Learning arbitrary stimulus-reward associations for naturalistic stimuli involves transition from learning about features to learning about objects

Shiva Farashahi [1,2]*, Jane Xu [1]*, Shih-Wei Wu[3,4], Alireza Soltani [1]

[1]Department of Psychological and Brain Sciences, Dartmouth College, NH, 03755

[2]Current Address: Flatiron institute, Simons Foundation, New York, NY 10010

[3]Institute of Neuroscience, National Yang-Ming University, Taipei, Taiwan

[4]Brain Research Center, National Yang-Ming University, Taipei, Taiwan

** Equal contribution*

Corresponding author: AS, Department of Psychological and Brain Sciences, Dartmouth College, Hanover NH 03755, soltani@dartmouth.edu

Manuscript information: 5 figures; 2 tables, 34 pages.
Abstract 276 words
Manuscript 7700 words including figure captions and Materials and Methods

Abstract

Most cognitive processes are studied using abstract or synthetic stimuli with specific features to fully control what is presented to subjects. However, recent studies have revealed enhancements of cognitive capacities (such as working memory) when processing naturalistic versus abstract stimuli. Using abstract stimuli constructed from distinct visual features (e.g., color and shape), we have recently shown that human subjects can learn multidimensional stimulus-reward associations via initially estimating reward value of individual features (feature-based learning) before gradually switching to learning about reward value of individual stimuli (object-based learning). Here, we examined whether similar strategies are adopted during learning about naturalistic stimuli that are clearly perceived as objects (instead of a combination of features) and contain both task-relevant and irrelevant features. We found that similar to learning about abstract stimuli, subjects initially adopted feature-based learning more strongly before transitioning to object-based learning. However, there were three key differences between learning about naturalistic and abstract stimuli. First, compared with abstract stimuli, the initial learning strategy was less feature-based for naturalistic stimuli. Second, subjects transitioned to object-based learning faster for naturalistic stimuli. Third, unexpectedly, subjects were more likely to adopt feature-based learning for naturalistic stimuli, both at the steady state and overall. These results suggest that despite the stronger tendency to perceive naturalistic stimuli as objects, which leads to greater likelihood of using object-based learning as the initial strategy and a faster transition to object-based learning, the influence of individual features on learning is stronger for these stimuli such that ultimately the object-based strategy is adopted less. Overall, our findings suggest that feature-based learning is a general initial strategy for learning about reward value of all types of multi-dimensional stimuli.

Keywords: value-based learning, curse of dimensionality, naturalistic tasks, reinforcement learning.

1. Introduction

A hallmark of human cognition is the ability to attribute reward outcomes to cues or events that precede them, or to choices that lead to those reward outcomes. Attributing reward outcomes to stimuli and actions allows the brain to learn and compute the so-called stimulus and action values, respectively, which we collectively refer to as "reward value" for simplicity. Choices faced in the real world, however, are often objects consisting of many different features or attribute dimensions (e.g., color, shape, texture, etc.), each of which could potentially take many values and carry different information about reward outcomes.

Learning about multi-dimensional stimuli is not a trivial problem given that humans and other animals have limited cognitive abilities in terms of the number of features or objects that can be held in working memory or attended at a time. In addition, the set of possible associations grows supra-linearly as the dimensionality of attributes increases, which is often referred to as the "curse of dimensionality" (Barto & Mahadevan, 2003; Diuk, Tsai, Wallis, Botvinick, & Niv, 2013; Hastie, Tibshirani, & Friedman, 2001; Sutton & Barto, 1998). It has been proposed that humans overcome the curse of dimensionality by constructing a simplified representation of the stimuli and learning only a small subset of features (Niv et al., 2015; Wilson & Niv, 2012), or by extracting a set of rules to estimate reward value of options based on their features (Braun, Mehring, & Wolpert, 2010; Dayan & Berridge, 2014; Gershman & Niv, 2010). We have recently shown that during learning about multi-dimensional stimuli, humans initially adopt feature-based learning (i.e., learn reward value of individual features shared between different options) to tackle the curse of dimensionality before gradually transitioning to learning reward value of individual stimuli, which we refer to as object-based learning (Farashahi, Rowe, Aslami, Gobbini, & Soltani, 2018; Farashahi, Rowe, Aslami, Lee, & Soltani, 2017b).

Most studies of reward learning for multi-dimensional stimuli (including ours), however, have focused on abstract stimuli, such as fractals, colored shapes, Gabor patches, etc. (Farashahi et al., 2017b; Niv et al., 2015; Oemisch et al., 2019; Wilson & Niv, 2012; Wunderlich, Beierholm, Bossaerts, & O'Doherty, 2011). These simple stimuli have been adapted to avoid the complexity related to real-world stimuli and better control what is provided to the subjects in the experiments. Although this approach has led to great progress in understanding reward-based learning, it remains unclear whether findings based on abstract stimuli generalize to naturalistic stimuli.

Recent work using naturalistic stimuli has provided evidence that other cognitive abilities such as working memory and visual search are enhanced when processing real-world objects rather than abstract stimuli (Brady, Störmer, & Alvarez, 2016; Brady et al., 2019; Spachtholz & Kuhbandner, 2017). In addition, there is also evidence that naturalistic stimuli can evoke a faster response compared to abstract stimuli (Arntzen & Lian, 2010; Battistoni, Kaiser, Hickey, & Peelen, 2018). These findings are significant because both working memory and visual search can contribute to reward learning. For example, limited capacity of working memory has been shown to decrease the speed of learning (Collins, Brown, Gold, Waltz, & Frank, 2014; Collins, Ciullo, Frank, & Badre, 2017; Collins & Frank, 2012; Otto, Raio, Chiang, Phelps, & Daw, 2013). In addition, analyses of visual search between abstract and naturalistic stimuli suggest that naturalistic stimuli tend to be processed faster because they are perceived to be more salient (Battistoni et al., 2018; Kaiser, Oosterhof, & Peelen, 2016; Thorpe, Fize, & Marlot, 1996). Increased saliency of naturalistic stimuli may lead to more object-based learning when tackling the curse of dimensionality. Together, these findings suggest that using naturalistic stimuli could lead to an overall improvement in learning and/or could bias learning strategy toward object-based learning.

To test these alternative hypotheses and further explore learning about reward value of naturalistic stimuli, here, we examined learning in a multi-dimensional environment that resembles naturalistic settings. Similar to our previous study (Farashahi et al., 2017b), human subjects learned reward value of multi-dimensional visual stimuli through feedback. To construct naturalistic stimuli, we used photos of athletic shoes with color and shoe type as the two task-relevant features. We found that similar to abstract stimuli, subjects initially adopted feature-based learning before systematically transitioning to object-based learning. We also observed three key differences in learning about naturalistic versus abstract stimuli. First, subjects initially adopted the feature-based strategy less often when learning about naturalistic stimuli. Second, the transition from feature-based to object-based learning was faster for naturalistic stimuli. Third, subjects were less likely to use the object-based strategy for naturalistic than abstract stimuli both at the steady state and overall.

2. Materials and Methods

2.1. Subjects

All subjects gave written informed consent prior to participating in the experiment in accordance with the procedures approved by the Dartmouth College Institutional Review Board. No subject had a history of neurological or psychiatric illness. A total of 46 subjects (29 females) were recruited from the Dartmouth College student population (ages 18–22 years). Among them, 23 subjects (15 females) performed two sessions of the experiment that involved learning about naturalistic stimuli only (first cohort of subjects). The other 23 subjects performed the experiment (four sessions) that involved learning both naturalistic and abstract stimuli: they completed two sessions with naturalistic stimuli on one day and two sessions with abstract stimuli on another day (second cohort of subjects). Data in the first cohort of subjects was obtained to compare learning about naturalistic stimuli with our previous study on abstract stimuli (Farashahi et al., 2017b). We then collected data from the second cohort of subjects to perform within-subject comparisons and to have identical task design between naturalistic and abstract stimuli.

