Reward Count(s): Negative Recency in Probabilistic Experience-Based Learning

René Schlegelmilch (r.schlegelmilch@uni-bremen.de)
Department of Psychology, University of Bremen, GER and University of Zurich, CH

Andy Wills
Department of Psychology, University of Plymouth, UK

Bettina von Helversen
Department of Psychology, University of Bremen, GER and University of Zurich, CH

Abstract

Learning how to make decisions from experience is often studied using probabilistic outcome prediction or choice tasks, as in conditioning, reward learning, or risky gambles (e.g., response A provides reward in 75% of the cases, response B in 25% over repeated trials with feedback). One debated phenomenon in such tasks is that of negative recency, describing that learners expect the rare event after observing a streak of common events (e.g., Gamblers fallacy). Here, we show that this behavior, despite instructing participants to use a visual stimulus, also occurs in probabilistic single-cue conditioning training, where participants predicted whether digging at a specific location on a plane (visual cue) leads to finding a Vase or Nothing (events), when they received reward for correct predictions. We manipulated reward magnitude in three conditions (equal for both common and rare events vs. high for common event vs. high for rare event, between factor). We further manipulated whether the label of the rare event was framed as event (finding a Vase) or non-event (finding Nothing; between factor). The results suggest, that reward magnitude affected the emergence of negative recency, being most prevalent when correctly predicting the rare event yielded a high reward, and least prevalent when the common event yielded a high reward. Interestingly, the event label instead rather affected *when* the rare event was expected, such that common Vase runs were expected to end earlier than common Nothing runs. We discuss the findings from conditioning and economic perspectives, generally concerning experience-based learning.

Keywords: probability learning, conditioning, negative recency, reward magnitude

Introduction

How humans learn taking actions given a stimulus and different choice alternatives has inspired research on experience-based learning over a century, in classic conditioning (e.g., learning that blue things predict an event A, green things predict nothing), and economic decisions (e.g., choosing/predicting event A yields high reward, but event B low reward). A common question is how people act when these stimulus-event or action-reward relations are uncertain and outcome probabilities need to be learned. One shared perspective assumes that reinforcing choices in the presence of specific stimuli or options leads to repeating this behavior, a.k.a. reinforcement learning (RL; see Rescorla & Wagner, 1972; Bush & Mosteller, 1951; Pothos & Wills, 2011). However, recent debates question this view as people often show unexplained behavior known as probability matching (for reviews see Estes, 1964; Koehler & James, 2014) when there is uncertainty in how often reinforcement occurs. This is usually the case if feedback is probabilistic (e.g., action A leads to a higher reward than action B in 75% of the cases, which probability matching enacts). In recent economic studies, discussions re-emerge on how this reflects a variant of the so-called Gamblers fallacy (e.g., Plonsky et al., 2015; Szollosi et al., 2022; Schulze et al., 2020). That is, after observing the commonly reinforced event several times in a row, participants begin to expect the rare event. We here provide evidence from an experience-based event-prediction task, that the magnitude of payoff for correctly predicting the rare event moderates this trend, despite participants being instructed to use a visual stimulus.

As a classic example, imagine you have to predict in which color a light-bulb will come on (color y vs. color b) in about 400 repeated trials, as in the verbal conditioning study by Nicks (1959). This is a similar task to predicting the next winning color at the Roulette table but without monetary payoffs. The (unknown) probability of observing y is 70% (random), and an observed event sequence could be *yyybyby*. In contrast to the predictions of standard RL, Nicks (1959) found that the expectancy of the rare b event increased with the run length (or streak) of common y events. In conditioning research, this has been debated as *negative-recency* effect (see also Estes, 1964; Fernberger, 1920; Jarvik, 1951). As mentioned, it has recently be re-discovered in risky gambles (e.g., Plonsky et al., 2015; Szollosi et al., 2022), but also brought back to the attention of researchers in conditioning (Perruchet, 2015; Myers, 2014).

Indeed, a plethora of discussions circle around the phenomenon of negative recency, which we review from a general perspective in Schlegelmilch, Wills, & von Helversen (2023). However, due to its neglect in recent decades, two still understudied questions that stand out are (a) whether this effect also occurs when participants are explicitly instructed to use visual predictors of events and (b) how the magnitude of reward associated with both events influences the prevalence of negative recency (see also Estes, 1964; Tune, 1964). Since most studies on conditioning and economic decisions entertain the idea that participants learn stimulus-event, or action-reward associations, it seems due scrutinizing which factors underlie learning about sequential event patterns and negative recency.

