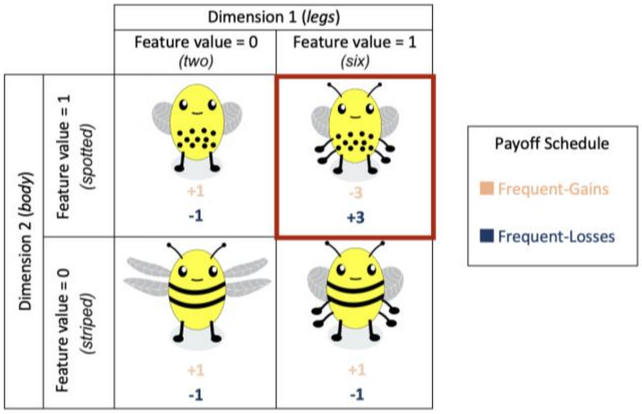

Figure 2: Example reward structure in Experiments 1 and 2, with *legs* and *body* as the relevant dimensions. The target combination is outlined in red. Here, a possible 1D rule involves approach-avoid decisions based on only the *legs* dimension.

In the frequent-gains condition, stimuli with the target combination of features were "dangerous" and approaching them led to a loss of 3 points; the remaining stimuli were "friendly" and approach led to a gain of 1 point. Conversely, in the frequent-losses condition, stimuli with the target combination were "friendly" and approach led to a gain of 3 points; the remaining stimuli were "dangerous" and approach led to a loss of 1 point.

Procedure Participants were first presented with a cover story which described that there were friendly and dangerous bees, and that their goal was to collect as much honey as possible from friendly bees, which would then translate into points for a bonus payment. Participants were informed that it was possible to predict which bees are friendly and which are dangerous using each bee's features.

Participants then completed a learning phase, beginning with a balance of 50 points. The learning phase consisted of eight blocks of 16 trials, so that each possible combination of the four dimensions (i.e., each of 16 unique stimuli) was encountered eight times. Hence, the total values of points gains/losses associated with 2D and 1D rules were identical across payoff conditions. Within each block, stimuli were pseudo-randomized such that in the frequent-gains condition, every eight stimuli contained six friendly and two dangerous bees; while in the frequent-losses condition, every eight stimuli contained two friendly and six dangerous bees.

On each trial, participants were presented with a bee stimulus and decided whether to approach ("harvest") or avoid it. There was no time limit to respond. Approach led to a gain or loss in points, depending on the payoff condition and the stimulus, and avoidance led to no change in points.

Participants in the *contingent feedback* group were shown the number of points they earned after approaching the bee, but received no feedback about outcomes if they avoided it. Those in the *full feedback* group were also given feedback after approaching the bee; however, if they avoided it, they were still shown whether the bee was friendly or dangerous. Following feedback, a new stimulus was displayed with an onscreen counter updated to reflect the participant's current point balance.

A test phase consisting of two blocks of 16 trials followed the learning phase, such that each combination of the four stimuli features was encountered twice. Trials were identical to those in the learning phase, but no feedback about the outcome of approach or avoid decisions was given.

Results and Discussion

Classifying Participant Behavior The 1D and 2D rules in the frequent-gains and frequent-losses conditions led to different expected patterns of approach-avoid decisions across a 16-trial block.

To illustrate how choice patterns for 1D and 2D rule-use were calculated, let us consider the frequent-gains payoff schedule shown in Figure 2. According to the 1D rule, within a 16-trial block, a participant would approach items based only on feature values from a single relevant dimension; if the 1D rule is defined on Dimension 1 (*legs*), they would avoid all items with a feature value of 1 on Dimension 1 (*six legs*), and only approach those with a feature value of 1 on the same dimension (*two legs*). That is, the participant would approach only eight items (i.e., only 2/3 of all "friendly" items), and avoid the other eight items (i.e., erroneously avoid 1/3 of all "friendly" items). (Recall that the 16 items comprising any given block were all distinct from one another.)

In contrast, if a participant made choices according to a 2D rule in a single 16-trial block, they would avoid only four items (e.g., those with feature combination [1,1] on dimensions 1 and 2 respectively). They would approach the 12 remaining unique items (e.g., those with feature combinations [0,1], [1,0] and [0,0] on dimensions 1 and 2).