Due to the learning nature of our experimental paradigm, we used a performance threshold to exclude subjects whose performance—defined by the average probability of choosing the more rewarding stimulus in each trial—were not distinguishable from chance level. More specifically, we excluded subjects whose average performance was below 0.5439 (equal to 2 s.e.m from chance level of 0.5 based on the average of 576 trials after excluding the first 30 trials of each session). This resulted in the exclusion of 5 out of 23 participants in the first cohort of subjects and 3 out of 23 participants in the second cohort of subjects. The data from the remaining 38 subjects were used for the results presented here. We did not perform data analysis on the excluded subjects due to the small sample size (8 subjects). All data used in this manuscript can be downloaded from http://ccnl.dartmouth.edu/DataShare/NatStiLer.zip.

Subjects were compensated with "t-points" (1 t-point/hour), which are extra credit points for classes within the Department of Psychological and Brain Sciences at Dartmouth College. Based on their performance, subjects were additionally rewarded up to $10 per hour. The experiment was written in MATLAB using the Psychophysics Toolbox Version 3 (Brainard, 1997) and presented using an OLED monitor.

2.2 Stimuli

We used both naturalistic and abstract stimuli. These stimuli were used in both the choice and estimation tasks described below. For naturalistic stimuli, we used pictures of shoes worn for different sports and outdoor activities. These stimuli had two task-relevant features for assigning reward probabilities: type of shoe (soccer shoe, basketball shoe, etc.) and color (blue, red, etc.; **Fig. 1c**). There were two possible sets of naturalistic stimuli (**Fig. 1c** left and right panels), each containing 9 pictures of shoes (3 shoe types × 3 colors). The order in which these two sets were used in the two experimental sessions was randomly determined for each participant.

For abstract stimuli, we used colored shapes similar to those of our previous study (Farashahi et al., 2017b). Specifically, abstract stimuli were drawn from a set of 9 objects that were constructed using combinations of three distinct patterns and three distinct shapes. For each subject and each session, three patterns and shapes were selected randomly and without replacement from a total of eight patterns and eight shapes (**Fig. 1e**). Importantly, we used the same reward probabilities for the task with abstract stimuli and the task with naturalistic stimuli.

2.3. Experimental procedure

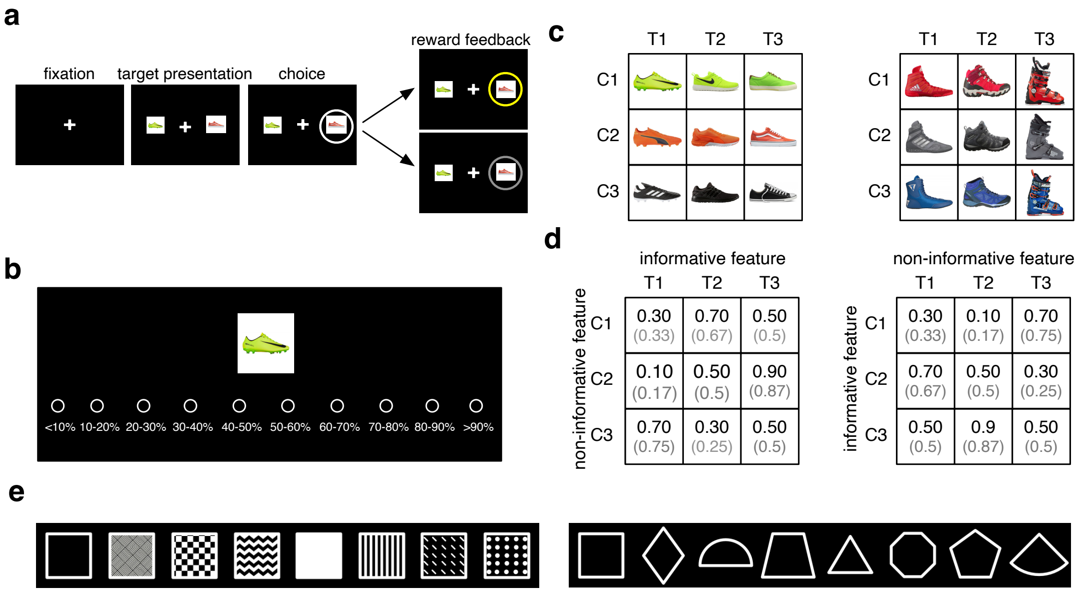
Overall, the experimental paradigm was identical to Experiment 3 in Farashahi et al. (2017b) except that subjects were required to learn a total of 9 (instead of 8) stimuli. Each participant in the first cohort of subjects completed two sessions of the task with naturalistic stimuli in one day. Participants in the second cohort of subjects completed two sessions of the task with naturalistic stimuli one day and two sessions with abstract stimuli on a separate day. The order of stimulus type (abstract and naturalistic) was randomly determined for each participant. Each session lasted about 30 minutes and consisted of 288 choice trials that were interleaved with 8 estimation bouts presented after trials 22, 43, 65, 86, 144, 216, 259, and 288 of the choice task.

In each trial of the choice task, the subjects were presented with a pair of stimuli and were asked to choose the stimulus that they believed would provide the most reward (**Fig. 1a**). The chosen stimulus was rewarded (independently of the other presented object) based on its assigned reward probability (**Fig. 1d**). Of the two features in each stimulus, one feature was partially informative of reward probability associated with the stimuli (e.g., shoe type [T] in the left panel and color [C] in the right panel of **Fig. 1d**), while the other feature was not. Hence, stimulus reward probability could not be determined by combining the reward probability assigned to individual features, resulting in a moderately non-generalizable environment. For example, while stimuli containing T3 feature were overall more rewarding than objects containing T2 feature (left panel in **Fig. 1d)**, stimulus C1T3 was less rewarding than stimulus C1T2. In addition, the average reward probability of stimuli containing a given non-informative feature (in this case C1) was equal to 0.5 (the average reward probability for C1T1, C1T2, and C1T3 objects was equal to 0.5). We constructed a non-generalizable reward environment because a fully generalizable environment is not realistic and could push subjects to solely adopt feature-based learning (Farashahi et al., 2017b). It is worth noting that these reward probabilities were adjusted by a small amount due to a limited number of trials for delivering reward with a certain probability. However, the general structure of reward assignments stayed the same throughout the experiment for the experienced reward.

During each bout of the estimation task, each consisting of 9 trials, subjects provided their estimates of reward probability for each individual stimulus. Possible values for these estimates ranged from 5% to 95% (the average value of each interval shown in **Fig. 1b**) in 10% increments.