Already Restle (1966) has shown that participants are able to learn to predict two non-monetary alternating events (b and y) when simply observing that each occurred after a specific

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

run length of the other event (e.g., alternation from *b* to *y* after observing *bbb*, and vice versa after observing *yyy*). Restle (1966) and other researchers therefore argued that people count the run-lengths of events (see also Goodnow & Pettigrew, 1955), while more recent approaches argue that people memorize whole sequences of patterns (Plonsky et al., 2015; Feher da Silva et al., 2017). However, Schulze et al. (2020) collected verbal reports about how participants would approach a monetary gamble task, finding, in line with early theories, that most would start counting event runs to predict event probabilities and their alternation.

Intriguingly, this not only concerns gambling scenarios, but also widely applied methods in classic conditioning, as used in our study, which employ probabilistic feedback. However, such learning tasks typically also involve putative visual predictors of events, and a question is, whether participants still would show negative recency behavior. For example, in conditioning, a typical focus lies on investigating the formation of cue-event associations, by repeatedly presenting a constant visual stimulus (cue) paired with a subsequent event. A common method is to analyze the average change in event expectancy over those trials, while negative recency is thereby often neglected. The presence of negative recency, however, would question the classic idea, that average learning curves reflect the formation of cue-event associations, as participants could also make use of the length of event runs as predictor.

The basic task-design used in our study reflects this single-cue conditioning paradigm (for an overview, see Mackintosh, 1974, see also Dunsmoor & Murphy, 2015; Lee et al., 2018; Lovibond et al., 2020), exemplified in Fig. 1A. It contains a training phase with trial-and-feedback learning. A visual cue on a continuous dimension is presented (a horizontal location), telling participants that this represents a digging site, and their task is to predict whether they will excavate a Vase or Nothing at this location, happening in repeated trials. After a choice, immediate feedback tells whether Vase or Nothing occurred. But this was probabilistic, such that one of both events was more common with 75% probability. In addition, if participants correctly predicted the event in a given trial, they received a reward in Thalers, the currency in this experiment. However, if incorrect, they received no reward, but were informed how many Thalers they would have earned if they predicted the other event in this trial.

Crucially, as in single-cue conditioning, the visual stimulus (location) was virtually constant such that it did not predict *when* the rare event can be observed. In principle, we argue, that for the given scenario, behavior should first be RL like (also called positive recency) if participants are instructed to use a visual predictor. However, once participants observed that the visual predictor is unreliable, they might search for other (hidden) predictors to achieve perfect accuracy and collect rewards, such as the run length of the common event (see also Estes, 1964). In other words, behavior should, over learning, change from positive to negative recency, which we test here. Crucially, already Tune (1964) speculated, similar to Estes (1964), that negative recency might only occur, when there is some interest in the rare event, which to our knowledge has not been studied. Our hypothesis is, that increasing the payoff associated to the rare event (relative to the common event) increases interest in it, leading to stronger negative recency.

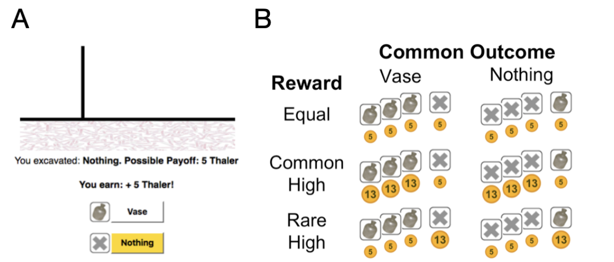

Figure 1. Training Trial Example and Study Design. (A) After presentation of the line location, participants predicted the event of an excavation (Vase or Nothing). Feedback is probabilistic (e.g., 75% Vase, 25% Nothing). The illustrated trial feedback shows that 'Nothing' was predicted and was correct (becoming yellow). Correct predictions yield reward in Thalers according to the study design; incorrect predictions yield 0 Thalers, but the possible reward is disclosed (reward associated to observed event). (B) Study design (fully crossed between): Reward magnitude (Equal vs. Common High vs. Rare High) X label of the events (common = Vase vs. Nothing). Coins (5 and 13) reflect reward in Thalers, later converted to a $ bonus.