For the frequent-losses condition, choice patterns for 1D and 2D rule use was calculated in a similar manner. Following on from the previous example where participants form a 1D rule based on Dimension 1, participants could now *approach* all items with a feature value of 1 on Dimension 1 (i.e., [1,1], [1,0] on dimensions 1 and 2 respectively; 8 out of 16 unique items in a given block) and avoid the remaining 8 items. On the other hand, if a participant followed a 2D rule, they would approach 4 items (e.g., those with feature combination [1,1] on dimensions 1 and 2) and avoid the 12 remaining unique items (e.g., those with feature combinations [0,1], [1,0], and [0,0]).

In the learning phase, if choices conformed to either the 1D or 2D pattern of approach-avoid decisions on at least 15 out of 16 trials within a given block, they were classified as using that rule. Since there are two possible 1D rules, we calculated the number of trials on which choices adhered to each of the possible 1D rules, and then took the maximum over these two quantities. (Note that classification of a block as adhering to a 1D or 2D rule are thus mutually exclusive.) If the criterion for neither rule was met, participants were considered as using an "unclassified" rule.

In the test phase, since no feedback is given and thus no learning occurs, we combined the two test-phase blocks (i.e., blocks 9 and 10) into a 32-trial block. Consequently, the criteria for classification as following a 1D or 2D rule during the test phase was 30 rule-congruent responses out of 32 trials; otherwise, the participant was "unclassified".

This method of rule classification was identical to that used by Rich and Gureckis (2018).

Task Performance During Learning and Test Phases Participants' rule use on blocks across the learning and test phase is shown in Figure 3, which show the proportion of participants using a 1D or 2D rule on a given block. (Note that we only present the data for 1D and 2D rule use here, since unclassified rule use is a complement of the other two proportions, and of less relevance to our research question.)

Since rule classification consisted of three mutually exclusive categories, we used multinomial logistic regression to analyze rule use. Using likelihood ratio tests, we conducted a forward selection procedure on the variables of interest. From the forward selection procedure, we selected a reduced parsimonious model from which we then extracted test statistics associated with each predictor (Matuschek et al., 2017 for a discussion of parsimonious vs. maximal models).

We found that receiving full feedback, instead of contingent feedback, increased participants' likelihood of using a 2D rule rather than a 1D rule by more than 2 times in the learning phase (OR = 2.97, 95% CI [2.10, 4.19], $p < .001$); and by the test phase, by more than 4 times (OR = 4.12, 95% CI [1.64, 10.31], p = .003). Replicating Rich and Gureckis (2018), these results indicate when feedback was not contingent upon approach decisions (in a baseline *full feedback* condition), most people experiencing a payoff schedule with frequent gains learned the optimal two-dimensional rule. However, in the contingent feedback frequent-gains condition, which replicated the payoff schedule used by Rich and Gureckis (2018), a majority tended to adopt an incomplete one-dimensional rule (39.1% during test) – indicating the development of a learning trap.

Turning to our primary question of whether payoff schedule valence (*frequent-gains* vs. *frequent-losses*) affects learning trap development, we found that in the contingent feedback conditions, a frequent-losses payoff schedule increased the likelihood of using a 2D rule as compared with a frequent-gains schedule, by more than 3 times in the learning phase, OR = 3.89, 95% CI [2.72, 5.55], $p < .001$; and in the test phase, by more than 8 times , OR = 8.23, 95% CI [2.84, 23.89], $p < .001$. This indicated that a learning environment with frequent small losses reduced the use of an incomplete 1D rule, and encouraged the use of a complete 2D rule, relative to an environment with frequent small gains.

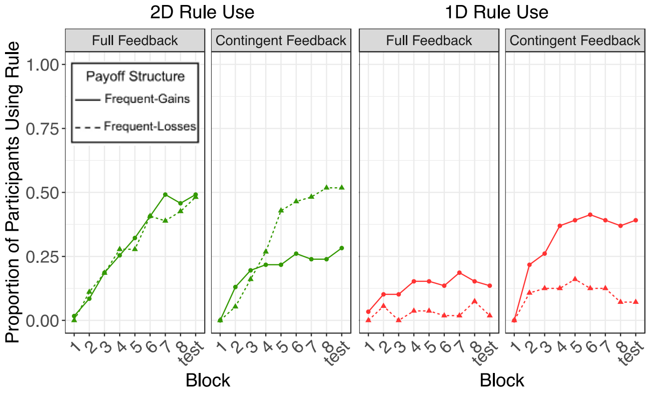

Figure 3: The proportion of participants in each condition classified as adopting the correct 2D rule, or falling into the 1D learning trap, over the course of Experiment 1.