2.4. Data analysis and model fitting

To examine the strategy adopted by subjects to estimate reward probabilities associated with different stimuli, we used two methods based on subjects' responses in the estimation trials. First, we fit a Generalized Linear Model (GLM) on subjects' estimates of reward probabilities using the following regressors: the actual reward probability assigned to each stimulus (the object-based regressor), the reward probability calculated by combining reward probability of individual features using the Bayes' theorem (the feature-based regressor; see Eq. 1 in Farashahi et al., 2017b), and a constant. The constant (bias) term in this model quantifies subjects' overall bias in estimating reward probability, and the other two terms determine the influence of feature-based and object-based strategies on probability estimation. We used the ratio of the regression coefficient associated with the object-based regressor to the sum of the regression coefficients associated with the object-based and feature-based regressors to quantify the relative weight of object-based strategy on learning.



non-informative feature	T1	T2	T3
C1	0.30 (0.33)	0.70 (0.67)	0.50 (0.5)
C2	0.10 (0.17)	0.50 (0.5)	0.90 (0.87)
C3	0.70 (0.75)	0.30 (0.25)	0.50 (0.5)

informative feature	T1	T2	T3
C1	0.30 (0.33)	0.10 (0.17)	0.70 (0.75)
C2	0.70 (0.67)	0.50 (0.5)	0.30 (0.25)
C3	0.50 (0.5)	0.9 (0.87)	0.50 (0.5)

Fig. 1. Reward probabilities and stimuli used in the experiment. (**a**) Timeline of the choice trials. In each trial, subjects chose between two options (i.e., shoes that differed in type and color, or two shapes that differed in shape and color) and were provided with feedback on the chosen option. Reward or no reward is indicated by yellow and grey rings, respectively. (**b**) A sample estimation trial. In each estimation trial, subjects estimated the probability of reward associated with a given stimulus by pressing one of ten keys on the keyboard associated with probability ranging from less than 10% to more than 90%. (**c**) Two sets of stimuli used in the experiment comprised of naturalistic stimuli with two task-relevant features (C: color; T: type). The order in which each set of stimuli were assigned to two experimental sessions was pseudo-randomized across subjects. (**d**) Stimulus-reward associations. Two sets of reward probabilities were assigned to the two sets of stimuli shown in (b). For each set of stimuli, only one feature was informative. An informative feature indicates that the average reward probability would change as a function of that feature. Importantly, reward probabilities assigned to the shoes could not be determined by combining the reward probability of individual features and thus, the reward environment was non-generalizable. Numbers in parentheses show the actual probabilities of reward obtained on each stimulus (by the subjects) due to limited resolution for reward assignment. For the set on the left, shoe type was on average informative about reward (average probability of reward = {0.36, 0.5, 0.63}), whereas color was not informative of reward probability (average probability of reward = {0.5, 0.5, 0.5}). The opposite is true for the set on the right. (**e**) The sets of possible patterns (left set) and shapes (right set) used in building abstract stimuli. For each session of the experiment, only three of these shapes were used for a given subject (randomly chosen without replacement). Other aspects of the task and reward probabilities were similar for abstract stimuli and naturalistic stimuli.

Second, to determine whether subjects' probability estimates were closer to estimates based on the feature-based or object-based strategy, we computed the correlation between subjects' estimates and the actual reward probability assigned to each stimulus as well as subjects' estimates and the reward probabilities calculated by combining the reward probability of individual features using Bayes' theorem. We then used the outcome correlation coefficients to determine the fractions of subjects whose estimates follow the feature-based or object-based learning strategy more strongly in each condition or over time. That is, each subject was assigned as a feature-based or object-based learner based on comparing the correlation coefficients mentioned above.

In addition, we fit two GLMs to test the effect of stimulus type (abstract vs. naturalistic), time (trial number), and the interaction of stimulus type with time. First, we performed a logistic regression analysis to predict the fraction of subjects whose estimates were more correlated with actual reward probabilities than reward probabilities calculated based on features and subjects' estimates of reward probabilities, using time and stimulus type as independent variables. Second, we fit a GLM on the difference between two correlation coefficients: the correlation between subjects' estimates and object-based predictions, and the correlation between subjects' estimates and feature-based predictions. The regressors in this model were time (trial number) and stimulus type (natural or abstract stimulus). For both models, we also considered the interaction of time and stimulus type.

To estimate the time course of performance as well as the time course of relative weight and fraction of subjects, we fit data using an exponential function based on the following equation:

$$y(t) = \ y_{ss} - (y_{ss} - y_0)exp^{(\frac{-t}{\tau})} \qquad \text{(Eq. 1)}$$

where y_0 and y_{ss} are the initial and steady-state values of performance, τ is the time constant for approaching steady state, and t represents the trial number in a session.

Finally, we also used six different reinforcement learning (RL) models based on object-based or feature-based learning strategies to fit individual subjects' choice behavior in order to identify the learning strategy adopted by each subject (see below for more details). These models were fit to experimental data by minimizing the negative log likelihood (LL) of the predicted choice probability given different model parameters using the '*fminsearch*' function in MATLAB (MathWorks, Inc., Natick, MA). To avoid finding local minima for the fit of experimental data, we repeated fitting of each dataset with at least 10 different sets of initial parameters and picked the best fit. Based on the examination of the fitting results, we found 10 initializations to be sufficient to avoid local minima. We performed model comparison using both Akaike information criterion (AIC) and Bayesian information criterion (BIC). The smaller value for each measure indicates a better fit of choice behavior.

In addition, to compare the ability of different models in fitting choice behavior over time, we also used AIC and BIC per trial (Farashahi et al., 2018), denoted as $\mathrm{AIC_p}$ and $\mathrm{BIC_p}$:

$$AIC_p(t) = -2LL(t) + 2k/N_{trials} \quad \text{(Eq. 2)}$$

$$BIC_p(t) = -2LL(t) + 2klog\,(N_{trials})/N_{trials} \quad \text{(Eq. 3)}$$

where k indicates the number of parameters in a given model, *t* represents the trial number, $LL(t)$ is the log-likelihood in trial *t*, and N_{trials} is the number of trials in the experiment. The logic behind these definitions is that penalties included in AIC and BIC are based on the sum of the log likelihoods over all trials (data), and thus, by dividing the penalty terms by the number of trials we ensure that the sum of AICp(t) and BICp(t) over all trials would be equal to AIC and BIC, respectively. The smaller values for these measures indicate a better fit of choice behavior. As we show here, these measures can be used to detect a transition between feature-based and object-based learning.

Finally, to confirm our results based on AIC and BIC, we applied the variational Bayesian model selection (BMS) approach in order to identify the most likely models that could account for our data. Specifically, the BMS approach treats different models as random variables and estimates the parameters of a Dirichlet distribution, which describes the probabilities from which models are sampled across all subjects. These probabilities translate to the probability of one model being more likely than any other model (Stephan et al., 2005). To avoid overfitting of data and reducing the effect of outliers, we randomly sampled 80% of the data to estimate the likelihoods and repeated this procedure 50 times to calculate the average likelihood of all models. All behavioral analyses and model fitting were done using custom codes written in MATLAB 2018a (MathWorks, Inc.).

2.4.1. Object-based RL models

Using standard RL models (Sutton & Barto, 1998), the reward value of each stimulus was estimated based on reward feedback following the subjects' choice in each trial. In the context of this study, reward value is equal to the reward probability associated with each stimulus. We fitted two types of models, referred to as uncoupled object-based RL and coupled object-based RL. In the uncoupled object-based RL, only the reward value of the chosen object was updated in each trial. This update was done via separate learning rates for rewarded or unrewarded trials using the following equation:

$$V_{chosS}(t+1) = V_{chosS}(t) + \alpha_{rew}\big(1 - V_{chosS}(t)\big)\,, \quad if\; r(t) = 1$$

$$V_{chosS}(t+1) = V_{chosS}(t) - \alpha_{unr}\big(V_{chosS}(t)\big), \quad if\; r(t) = 0 \qquad \text{(Eq. 4)}$$

where *t* represents the trial number, V_{chosS} is the estimated reward value of the chosen stimulus, $r(t)$ is the trial outcome (1 for a rewarded outcome, 0 for an unrewarded outcome), and α_{rew} and α_{unr} are the learning rates for rewarded and unrewarded trials. The value of the unchosen stimulus is not updated in this model.