More generally, it seems noteworthy, that introducing choice-contingent rewards into a single-cue conditioning task becomes practically identical to standard probability learning (see Estes, 1964) or probabilistic reward learning (i.e., which response leads to reward, e.g., Feher da Silva et al., 2017; Shanks et al., 2002). That is, in the latter, there is no putative causal cue, and in our study, the visual stimulus is virtually constant (non-diagnostic). The current design, thus, could be conceptualized as stimulus-event | response-reward learning, which allows connecting the perspectives of conditioning and probabilistic reward learning. However, to our knowledge, the influence of reward magnitude on negative recency has also not been studied in reward learning tasks.

In probability learning tasks, it is more common to take a perspective concerning how people form a preference for the choice options, such as estimating their long-term utility by integrating both probability and magnitude of the options' payoffs, which is typically tested via the participants' average choice rates. For example, if reward is paid for correctly predicting each event (e.g., as in Shanks et al., 2002; Feher da Silva et al., 2017), and the Vase event occurs in 75% of the trials, then choosing the Vase event should be strategically preferred over the alternative, because it has a higher expected value, which would enact a strategy known as probability maximizing (see Koehler & James, 2014).

Thus, it seems uncontroversial to expect that participants more often choose the more frequently occurring event (the common event) if both events are associated to the same immediate reward (e.g., 5 Thalers). However, manipulating the reward of these events (cet. par.) might induce a bias towards predicting the event associated to the higher reward in general. In this vein, the question is whether this happens in terms of a plain choice bias independent from the run lengths of events, or alternatively, whether this is driven by negative recency leading to more frequent rare-event choices, the longer the common-event run becomes. To test how reward magnitude affects the event predictions, we therefore employed three conditions (between subjects). In an equal rewards condition, similar to Shanks et al. (2002), *correctly predicting* each common and rare event led to the same reward of 5 Thalers in a given trial. However, in a Common High condition, correctly predicting the common event yielded 13 Thalers and the rare event 5 Thalers. Vice versa, in a Rare High condition, correctly predicting the common and rare events yielded 5 and 13 Thalers, respectively.

Finally, conditioning and reward learning sometimes confound whether 'something' (common event or reward) happens or 'nothing' happens (rare non-events or no reward), which could as well affect stimulus-event predictions (traditionally discussed in terms of inihibtory mechanism; e.g., Honig et al., 1963; Lovibond et al., 2020; Mackintosh, 1974). That is, the distinction between event labels and their associated rewards highlights, that our learning task includes two aspects that might concern cognitively distinct processes. First, learning of how likely a (non-monetary) event is either given a stimulus or the length of event runs, and, second, of how desirable the corresponding event is (reward magnitude). In turn, the common-event labels (Vase vs. Nothing) and their rewards (Equal vs Common High vs. Rare High) could affect either one or both of these mechanisms. Thus, to delineate the impact of reward magnitude from that of event/non-event coding, we also manipulated the labels of the commonly observed (75%) event either being 'Vase' or 'Nothing' (between, fully crossed). Fig. 1B illustrates the full design. Note, we here mainly focus on our central hypotheses regarding the effects of reward magnitude.

Experiment

Our task combines probabilistic stimulus-event and response-reward learning, which allows studying the potential interaction of corresponding learning processes. Note, we preregistered this study to investigate how reward and event labels influence event probability estimates during a test phase after training, where participants also responded to line stimuli that were unobserved in prior training (`osf.io/zswqm/`). However, the following report deviates from the pre-registration, and exclusively focuses on the training phase. The study was approved by the ethics committee of the Faculty of Arts and Social Sciences, University of Zurich (Nr. 19.2.19).

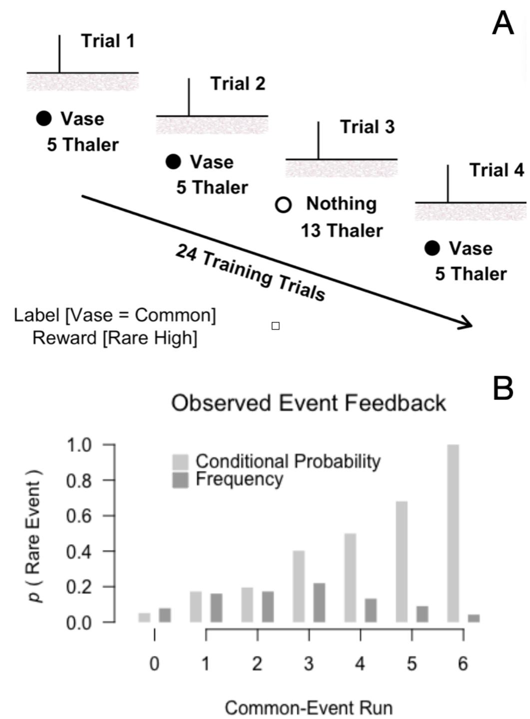

Figure 2. Trial-And-Feedback Procedure. (A) Example of trial sequence. A rare event was observed once in every 4 trials (1-4, 5-8, etc). (B) Resulting observed rare-event regularity in terms of conditional probabilities (i.e., rare next, given common-run length) and overall frequency (proportion of trials with corresponding common-run length; 1 = common event in last trial, 2 = in last two trials etc., 0 indicates previous rare feedback).