The results of Experiment 1 therefore appear congruent with the account that loss attention led to those in a frequent-losses condition to explore different feature combinations on early trials, which then increased the likelihood of discovering that a combination of features from two dimensions are needed to optimise reward.

Experiment 2

Experiment 1 found an advantage for those in a learning environment with frequent losses, showing that they were less likely to fall into the one-dimensional learning trap, and more likely to learn a complete two-dimensional rule on the current task. However, another explanation for this advantage is that individuals may have prior beliefs about the proportion of good versus bad items that lead to a higher likelihood of avoidance. As such, individuals in the frequent-gains condition may inadvertently fall into the learning trap. If this was true, having some prior knowledge of the relevant base rates in the frequent-losses (i.e., many dangerous, few friendly) and frequent-gains (i.e., many friendly, few dangerous) conditions should decrease 1D responding, and reduce differences between the frequent-gains and frequent-losses conditions in 2D rule use.

As such, the primary goal of Experiment 2 was twofold: first, to replicate the finding suggesting an effect of outcome valence; and second, to follow up on one possible alternative explanation that might lead to differences in how people respond in the two payoff conditions.

Method

Participants 198 US adults (M_{age} = 38.08 years, 127 male, 71 female) participated online through Amazon Mechanical Turk. The payment scheme was identical to Experiment 1 (*M* = $3.48 AUD), and those who had previously completed Experiment 1 were excluded from participation.

Materials and Procedure Participants completed the same decision-based category learning task used in Experiment 1, with two differences. First, given that the primary aims of Experiment 2 concerned further investigating the learning trap that arises from choice-contingent feedback, the full feedback condition was omitted.

Second, we added groups who received additional instructions about the base rates of friendly and dangerous bees. In the *no base-rate tip* condition, task instructions were identical to those in Experiment 1. In the *base-rate tip* condition, participants were explicitly instructed about the relative proportions of positive versus negative stimuli during task instructions. In this condition, a comprehension check question was added after the instructions to ensure that participants remembered the base rate information given.

Results and Discussion

Task Performance Participants' rule use over the course of the learning and test phase is shown in Figure 4.

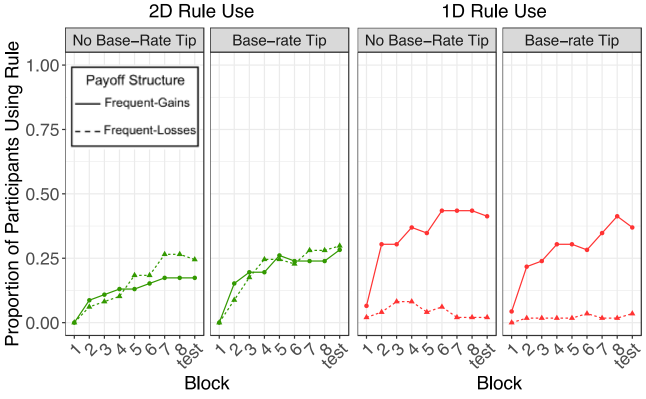

Figure 4: The proportion of participants in each condition classified as adopting the correct 2D rule or falling into the 1D learning trap, over the course of Experiment 2.

Replicating the key finding of Experiment 1, the 1D learning trap was attenuated in a contingent-feedback frequent-losses condition. Compared to participants in contingent frequent-gains conditions, participants who experienced the frequent-losses payoff conditions were more than 10 times as likely to use a 2D rule over a 1D rule during the learning phase (OR = 10.59, 95% CI [6.57, 17.06], $p <$ .001); by the test phase, they were more than 16 times as likely (OR = 16.57, 95% CI [4.49, 61.10], $p < .001$).

Comparison of test-phase rule use in the present frequent-losses conditions (i.e., *tip* and *no tip*) with the corresponding frequent-losses condition in Experiment 1 (i.e., *contingent, frequent-losses*) suggests that 2D rule use was relatively less common in the present experiment (27.16% in Experiment 2, vs. 51.79% in Experiment 1). Nevertheless, statistical analysis still indicated that the effect of payoff schedule on two-dimensional rule use was robust.