In the coupled object-based RL, reward values of both stimuli presented in a given trial were updated, but in opposite directions (if the subject incorrectly assumes that reward assignments on the two stimuli are anti-correlated). That is, while the reward value of the chosen object was updated based on Eq. (4), the value of the unchosen stimulus was updated based on the following equation:

$$V_{uncS}(t+1) = V_{uncS}(t) - \alpha_{rew}\big(V_{uncS}(t)\big), \quad if\; r(t) = 1$$

$$V_{uncS}(t+1) = V_{uncS}(t) + \alpha_{unr}\big(1 - V_{uncS}(t)\big), \; if\; r(t) = 0 \qquad \text{(Eq. 5)}$$

where V_{uncS} is the estimated reward value of the unchosen stimulus.

The estimated value functions were then used to compute the probability of choosing between the two stimuli in a given trial based on a logistic function:

$$logit\ P_L(t) = \big(V_L(t) - V_R(t)\big)/\sigma + bias \qquad \text{(Eq. 6)}$$

where P_L is the probability of choosing the stimulus presented on the left, V_L and V_R are reward values of the stimuli presented to the left and right, respectively, *bias* measures a response bias toward the left option to capture the subject's location bias, and σ is a parameter measuring the level of stochasticity in decision-making processes.

2.4.2. Feature-based RL models

In this set of models, reward value of each stimulus is computed by combining reward values of the features of that object, which are estimated from reward feedback using a standard RL model. The updating rules for the feature-based RL models are identical to the object-based RLs described above except that the reward value of the chosen (unchosen) stimulus is replaced by reward values of the features of the chosen (unchosen) stimulus.

As with the object-based RL models, the probability of choosing a stimulus is determined based on the logistic function of the difference between the estimated values for the stimuli presented:

$$logit\ P_L(t) = w_{shape}\big(V_{shapeL}(t) - V_{shapeR}(t)\big) + w_{color}(V_{colorL}(t) - V_{colorR}(t)) + bias \qquad \text{(Eq. 7)}$$

where V_{shapeL} (V_{colorL}) and V_{shapeR} (V_{colorR}) are reward values associated with the shape (color) of the left and right stimuli, respectively, *bias* measures a response bias toward the left option to capture any location bias, and w_{shape} and w_{color} determine the influence of the two features on the final choice as well as the overall stochasticity in choice (larger values of weights correspond to smaller stochasticity in choice). Note that these weights can be assumed to be learned over time through reward feedback (as in our models; see *RL models with decay* below) or could reflect differential processing of the two features due to attention.

2.4.3. RL models with decay

Additionally, we investigated the effect of "forgetting" reward values of the unchosen (or not-presented) stimuli or features by introducing a decay in reward values. This feature has been shown to capture some aspects of learning (Ito & Doya, 2009), especially in multi-dimensional tasks. More specifically, reward values of the unchosen or not-presented stimuli or features decayed to 0.5 with a rate of d $(0 < d < 1)$ as follows:

$$V(t+1) = V(t) - d * (V(t) - 0.5) \qquad \text{(Eq. 8)}$$

where t represents the trial number and V is the estimated reward value of an unchosen stimulus or feature.

2.4.4. Hybrid RL model

To show that AICp(t) and BICp(t) can be used to detect a transition between feature-based and object-based learning, we performed additional simulations using a hybrid RL model. In this model, the subjective value of each option is the weighted sums of two sets of values: values based on a feature-based RL model with decay and values based on an object-learning RL model with decay. As a result, the probability of choosing between the two stimuli is equal to:

$$logit\ P_{\mathrm{L}}(t) = w(t)\left(\left(V_{shape\mathrm{L}}(t) - V_{shape\mathrm{R}}(t)\right)/2 + \left(V_{color\mathrm{L}}(t) - V_{color\mathrm{R}}(t)\right)/2\right)$$

$$+(1 - w(t))\left(V_{\mathrm{L}}(t) - V_{\mathrm{R}}(t)\right) \qquad \text{(Eq. 9)}$$

where $w(t)$ is the relative weight of the object-based to the feature-based component. The relative weight of the object-based to feature-based component monotonically increases over time as follows:

$$w(t) = \ w_{ss} - (w_{ss} - w_0)exp^{(\frac{-t}{\tau})} \qquad \text{(Eq. 10)}$$

where w_0 and w_{ss} are the initial and steady state values, and τ is the time constant. We set the w_0, w_{ss}, τ, $\alpha_{rew,}$ and α_{unr} to 0.3, 0.7, 100, 0.2, and 0.1, respectively, to match the observed choice behavior of the subjects in our experiments. We used this hybrid model to simulate choice behavior in our experiment with the same task parameters used for subjects.

3. Results

3.1. Effects of stimulus type on performance

We first examined performance (probability of choosing the more rewarding stimuli) to determine whether participants were able to perform the choice task correctly. We found that in choice trials, the average performance was significantly above chance level (abstract stimuli: mean±std = 0.60±0.08; naturalistic stimuli [first subject cohort]: mean±std = 0.61±0.05; naturalistic stimuli [second cohort of subjects]: mean±std = 0.62±0.07). Performance in each cohort of subjects and across all subjects quickly increased and plateaued after about 100 trials (**Fig. 2a, d, g**). These results demonstrate that participants were engaged in the task and were able to select the stimulus with a higher probability of reward in most trials.

We then compared the dynamics of learning for naturalistic and abstract stimuli by fitting the time course of performance with an exponential function (see Materials and Methods). We found that for the second cohort of subjects who performed the task with both naturalistic and abstract stimuli, the performance reached its steady-state at a faster rate for the naturalistic rather than abstract stimuli (naturalistic: τ = 52 trials, CI = [31, 62]; abstract: τ = 85 trials, CI = [76, 94]). However, the steady-state performance was not significantly different between the two types of stimuli in this cohort of subjects (0.65 and 0.63 for abstract and naturalistic stimuli, respectively; **Fig. 2a, d**).

We found similar results when considering data from both cohorts of subjects. More specifically, the subjects reached the steady-state performance at a faster rate when learning about naturalistic stimuli (naturalistic: τ = 60 trials, CI = [49, 71]; abstract: τ = 85 trials, CI = [75, 95]), whereas the steady states were not different between the two types of stimuli (equal to 0.64 and 0.65 for naturalistic and abstract stimuli, respectively; **Fig. 2g**).

Finally, we also fitted a Generalized Linear Model (GLM) on the overall performance in order to test for possible transfer of knowledge between the two sessions of the experiment. However, this analysis did not reveal any effect of stimulus type (abstract vs. naturalistic stimuli) or session number (first vs. second) on performance.

3.2. Subjects' estimates reveal the effects of stimulus type on learning strategy

Next, we used two GLMs to examine the effects of stimulus type (abstract vs. naturalistic), time (trial number), and their interaction on the subjects' probability estimates throughout the experiments (see Materials and Methods). First, we performed a logistic regression analysis on the fractions of subjects whose estimates were more correlated with actual reward probabilities than reward probabilities calculated based on features and subjects' estimates of reward probabilities. Second, we used a multiple regression model to predict the difference in the correlations of subjects' estimates and object-based predictions and subjects' estimates and feature-based predictions. Using these analyses, we found significant effects of time and stimulus type, suggesting an overall larger value for abstract than naturalistic stimuli and an increase in the use of object-based strategy over time (**Table 1**). Additionally, we found negative but non-significant regression coefficients for the interaction of time and stimulus type in both models. This result suggests that learning about abstract and naturalistic stimuli follow different time courses.