Method

Participants The experiment was carried out online using jsPsych (de Leeuw, 2018). We recruited 299 participants on amazon Mechanical Turk, randomly assigned to the six design cells. The study duration was about $M = 14.5$ minutes, compensated with $2.58 on average, comprised of a lump sum of $1 plus additional bonus contingent on the collected Thaler in training. As preregistered, we excluded 16 participants who incorrectly answered the questions for task comprehension twice, or who stated that their data should be excluded. We also excluded 32 participants who did not notice reward variations in both reward conditions in a manipulation check, leaving $N_{(E,V)} = 44$, $N_{(E,N)} = 43$, $N_{(C,V)} = 39$, $N_{(C,N)} = 43$, $N_{(R,V)} = 46$, and $N_{(R,N)} = 47$ (Reward: [E]qual vs. [C]ommon vs. [R]are High; Common Label: [V]ase vs. [N]othing).

Materials & Procedure After signing an informed consent, participants received instructions about the training phase, followed by control questions to ensure comprehension of the feedback and reward procedure. If incorrectly answered, a short summary of the instructions followed, and the questions were presented again, and repeated until all questions were answered correctly. Then, they completed the training and transfer phases. Upon completion, a manipulation check asked whether the potential rewards varied during training. Finally, after a debriefing, they indicated whether their data can be used or not.

The training procedure is illustrated in Fig. 2A. In 24 sequential training trials, six slightly different locations on a plane were marked by a vertical line. That is, the horizontal line was screen-centered, 450px wide, and segmented into 60 units, with one unit = width of the vertical line, and we presented only locations around unit 20 (i.e., 17 to 22), which were appearing as nearly identical. Each of the six locations was presented four times, and their order was randomized (without replacement). Out of the four respective presentations, each stimulus was paired three times with the common event, once with rare event (i.e., the location was non-diagnostic in predicting both events). In each trial, participants predicted whether 'Vase' or 'Nothing' will follow on a single location, by choosing from two buttons, followed by corrective feedback, presented below the stimulus, together with the reward obtained for the correct prediction according to the design in Fig. 1B. In incorrect trials, participants always received 0 Thalers, but we also uncovered the possible reward, such that they knew that they could have earned more by predicting the alternative event. To prevent that observed reward magnitude effects stem from being distracted by varying numbers, we also introduced slight trial-wise reward deviations in the 'Equal' condition, by randomly adding +.1, 0 or -.1 to the 5 Thaler.

Importantly, however, we randomized the trial sequences similar to Lee et al. (2018), resulting in rare events appearing uniformly over the 24 trials. Specifically, within every four trials (e.g., trials 1-4, 5-8 etc.), there was one and only one rare event for each participant. An important consequence of this randomization method is, that the temporal distance between two rare events becomes approximately regular, illustrated in Fig. 2B. Specifically, the mode-run-length of the common event was three, after which the rare event appeared most often (frequency; dark grey bars). Furthermore, the conditional probability that a run ended given the current run length increased monotonically, and a run definitely ended after six common events (light gray bars). Apparently, such a regularity implies, that there is something to learn about run length (e.g., as in Restle, 1966), which could be beneficial in collecting rewards.

Results

Fig. 3 shows the average choice probabilities over training. Clearly, the probability of predicting the common event differed between the reward conditions. On average, behavior in the 'Equal' rewards baseline approached the true base rate (75% common), aligning with previously observed probability matching trends (e.g., Shanks et al., 2002). As early hypothesized by Estes (1964), providing high reward for the common event increased the tendency towards always choosing common, while providing high reward for the rare event had the opposite effect. Also, common responding seemed generally lower when Nothing (non-event) was common than when Vase (event) was common, but there was no apparent interaction with reward magnitude.