An effect of base-rate tip was also found, but only for the learning phase. During the learning phase, receiving a tip instead of no tip increased a participant's likelihood of using a two-dimensional rule by a factor of two, OR = 2.06, 95% CI [1.42, 2.98], $p < .001$; by the test phase, no effect of base-rate tip was found ($p = .43$). Although a visual inspection of the data suggests that receiving a base-rate tip may contribute to increased 2D rule use in the frequent-gains condition (17.39% without tip, 28.26% with tip), levels of 1D rule use still remained high in the frequent-gains condition (41.30% without tip, 36.96% with tip); consequently, analysis did not indicate a significant interaction between base-rate tip and payoff condition (learning phase: $p = .29$; test phase: $p = .76$).

In summary, Experiment 2 generally replicated the primary findings of Experiment 1. An environment with a frequent-losses payoff schedule led learners to be more likely to use a complete 2D rule instead of falling into a 1D learning trap, relative to frequent-gains. Additionally, informing participants about the base rates of positive and negative stimuli prior to commencing the task did not completely account for the outcome valence effect on learning trap attenuation. Receiving information about base rates had a limited effect on attenuating the learning trap during the learning phase, but did not affect decisions by the test-phase.

General Discussion

The present experiments examined how we may attenuate the development of a suboptimal behavioral learning trap that emerges during experiential category learning. We replicated previous findings demonstrating a persistent learning trap (Rich & Gureckis, 2018), in which the choice-contingent nature of learning from experience prevents the correction of a learning error. In a frequent-gains environment (i.e., a category structure involving frequent gains and infrequent losses), we found that this behavioral learning trap frequently led people to selectively attend to only a single dimension, and prevented people from learning about the true two-dimensional structure of the task.

A novel finding that emerged across the two experiments was that the learning trap was attenuated in a frequent-losses environment (i.e., a structure involving frequent losses and infrequent gains), and that levels of learning the complete and optimal categorization rule was higher, relative to a frequent-gains environment. Moreover, Experiment 2 showed that this effect of payoff schedules persisted, even when we attempted to correct potentially erroneous prior beliefs about the base rates of positive versus negative stimuli.

The effect of frequent losses in encouraging the use of an optimal two-dimensional rule, as opposed to an incomplete one-dimensional rule, was shown to be statistically robust across the two experiments. Nevertheless, we would like to note that 2D rule use in the frequent-losses condition appeared less frequent in Experiment 2 than in Experiment 1. These results could suggest that while frequent losses helped people avoid the learning trap, this does not always mean that the optimal two-dimensional rule will be learned.

The present results are consistent with previous findings from simpler, gamble-type tasks that demonstrate a valence asymmetry in attentional and choice exploration (i.e., loss attention). These studies have found loss-induced increases in attention both when operationalized through choice (e.g., Yechiam & Hochman, 2013), and through process-tracing measures of attention such as cursor-tracking tools (e.g., Lejarraga et al., 2019). Following on from these findings, we speculated that increased "loss attention" in a learning environment with frequent losses (and infrequent gains) may enhance attentional exploration of the relevant category dimensions. Present results were consistent with this notion; frequent losses appear to have increased attention to learning the specific combinations of features that predicted gain or loss outcomes, thereby increasing the likelihood of learning a complete two-dimensional rule.

It is notable that a frequent-losses environment was able to attenuate the development of a one-dimensional learning trap, given the previously demonstrated persistence of this trap. With the goal of slowing the narrowing of attention to a single feature dimension, Rich and Gureckis (2018) tried three separate "interventions" within the choice-contingent feedback paradigm: introducing stochasticity to exemplar-outcome associations, random occlusion of dimensions, and adding individuating features to each unique exemplar during learning. All three had limited success; none of these strategies effectively aided the learning of optimal 2D rule, and any attenuation in 1D responding was likewise accompanied by reductions in 2D responding, indicating that participants found it difficult to learn any type of dimensional rule. Our findings thus show that loss-induced increases in choice and attentional exploration may help people overcome a persistent error in experiential learning.

Future Directions

An important question that remains is exactly *how* people learn about the category environment – and consequently how learning traps develop or are attenuated – on a trial-by-trial basis. By assessing how formal models of category learning relate to our data, we may be able to better understand the mechanisms that explain the formation of learning traps, and their attenuation, in this study.