To further investigate the possible interaction between stimulus type and time, we explored the adoption of the two learning strategies over the course of the experiment. First, using estimates of reward probabilities, we confirmed the previously observed transition from feature-based to object-based learning (**Fig. 2b, e**). More specifically, using a GLM to predict subjects' estimates, we found that for the abstract stimuli, the relative weight of the object-based strategy (i.e., the weight of the object-based divided by the sum of the weights for the object-based and feature-based strategies) was smaller than 0.5 during the initial estimation bouts and gradually increased and became larger than 0.5 over time (relative weight for the first two estimation bouts = 0.33, 95% CI = [0.24, 0.45], $p = 0.04$, $d = 0.29$, $N = 40$; relative weight for the last two estimation bouts = 0.80, 95% CI = [0.7, 0.9], $p = 0.01$, $d = 0.38$, $N = 40$; **Fig. 2b**). We found a similar pattern for learning with naturalistic stimuli (relative weight for the first two estimation bouts = 0.22, 95% CI = [0.15, 0.29], $p = 0.02$, $d = 0.46$, $N = 40$; relative weight for the last two estimation bouts = 0.62, 95% CI = [0.55, 0.68], $p = 0.03$, $d = 0.32$, $N = 40$; **Fig. 2e**). This result shows that initially, subjects' estimates of reward probabilities were more strongly influenced by the feature-based strategy but later on, were more affected by the object-based strategy for learning and computing reward probabilities of different stimuli.

Independent variables for predicting fraction of subjects	stimulus type (abstract vs. naturalistic)	time (trial #)	stimulus type×time
Regression coefficients and corresponding p-values	0.37±0.13 p = 0.016	0.004±0.001 p = 0.013	-0.004±0.0004 p = 0.09

Independent variables for predicting difference in correlation of subjects' estimate with object-based vs. feature-based learning	stimulus type (abstract vs. naturalistic)	time (trial #)	stimulus type×time
Regression coefficients and corresponding p-values	0.26±0.11 p = 0.033	0.003±0.001 p = 0.025	-0.004±0.0005 p = 0.11

Table 1. The effects of time and stimulus type on subjects' estimates of reward probabilities. Reported are the values of regression coefficients (mean±s.e.m.) and corresponding p-values for a logistic regression model that predicts the fraction of subjects whose estimates were more strongly correlated with object-based than feature-based predictions (top), and a linear regression model that predicts the difference between the correlation of subjects' estimates and object-based predictions and the correlation of subjects' estimates and feature-based predictions. In both models, we used stimulus type, time, and the interaction of stimulus type with time as regressors.

Consistent with these results, correlation analysis revealed that during the first two estimation bouts, the probability estimates of less than half of the subjects were more correlated with the actual reward probabilities than the reward probabilities calculated based on feature values, but this fraction increased over time for both abstract stimuli (comparison of fractions in first two estimation vs. last two estimation bouts: $\chi2$ (1) = 12.25, $p = 4.6\times10^{-4}$, N = 40; **Fig. 2c**) and naturalistic stimuli (comparison of fractions in first two vs. last two estimation bouts: $\chi2$ (1) = 5.85, p = 0.015, N = 40; **Fig. 2f**).

We found similar results when considering data from all subjects who performed the learning task with naturalistic stimuli (first and second cohorts of subjects). The relative weight of the object-based strategy was smaller than 0.5 during the initial estimation bouts but gradually increased and became larger than 0.5 (relative weight for the first two estimation bouts = 0.35, 95% CI = [0.31, 0.40], p = 0.02, d = 0.41, N = 76; relative weight for the last two estimation bouts = 0.64, 95% CI = [0.62, 0.69], p = 0.01, d = 0.58, N = 76; **Fig. 2h**). Additionally, the probability estimates of less than half of the subjects were more correlated with the actual reward probabilities than the reward probabilities calculated based on features, and this fraction increased over time (comparison of fractions in first two vs. last two estimation bouts: $\chi2$ (1) = 47.04, $p = 7.0\times10^{-12}$, N = 76; **Fig. 2i**).

To identify the similarities and differences between learning about naturalistic and abstract stimuli, we next compared our measures within participants in the second cohort of subjects. Comparison between the relative weights of object-based strategy on estimated reward probabilities did not reveal any significant difference between the initial estimation bouts (difference in the relative weight of object-based strategy between abstract and naturalistic stimuli = 0.11, 95% CI = [-0.02, 0.23], p = 0.11, d = 0.12, N = 40). However, we found the relative weight of the object-based strategy during the last two estimation bouts to be significantly larger for abstract than naturalistic stimuli (difference = 0.18, 95% CI = [0.09, 0.26], p = 0.03, d = 0.26, N = 40). Comparison between the relative weights of the object-based term on probability estimates during the first two estimation bouts revealed that subjects' initial strategy was not different between learning about naturalistic and abstract stimuli (difference in the relative weight of the object-based strategy between abstract and naturalistic stimuli = 0.02, 95% CI = [-0.07, 0.11], p = 0.27, d = 0.09, N = 76). Moreover, the relative weight of the object-based strategy during the last two estimation bouts was larger for abstract than for naturalistic stimuli (difference = 0.15, 95% CI = [0.03, 0.23], p = 0.04, d = 0.13, N = 76).

By examining the fraction of subjects whose reward-probability estimates were more correlated with actual reward probabilities associated with the stimuli than the reward probabilities calculated based on features, we found that in the early stages of learning, a larger fraction of subjects followed an object-based strategy for naturalistic rather than abstract stimuli (the difference in fractions between naturalistic and abstract stimuli during the first two estimation bouts = 0.21, $\chi2$ (1) = 6.25, p = 0.01, N = 40). We also found that toward the end of the experiment, a slightly larger proportion of subjects provided probability estimates that were more strongly correlated with the object-based strategy when learning about abstract stimuli (the difference in fraction between naturalistic and abstract stimuli during the last two estimation bouts = -0.19, $\chi2$ (1) = 7.34, p = 0.0016, N = 40).