Thus, to first take the classic approach, we tested these learning-curve effects, by conducting a standard logistic regression (R glm; R Development Core Team, 2008). Other methods led to very similar results (as for the following analyses), and we report this method for brevity. We included all main effects and interactions for trial (mean centered), reward magnitude, and label. The model ANOVA indicated a significant effect of trial $\chi^2(1) = 19.017$, $p < .001$, a main effect of reward magnitude, $\chi^2(2) = 123.186$, $p < .001$, and a main effect of event label, $\chi^2(2) = 35.019$, $p < .001$. Furthermore, there was an interaction between trial and reward magnitude, $\chi^2(2) = 24.744$, $p < .001$, while all other effects were non-significant ($p > .1$). Thus, as indicated in Fig. 3, participants in the Common High condition most quickly increased in their probability of predicting the common event over trials, but participants in the Rare High condition tended towards predicting the Rare Event, with the baseline falling in between. In contrast, the event label (common event = Vase vs. Nothing) only exerted a bias towards predicting the Vase event, which did not seem to change over learning.

However, while the learning-curve analysis suggests a mere bias towards predicting the more rewarding event, there was evidence that this was due to negative recency, illustrated in Fig. 4. Each panel shows the probability of predicting the rare event when observing common-event runs (x-axes). The four panels separate these recency trends depending on how many rare events were generally observed before the given run (e.g., Common Only = first trials before ever observing a rare event vs. 5 Rare Events = near the end of training). In Fig. 4A, the run length of 0 reflects behavior in trial 1, while

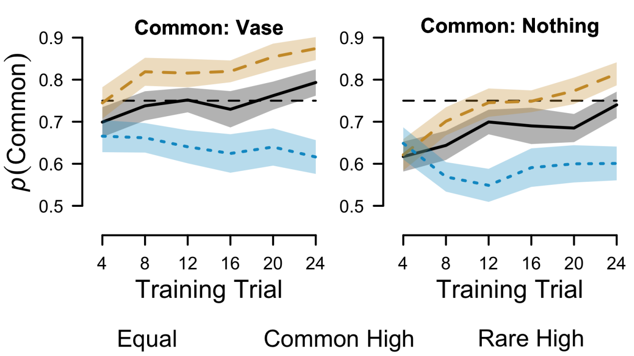

Figure 3. Training Performance. Probability of predicting the common event (y-axis) in each condition over training bins (6 blocks, 4 trials each). Horizontal dashed line (left) highlights true p(Common). Shaded areas = $+-1SE$.

in the other panels 0 indicates the response probability directly after observing the most recent rare event, and increasing run lengths reflect the series of common events thereafter.

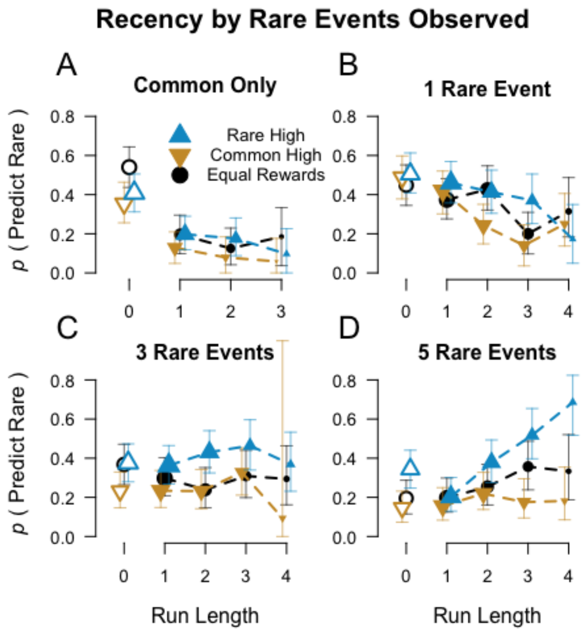

Figure 4. Negative Recency Effect in Training. Probability of predicting the rare event (y-axis), depending on run length of the common event (x-axis). Panels represent runs 1, 2, 4 and 5 (see text). Error bars indicate simulated 95%CIs of the choice probabilities in A, and estimated CIs in C to D from the regression model.

As can be seen, before observing a rare event (Fig. 4A), the initial choices reflected guessing (run length 0), but directly after observing the common event in trial 1 (run length 1), participants strongly expected another common event to follow, in line with the idea of stimulus-event reinforcement. After observing the first rare event (Fig. 4B), a similar but more continuous trend emerged. Crucially, the more rare events the participants encountered in previous training (Fig. 4C and D), the stronger negative recency became, seemingly moderated by reward magnitude.