One model that has been assessed against behavior on the *frequent-gains* version of the present task is a modified version of the popular exemplar-based connectionist categorization model, ALCOVE (Kruschke, 1992), called ALCOVE-RL (Jones & Cañas, 2010; Rich & Gureckis, 2018). Similar to ALCOVE, ALCOVE-RL classifies a new stimulus based on its similarity to category exemplars stored in memory. Since not all features are relevant to discriminating between categories, a key feature of ALCOVE-RL is a selective attention mechanism, which shifts attention following errors in predicting category outcomes. In addition to the core ALCOVE architecture, ALCOVE-RL includes a reinforcement learning and choice-contingent feedback mechanism, such that network nodes are updated when instances are approached and outcome feedback is received, but not when they are avoided. ALCOVE-RL was able to successfully simulate the higher rate of learning a suboptimal 1D rule in the frequent-gains contingent feedback condition as compared with the full feedback condition. However, the model underpredicted the extent to which the learning trap developed when compared to data in the frequent-gains version of the present task.

Future modelling work will involve refinement of ALCOVE-RL's feedback mechanism to better simulate actual learning, and extension of the model to explain the current findings of learning trap attenuation in the *frequent-losses* conditions. It will also be useful to examine whether the current data can be better explained by other formal models of category learning, such as ATRIUM (Erickson & Kruschke, 1998), which is a hybrid exemplar- and rule-based model. One key feature of ATRIUM is that it can learn a general categorization rule that applies to most training stimuli, but separate rules that apply to "exception" items. In the current context, ATRIUM may fall into a learning trap by learning an incomplete and simpler rule on early training trials and – due to the choice-contingent nature of feedback – failing to identify important exceptions to this rule.

Implications

The present work addresses an important intersection between category learning and reinforcement learning that is ubiquitous in everyday scenarios, where we must learn to represent a multidimensional world in a way that helps us make decisions. As noted by Radulescu, Niv, and Ballard (2019), current accounts of reinforcement learning still cannot fully account for how we represent multidimensional environments to maximize favorable outcomes. On the other hand, little work has been done in category learning to address the same question, despite the clear relevance of categorization in accomplishing this important task; category learning studies that *have* considered the role of outcomes in learning performance have noted the pitfall of considering rewards but not losses (e.g., Schlegelmilch & von Helversen, 2020), even though both occur in real-world learning.

To this end, we extend an emerging literature that addresses how suboptimalities in learning may emerge in the intersection between reinforcement and category learning, by considering an addition theoretical dimension – the valence of one's learning environment. Present findings thus offer a starting point to suggest how we may overcome a persistent and consequential bias in human experiential learning. Given the implications of learning traps for our everyday lives, the understanding of mechanisms that influence their formation – which, in turn, informs how we may be able to prevent such traps – constitutes a compelling area for future inquiry.

Acknowledgements

We would like to thank Dr. Nicole Cruz for sharing her statistical expertise. This work was supported by an Australian Research Council Discovery Grant DP190101224 to BKH.

References

Denrell, J. (2005). Why most people disapprove of me: Experience sampling in impression formation. *Psychological Review, 112*(4), 951–978. https://doi.org/10.1037/0033-295X.112.4.951

Denrell, J., & March, J. G. (2001). Adaptation as information restriction: The hot stove effect. *Organization Science, 12,* 523–538.

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). Free University of Bozen-Bolzano.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General, 127*, 107–140.

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.

Jones, M., & Cañas, F. (2010). Integrating reinforcement learning with models of representation learning. *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society*.

Lejarraga, T., Schulte-Mecklenbeck, M., Pachur, T., & Hertwig, R. (2019). The attention–aversion gap: How allocation of attention relates to loss aversion. *Evolution and Human Behavior*, *40*(5), 457–469.

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*(3), 419–435.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language, 94*, 305–315.

Pachur, T., Schulte-Mecklenbeck, M., Murphy, R. O., & Hertwig, R. (2018). Prospect theory reflects selective allocation of attention. *Journal of Experimental Psychology: General, 147*(2), 147–169.

Radulescu, A., Niv, Y., & Ballard, I. (2019). Holistic reinforcement learning: The role of structure and attention. *Trends in Cognitive Sciences, 23*(4), 278–292.

Rich, A. S., & Gureckis, T. M. (2018). The limits of learning: Exploration, generalization, and the development of learning traps. *Journal of Experimental Psychology: General, 147*(11), 1553–1570.

Schlegelmilch, R., & von Helversen, B. (2020). The influence of reward magnitude on stimulus memory and stimulus generalization in categorization decisions. *Journal of Experimental Psychology: General, 149*(10), 1823–1854.

Yechiam, E., & Hochman, G. (2013). Loss-aversion or loss-attention: The impact of losses on cognitive performance. *Cognitive Psychology*, 66(2), 212–231.