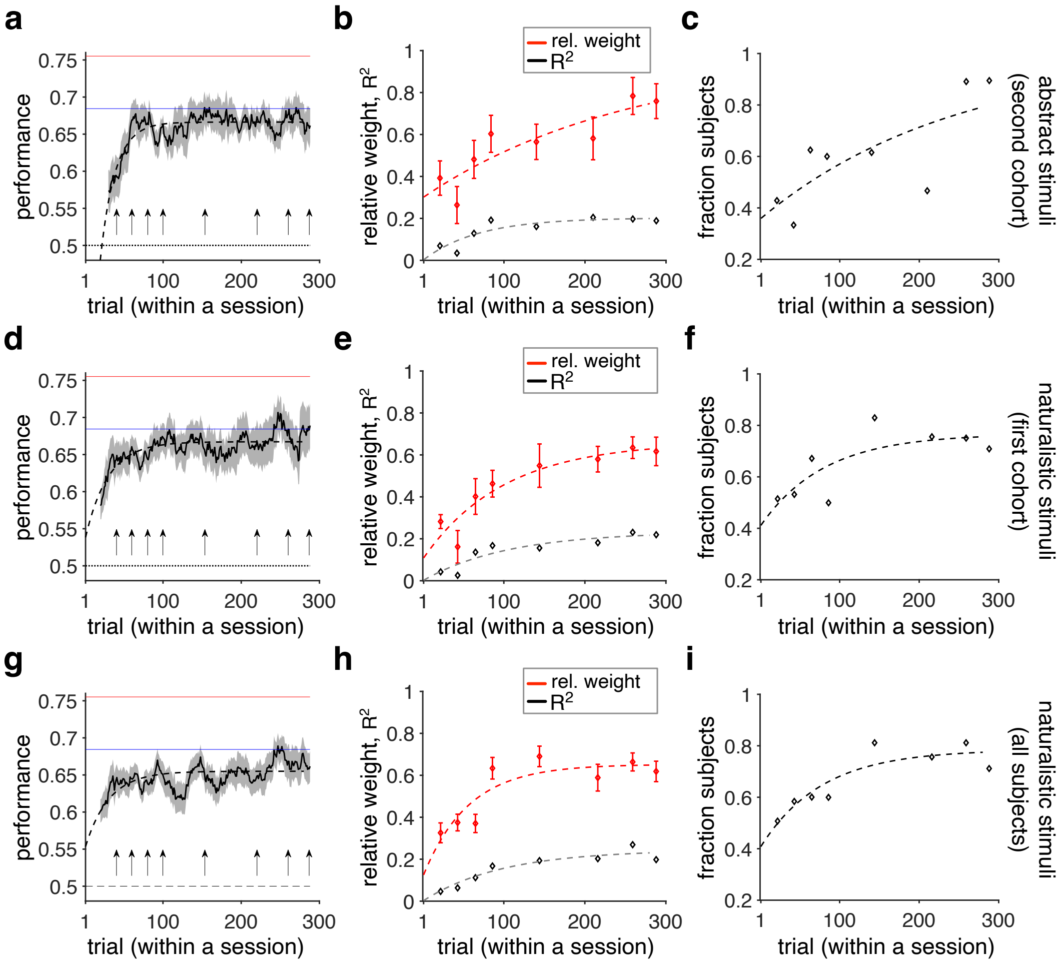


Figure 2. Transition from feature-based to object-based learning occurs faster when learning from naturalistic stimuli. (**a**) The time course of performance for learning abstract stimuli. Plotted is the probability of choosing the more rewarding option in each trial (shaded areas indicate s.e.m.). The dotted line shows chance performance and the dashed line shows the fit of data based on an exponential function. The red and blue solid lines show the maximum performance using the feature-based and object-based RLs, respectively, assuming that the decision maker selects the more rewarding option based on a model approach in every trial. Arrows mark the locations of estimation bouts throughout a session. (**b**) The time course of the strategy used to estimate reward probabilities based on fitting subjects' estimates of reward probabilities. Plotted is the relative weight of object-based to the sum of the object-based (in red) and feature-based terms and explained variance in estimates (R^2, black curve) over time. The error bars demonstrate the confidence interval and the dashed lines show extrapolation based on an exponential fit. **(c)** The fraction of subjects who showed a stronger correlation of probability estimates with actual reward probabilities than with the probabilities estimated using reward probabilities of stimuli's features. The dashed line shows extrapolation based on an exponential fit. **(d–f)** Similar to panels a–c but for learning from naturalistic stimuli in the same cohort of subjects. (**g–i**) Similar to panels a–c but for learning from naturalistic stimuli across all subjects (first and second cohorts of subjects).

These results hold when considering data from both cohorts of subjects. That is, in the early stages of learning, a larger fraction of subjects followed the object-based strategy for naturalistic stimuli (the difference in fraction between naturalistic and abstract stimuli during the first two estimation bouts = 0.23, $\chi 2$ (1) = 8.65, p = 0.0034, N = 76). We also confirmed that toward the end of the experiment, a slightly larger fraction of subjects provided probability estimates that were more strongly correlated with the object-based strategy when learning about abstract stimuli (difference in fractions between naturalistic and abstract stimuli during the last two estimation bouts = -0.16, $\chi 2$ (1) = 9.34, p = 0.0029, N = 76).

Together, the results from the above analyses indicate that subjects transitioned from primarily using feature-based learning to object-based learning for both types of stimuli. An interesting difference between learning about abstract and naturalistic stimuli was that, even though the subjects initially adopted a more object-based strategy when learning about naturalistic stimuli, they reached a higher level of object-based learning for abstract stimuli.

Having these results, we next compared the rate of transition from feature-based to an object-based strategy for naturalistic and abstract stimuli using our measures. First, we fit the relative weight of the object-based term over time (based on an exponential function) and found that subjects transitioned to object-based learning at a faster rate when learning about naturalistic stimuli (τ = 227, 85 and 65 trials for abstract and naturalistic stimuli in the second cohort of subjects and naturalistic stimuli across all subjects, respectively; **Fig. 2b, e, h**). Moreover, to estimate errors related to the reported time constants, we fitted a GLM to individual subjects' estimates of reward probabilities. Consistent with our previous results, we found that subjects transitioned to object-based learning at a faster rate when learning about naturalistic stimuli (τ = 216, 95% CI = [277, 158], 93, 95% CI = [132, 54], and 52, 95% CI = [24, 79] trials for abstract and naturalistic stimuli in the second cohort of subjects and naturalistic stimuli across all subjects, respectively). However, we note that fitting GLM this way, as opposed to fitting GLM to all subjects' estimates of reward probabilities, is prone to serious overfitting (3 parameters for fitting 9 data points).

Similarly, the fraction of subjects with a stronger correlation between estimated reward probabilities and actual reward probabilities reached a plateau faster for naturalistic stimuli (τ = 207, 80, and 76 trials for the time constant of abstract and naturalistic stimuli in the second cohort of subjects and naturalistic stimuli across all subjects, respectively; **Fig. 2c, f, i**). Together, results based on different types of measures illustrate that subjects learned at a faster rate when faced with naturalistic stimuli compared to abstract stimuli.

Next, to assess if feature identity impacted learning, we compared the subjects' assignment of reward probabilities between the two task-relevant features: color and shoe type. This is because color is a low-level visual feature compared to shoe type, which is a high-level concept. Therefore, we compared learning between sessions when either color or shoe type was the informative feature within individual subjects (each subject performed the task twice, once with color and once with shoe type as the informative feature). However, we did not find any significant difference in the relative weight of the object-based term between color or shoe type as the informative feature (the difference in estimated weights between color and shoe type = 0.05, CI = [-0.12, 0.15]; **Fig. 3a**). Moreover, when comparing the fraction of subjects whose reward-probability estimates were more correlated with actual reward probabilities than reward probabilities calculated based on features, we did not find any evidence for the type of informative feature in any of the estimation bouts (the difference in the fraction of subjects for color and shoe type = 0.10, $\chi2$ (1) < 2.51, $p > 0.11$; **Fig. 3b**).

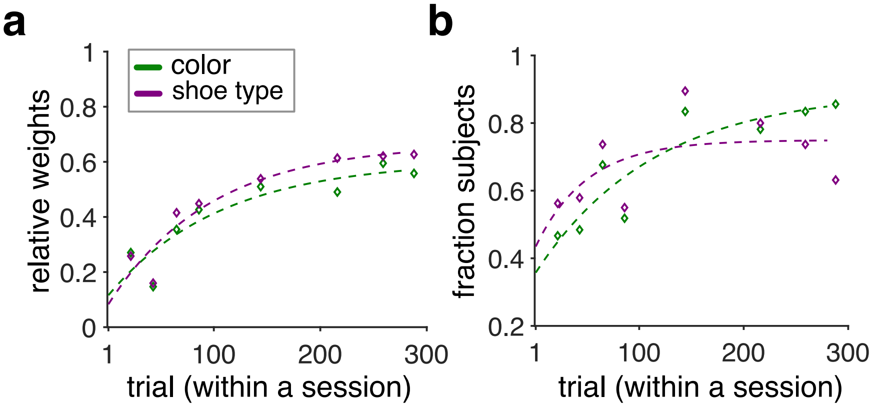


Figure 3. Transition from feature-based to object-based learning was not different between the sessions with color and shoe type as the informative features. (**a**) Plotted is the relative weight of object-based to the sum of the object-based and feature-based terms for sessions with color and shoe type as the informative features. Dashed lines show the fit of data based on an exponential function that allows extrapolation over the entire course of the experiment. (**b**) The fraction of subjects who showed a stronger correlation between reward-probability estimates and actual reward probabilities than the probabilities estimated using reward values of features for sessions with color and shoe type as the informative features.