To test this, we again used a standard logistic regression, now, however, coding the trials in terms of the common-event run length, and the number of rare events observed so far (factor coding), as done in Fig. 4, and as before, further including reward magnitude and event label. However, we restricted the run lengths to those larger than 0 to focus on behavior after participants observed the common events where negative recency is signified by an increase in rare-event predictions. We also only included factor levels for which there were more than 10 observations in each resulting design cell. This included run lengths 1 to 4 in general (as depicted in Fig. 4), but excluded the first run (i.e., Common Only, Fig. 4A) and the very last run for which only run lengths 1 and 2 were available. Otherwise, the method would result in model-convergence and effect-identification issues.

As before, the model ANOVA indicated main effects of reward magnitude $\chi^2(2) = 88.099$, $p < .001$, and event label, $\chi^2(1) = 14.576$, $p < .001$. Importantly, according to the hypothesis that behavior changes from positive to negative recency in general, there was a significant interaction of run length and the number of observed rare events so far $\chi^2(12) = 54.916$, $p < .001$. However, as also seen in Fig. 4B to D, and in line with the hypothesis that negative recency depends on the interest in the rare event, there was a significant three-way interaction between reward magnitude, run length and the number of observed events so far, $\chi^2(24) = 39.359$, $p = .025$. Finally, there was a significant interaction between event labels and run length, $\chi^2(3) = 8.659$, $p < .034$. Different from reward magnitude, however, this interaction did not concern the strength of negative recency in terms of rare-event expectancy increasing with run length. Instead, when Vase was the common event, the probability of expecting the rare Nothing event generally peaked at run length 2, but when Nothing was the common event, expectancy for the rare Vase event peaked at run length 3 (not shown). Overall, thus, the influence of reward magnitude on choice preference seemed driven by negative recency moderated by reward magnitude. In contrast, event labels seemed to induce a bias of expecting the Vase event in general, but also to expecting common-Vase runs to break earlier than common-Nothing runs.

Given there was noisy a regularity such that a rare event in all conditions most often appeared at a common-event run length of three, a consequent question is whether the negative recency strategy helped participants collecting reward. Thus, as an additional exploration, acknowledging individual differences in these strategies, we applied a method similar to that by Szollosi et al. (2022). Specifically, we assumed that after observing three rare events, participants should stabilize on a trend, which could be negative recency, or one of the two most commonly investigated alternative patterns, namely, positive recency (RL like), or the usually rational strategy of probability maximizing (PC, always Predicting the Common event). We therefore passed runs four to seven into a logistic regression, separately for each individual, estimating the continuous increase in rare-event expectancy with increasing run length (neglecting other factors). We then classified whether a participant enacted negative recency (increasing expectancy), positive recency (decreasing expectancy), or PC (exactly-zero effect, due to never predicting rare). Overall, the shares of the strategies, respectively, were 48%, 35%, and 16%.

To evaluate the success of each strategy, Fig. 5 shows the participants' obtained scores in % of the maximal achievable payoff in Thalers in these trials. As reference, the horizontal lines indicate how many % Thalers would have been possible under two hypothetical strategies, namely, pure maximizing (Predict the Common, PC), and random probability matching with 75% common choices (RPM). As can be seen, participants who exhibited positive recency (RL; squares) underperformed in most cases. Maximizing participants (circles) performed better than an RPM strategy and other participants in the Equal Rewards and Common High condition. However, in the Rare High condition, enacting negative recency (diamonds), indeed, was en par with PC, and even allowed to achieve the best results for a number of participants, even though the sequential regularity was noisy.

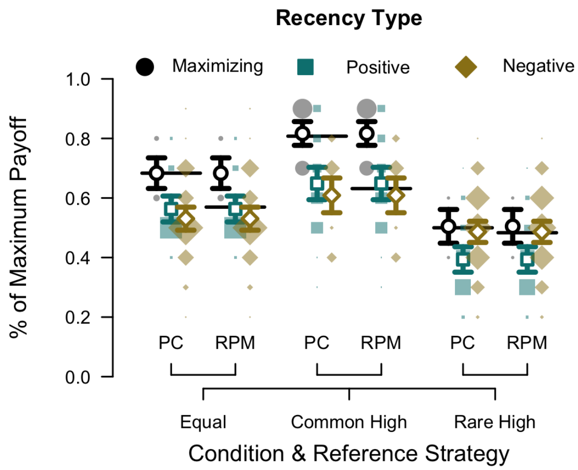

Figure 5. Obtained Payoffs by Condition and Recency. The relative Score (y-axis) indicates individual payoffs relative to the maximum amount of Thalers achievable. Horizontal lines indicate achievable payoffs according to hypothetical strategies, PC (always predict common event / maximizing) and RPM (random probability matching with 75% common responding). Recency Type indicates sub-groups enacting positive or negative recency, or maximizing (PC). Hollow symbols = means with 95%CIs. Colored symbols in background reflect histograms of the individual scores (symbol size reflects participant frequency).

Discussion

The central goal of our study was to understand how reward pursuit during visual stimulus-event conditioning affects learning behavior. We therefore employed a single-cue training task in which participants predicted whether to observe a Vase or Nothing event (e.g., with Vase observed in 75% of trials, being the common event). We manipulated the magnitude of reward for correctly predicting the common and rare outcomes (Equal vs. Common High vs. Rare High), and the label of the rare event (Vase vs. Nothing). In sum, we found that instead of simply using the visual stimulus, most participants seemed to estimate how long the run of the common event was to predict *when* the next rare event will occur. This trend was strongest when the rare event yielded high reward, and weakest when the common event yielded high reward. Whether the rare event was labeled as an event (Vase) or non-event (Nothing), in contrast, revealed a bias in generally expecting more Vases, which was accompanied by an interaction, indicating that participants expected common Vase-runs to end earlier relative to common Nothing runs.

Overall, despite being non-preregistered, the analyses provide support for the hypotheses by Tune (1964) and Estes (1964), that negative recency depends on the interest in the predicted events, highlighting a substantial challenge for widely applied theories of classic conditioning and economic research, involving probabilistic feedback, of which we highlight two. First, similarly discussed by Perruchet (2015), it seems unfortunate that the phenomenon of negative recency is often neglected in studies investigating stimulus-event associations, which seems to alter the theoretical interpretation of learning behavior regarding standard RL. That is, most conditioning studies focus on the question how visual stimulus-outcome associations are formed using average learning curves. However, we suggest, in line with early theories (see Estes, 1964; Restle, 1966; Goodnow & Pettigrew, 1955; Jarvik, 1951), that between-trial variance betrays that most people count event runs instead. In a similar vein, visual stimulus-outcome associations are often also investigated regarding how they generalize to novel stimuli in post-training test phases (or extinction), using so-called generalization gradients (see Shepard, 1987). However, during such tests, human generalization responses have been found to be quite diverse (e.g., Lee et al., 2018; Schlegelmilch, Wills, & Lee, 2023), which might concern whether or not participants actually paid attention to the visual stimulus (see also Mackintosh, 1974) or to run length. Crucially, in test or extinction, event feedback and thereby run length is typically missing. The consequent question is, how individuals having used run length would proceed when the visual stimulus becomes important all of a sudden, perhaps inviting for unsupervised learning, warranting further investigation.

Second, from an economic perspective, our results support the argument that experience-based decisions that involve monetary prospects such as risky gambles (see Szollosi et al., 2022), or probabilistic action rewards (see Feher da Silva et al., 2017), involve the same learning processes as studied in conditioning (see Estes, 1964; Fiorina, 1971), leading to negative recency. In addition, we suggest that the strength of negative recency depends on the magnitude of reinforcement of rare events, which could extend common conceptions of choice preference and risk-taking behavior. Indeed, we agree with Szollosi et al. (2022) to question the typical view of calculating prospects, and instead to consider integrative psychological explanations of often so-called irrational behavior. At minimum, in line with recent arguments (e.g., Schulze et al., 2020), our results suggest that using run lengths to predict outcomes can be considered a rational strategy if there is a possibility that there is a sequential pattern, which would allow outperforming supposedly more rational ones like probability maximizing (e.g., as in our Rare High condition). At least in real life highly rewarding events are also more rare and there might be reasons for those events not reoccurring in closer succession. In general, we believe that integrating insights from conditioning, economic decisions and other areas studying experience-based decisions beyond RL is overdue.

References

Bush, R. R., & Mosteller, F. (1951). A mathematical model for simple learning. *Psychological review*, *58*(5), 313–323. doi:10.1037/h0054388

Dunsmoor, J. E., & Murphy, G. L. (2015). Categories, concepts, and conditioning: How humans generalize fear. *Trends in cognitive sciences*, *19*(2), 73–77.