3.3. Choice behavior reveals the effects of stimulus type on learning strategy

To identify the learning strategy adopted by the subjects during choice trials, we fit the choice data using six different RL models that relied on either object-based or feature-based approaches for updating reward probabilities. Specifically, in uncoupled feature-based RL models, the features associated with the selected stimulus are updated. In coupled feature-based RL models, however, the features associated with both the chosen and unchosen stimuli are updated (with the assumption of anti-correlated reward assignment). Similarly, in uncoupled object-based models, only the reward probability of the selected stimulus is updated, whereas the reward probabilities of both chosen and unchosen stimuli are updated in coupled object-based models. In RL models with decay, the reward probabilities of unchosen stimuli or features are lost over time (see Materials and Methods). For model comparison, we used goodness-of-fit measures in terms of AIC and BIC.

We found the best object-based and feature-based models to be those that incorporate the decay in value estimates over time (**Table 2**). More importantly, the object-based with decay model provided a significantly better fit than the feature-based with decay model when learning about naturalistic stimuli across all subjects and abstract stimuli in the second cohort of subjects (naturalistic stimuli: two-sided sign-rank test; BIC [feature-based with decay] − BIC [object-based with decay]: mean±s.e.m. = 45.7±18.2, $p = 0.02$, $N = 38$, $d = 0.65$; AIC [feature-based with decay] − AIC [object-based with decay]: mean±s.e.m. = 44.1±18.1; $p = 0.02$, $N = 38$, $d = 0.62$, abstract stimuli: BIC [feature-based with decay] − BIC [object-based with decay]: mean±s.e.m. = 48.4±19.1, $p = 0.04$, $N = 20$, $d = 0.38$; AIC [feature-based with decay] − AIC [object-based with decay]: mean±s.e.m. = 47.4±19.2; $p = 0.04$, $N = 20$, $d = 0.35$).

a)

Naturalistic stimuli						
Model	Coupled feature-based	Uncoupled feature-based	Feature-based with decay	Coupled object-based	Uncoupled object-based	Object-based with decay
# pars.	5	5	6	4	4	5
-LL	363.1±8.5	370.9±9.5	351.5±9.0	359.5±8.9	363.5±9.8	330.4±12.1*
AIC	736.1±16.9	751.8±19.1	715.0±18.0	727.0±17.8	734.9±18.6	670.9±24.2*
BIC	744.2±16.9	759.9.0±19.1	724.7±18.0	733.5±17.8	741.4±18.6	679.0±24.2*

b)

Abstract stimuli						
Model	Coupled feature-based	Uncoupled feature-based	Feature-based with decay	Coupled object-based	Uncoupled object-based	Object-based with decay
# pars.	5	5	6	4	4	5
-LL	365.7±10.1	372.0±10.4	354.8±9.9	361.4±8.9	367.6±11.0	332.1±11.6*
AIC	741.4±20.2	754.1±20.8	721.7±18.8	730.8±17.8	743.3±22.1	674.3±23.2*
BIC	746.4±20.2	759.1±20.8	727.7±18.8	734.8±17.8	747.3±22.1	679.3±23.2*

Table 2. Comparison of the goodness-of-fit measures. The object-based model with decay provides the best fit to the choice data. Reported are three measures for the goodness-of-fit, negative log likelihood (-LL), Akaike information criterion (AIC), and Bayesian information criterion (BIC) averaged over subjects (mean±s.e.m.) for three feature-based RLs and their object-based counterparts when learning from naturalistic stimuli across all subjects (a) and abstract stimuli in the second cohort of subjects (b). A smaller value indicates a better fit. The model providing the best fit in a given experiment and its object-based or feature-based counterpart are highlighted in cyan and orange, respectively. Each feature-based RL was compared with its object-based counterpart using a two-sided, sign-rank test, and (*) indicates the difference is significant at $p < 0.05$.

Additionally, we applied the variational Bayesian model selection (BMS) approach to identify the most likely models that could account for our data. We found the most likely object-based and feature-based models to explain the data were those incorporating the decay. More importantly, the object-based with decay model was more likely than the feature-based with decay model (**Fig. 4a-b**). Therefore, across all models, choice behavior was best accounted for by an object-based RL with decay, suggesting that subjects learned the reward probability of the chosen stimulus and forgot the reward probability of the unchosen and non-presented stimuli. These results illustrate that overall subjects' choice behavior was more compatible with an object-based strategy for learning.

To capture a potential change in the best model that accounted for choice data over time, we also computed the BICp and AICp over time for the best object-based and feature-based models (see Materials and Methods). We found that the object-based model provided a better fit mainly in the later stage of the experiment (**Fig. 4c, d**). The difference between the goodness-of-fit for the object-based and feature-based models was significantly different between early (1–50) and late (50–288) trials (Δ $BIC_p = \Delta$ $AIC_p = \Delta$ (-LL): mean±std = 0.14 ± 0.09; two-sided sign-rank test; $p = 0.03$, $d = 0.94$, $N = 38$). We note that the boundary for early versus late trials (at 100) was selected based on the time course of performance (**Fig. 2a, d, g**) but that the reported difference was significantly larger than zero ($p < 0.05$) for any boundary values between 80 and 120 as well. However, comparing naturalistic with abstract stimuli (**Fig. 4c, d)**, the difference in goodness-of-fit for the object-based and feature-based models between early and late trials was not significant (Δ $BIC_p = \Delta$ $AIC_p = \Delta$ (-LL): mean±std = 0.03±0.07; two-sided rank-sum test; $p = 0.28$, $d = 0.25$, $N = 58$). This observation can be explained by the fact that models are fit to the choice data from all trials. Therefore, fitting provides a set of parameters that captures choice behavior the best on average, and therefore change in behavior is not captured best in this measure. Together, results based on fitting choice behavior illustrate that similar to abstract stimuli, subjects transitioned from feature-based to object-based strategy during the time course of the experiment.

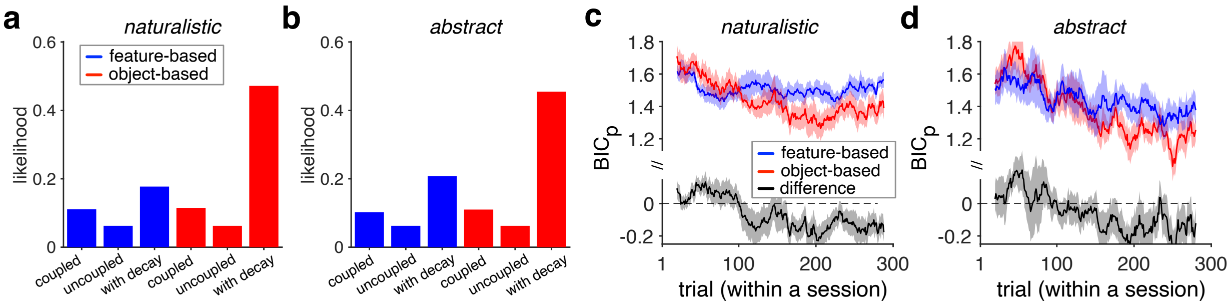


Fig. 4. Goodness-of-fit based on two best models shows similar transitions from feature-based to object-based learning for naturalistic and abstract stimuli. (**a–b**) Likelihood of different strategies adopted by humans when learning about naturalistic stimuli across all subjects (a) and abstract stimuli in the second cohort of subjects (b). Fitting choice behavior shows that subjects' choice behavior was more likely to be explained by an object-based learning strategy. (**c–d**) Plotted is the average BIC per trial across all subjects based on the feature-based model with decay, the object-based RL model with decay, and the difference between object-based and feature-based models when learning from naturalistic stimuli across all subjects (c) and abstract stimuli in the second cohort of subjects (d). We did not observe any significant difference between learning about naturalistic and abstract stimuli based on the goodness-of-fit measure.