Estes, W. K. (1964). Probability learning. In *Categories of human learning* (pp. 89–128). Elsevier. doi:10.1016/B978-1-4832-3145-7.50010-8

Feher da Silva, C., Victorino, C. G., Caticha, N., & Baldo, M. V. C. (2017). Exploration and recency as the main proximate causes of probability matching: a reinforcement learning analysis. *Scientific Reports*, *7*, 15326. doi:10.1038/s41598-017-15587-z

Fernberger, S. W. (1920). Interdependence of judgments within the series for the method of constant stimuli. *Journal of Experimental Psychology*, *3*(2), 126–150. doi:10.1037/h0065212

Fiorina, M. P. (1971). A note on probability matching and rational choice. *Behavioral Science*, *16*(2), 158–166. doi:10.1002/bs.3830160206

Goodnow, J. J., & Pettigrew, T. F. (1955). Effect of prior patterns of experience upon strategies and learning sets. *Journal of Experimental Psychology*, *49*(6), 381–389. doi:10.1037/h0049350

Honig, W. K., Boneau, C. A., Burstein, K., & Pennypacker, H. (1963). Positive and negative generalization gradients obtained after equivalent training conditions. *Journal of Comparative and Physiological Psychology*, *56*(1), 111–116. doi:10.1037/h0048683

Jarvik, M. E. (1951). Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *Journal of experimental psychology*, *41*(4), 291–297. doi:10.1037/h0056878

Koehler, D. J., & James, G. (2014). Chapter three - probability matching, fast and slow. In B. H. Ross (Ed.), (Vol. 61, p. 103-131). Academic Press. doi:10.1016/B978-0-12-800283-4.00003-4

Lee, J. C., Hayes, B. K., & Lovibond, P. F. (2018). Peak shift and rules in human generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(12), 1955–1970. doi:10.1037/xlm0000558

Lovibond, P. F., Lee, J. C., & Hayes, B. K. (2020). Stimulus discriminability and induction as independent components of generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(6), 1106–1120. doi:10.1037/xlm0000779

Mackintosh, N. J. (1974). *The psychology of animal learning.* New York, NY: Academic Press.

Myers, J. L. (2014). Probability learning and sequence learning. *Handbook of Learning and Cognitive Processes, ed. WK Estes*, 171–205.

Nicks, D. C. (1959). Prediction of sequential two-choice decisions from event runs. *Journal of Experimental Psychology*, *57*(2), 105–114. doi:10.1037/h0045193

Perruchet, P. (2015). Dissociating conscious expectancies from automatic link formation in associative learning: a review on the so-called perruchet effect. *Journal of Experimental Psychology: Animal Learning and Cognition*, *41*(2), 105–127. doi:10.1037/xan0000060.supp

Plonsky, O., Teodorescu, K., & Erev, I. (2015). Reliance on small samples, the wavy recency effect, and similarity-based learning. *Psychological Review*, *122*(4), 621–647. doi:10.1037/a0039413

Pothos, E. M., & Wills, A. J. (2011). *Formal approaches in categorization.* New York, NY: Cambridge University Press. doi:10.1017/CBO9780511921322

R Development Core Team. (2008). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org` (ISBN 3-900051-07-0)

Rescorla, R. A., & Wagner, A. R. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning II: Current research and theory*, *2*, 64–99.

Restle, F. (1966). Run structure and probability learning: Disproof of Restle's model. *Journal of Experimental Psychology*, *72*(3), 382–389. doi:10.1037/h0023637

Schlegelmilch, R., Wills, A. J., & Lee, J. C. (2023, Sep). *Is generalization general? a comparison of predictive and category learning and the influence of response format and framing.* PsyArXiv. Retrieved from `psyarxiv.com/eakwz` doi:10.31234/osf.io/eakwz

Schlegelmilch, R., Wills, A. J., & von Helversen, B. (2023). Towards re-unifying the theoretical perspectives on conditioning, category learning, probability learning and economic decisions from experience. doi:10.31234/osf.io/h4nu6

Schulze, C., Gaissmaier, W., & Newell, B. R. (2020). Maximizing as satisficing: On pattern matching and probability maximizing in groups and individuals. *Cognition*, *205*, 104382. doi:10.1016/j.cognition.2020.104382

Shanks, D. R., Tunney, R. J., & McCarthy, J. D. (2002). A reexamination of probability matching and rational choice. *Journal of Behavioral Decision Making*, *15*(3), 233–250. doi:10.1002/bdm.413

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323. doi:10.1126/science.3629243

Szollosi, A., Donkin, C., & Newell, B. R. (2022). Toward nonprobabilistic explanations of learning and decision-making. *Psychological Review*, *Advance online publication*. doi:10.1037/rev0000355

Tune, G. (1964). Response preferences: A review of some relevant literature. *Psychological bulletin*, *61*(4), 286–302. doi:10.1037/h0048618