We have previously used LL(t), AICp(t), and BICp(t) to compare competing models in terms of their ability to capture choice after a sequence of trials (Farashahi et al., 2017a) and at a given point in time during a session (Farashahi et al., 2017b, Farashahi et al., 2018). Nonetheless, we performed additional simulations to show that these measures can capture a transition between feature-based and object-based learning. More specifically, we simulated 50 instances of choice behavior in a hybrid model that includes both feature-based learning with decay and object-based learning with decay components and in which the relative weight of these two components continuously changes over time (see Materials and Methods for more details). We then fit the simulated choice data using an object-based model with decay and a feature-based model with decay and computed BICp(t) for fit based on these two models. We found that BIC$_p$(t) can detect the transition from feature-based to object-based learning at about the same time point (~70 trials) as when the object-based component became stronger than the feature-based component (i.e., when $w(t) > 0.5$; **Fig. 5**). This result shows that BICp (and similarly AICp) can detect a transition in the learning strategy over time.

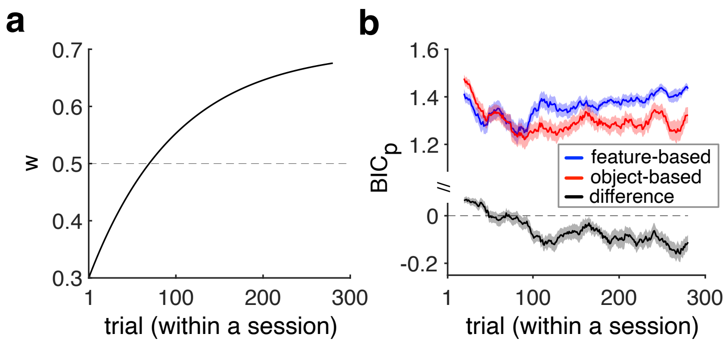


Figure 5. BICp can be used to detect transition from feature-based to object-based learning from simulated data. (**a**) Plot shows the relative weight of the object-based to the feature-based component in the hybrid model used to generate the simulated choice data. (**b**) Plot shows the average BICp(t) across all subjects for fit of simulated data based on the feature-based model with decay, the object-based model with decay, and the difference between BICp(t) in the two models. At the beginning of the session, the feature-based model provides a better fit as reflected in a smaller BICp, but later on, the object-based model provides a better fit. Importantly, the difference in BICp changes sign around the same time point (~70 trials) as the relative weight of the object-based to the feature-based component in the simulated data passes 0.5. This result shows that BICp can be used to detect a transition between different learning strategies over time.

4. Discussion

In this study, we investigated learning about reward value of naturalistic stimuli based on feedback in multi-dimensional reward environments. We confirmed our previous observations using abstract stimuli (Farashahi et al., 2017b) but also found significant differences between learning naturalistic and abstract stimuli. More specifically, our subjects initially adopted a feature-based learning strategy more strongly and slowly transitioned to an object-based strategy as they gained more experience through reward feedback. However, we found that compared with abstract stimuli, subjects initially adopted a less feature-based strategy and transitioned to an object-based strategy faster when learning about naturalistic stimuli. These findings validate our previous results that feature-based learning is a general initial strategy for both learning about reward value of multi-dimensional stimuli and tackling the curse of dimensionality.

RL theories have been widely adopted as the main framework to understand reward learning in human and non-human primates. However, it has been suggested that other cognitive processes such as working memory (WM) play a role in learning (O'Reilly & Frank, 2006). For example, WM capacity has been shown to be a limiting factor for learning from reward feedback (Collins et al., 2017; Collins & Frank, 2012). Moreover, although it is generally accepted that WM capacity is discrete and limited (Awh, Barton, & Vogel, 2007; Cowan, 2001; Fukuda, Awh, & Vogel, 2010; Miller, 1956; Rouder et al., 2008), a series of recent studies has shown that the capacity of WM is continuous (Alvarez & Cavanagh, 2004; Bays, Catalao, & Husain, 2009; Bays & Husain, 2008; Ma, Husain, & Bays, 2014) and can be almost unlimited for naturalistic objects (Brady et al., 2016). Based on the aforementioned findings, our observed faster rate of learning for naturalistic stimuli could be attributed to an increase in WM capacity for these stimuli.

Studies of interactions between WM and learning have also pointed to the influence of individual differences in WM capacity on the balance between model-free and model-based learning (Etkin, Büchel, & Gross, 2016; Otto et al., 2013; Schad et al., 2014; Wills, Graham, Koh, McLaren, & Rolland, 2011). Although these suggest that WM capacity might affect the speed of alternation between learning strategies, it is still unclear how WM capacity influences learning strategies. Here, we find that naturalistic stimuli bias the initial learning strategy toward object-based learning and result in a faster transition to object-based learning.

Additionally, naturalistic stimuli are more familiar and could be perceived as more salient than abstract stimuli (Battistoni et al., 2018; Kaiser et al., 2016; Thorpe et al., 1996), and thus, could result in a strong bias toward object-based learning. Nonetheless, we find that the heuristic feature-based strategy, which provides an approximation for reward value based on features, is still adopted as the initial learning strategy when learning about naturalistic stimuli. Learning about features has been shown to enhance learning speed (Gigerenzer & Goldstein, 1996; Jocham et al., 2016) and allows for generalization of values (Kahnt, Park, Burke, & Tobler, 2012; Kahnt & Tobler, 2016). Together, these findings suggest that, when adopting learning strategies, the demand for adaptability (Farashahi et al., 2017b; Farashahi et al., 2017a; Farashahi et al., 2019; Soltani & Izquierdo, 2019) and tackling the curse of dimensionality could be the more important factors than the saliency of naturalistic stimuli.

Our experimental design has a few limitations that can be addressed in future experiments. First, only a specific type of object (i.e., shoe) was used as naturalistic stimuli, which might have been more familiar to some subjects than others. Familiarity (e.g., repeated exposure to the same stimuli) has been shown to enhance the WM performance (Olson, Jiang, & Moore, 2005; Olsson & Poom, 2005). Measuring subjects' degree of familiarity to establish a baseline measure with given stimuli can be used in future studies to understand the effect of familiarity on learning from reward feedback. Another limitation of our study is the difference between the two task-relevant features: color is a low-level visual feature whereas shoe type is more high-level and conceptual. Although we did not find any difference between learning depending on the informative feature, the difference between our two task-relevant features could potentially bias learning toward a feature-based strategy because of recent studies suggesting that existing semantic knowledge impacts WM capacity (Bower, Karlin, & Dueck, 1975; Brady et al., 2019; Konkle, Brady, Alvarez, & Oliva, 2010; McWeeny, Young, Hay, & Ellis, 1987).

Finally, a novel aspect of our work is the use of naturalistic stimuli to study learning because so far only a limited number of studies have investigated cognitive processes using such stimuli instead of abstract stimuli (Battistoni et al., 2018; Boorman, Rajendran, O'Reilly, & Behrens, 2016; Brady et al., 2016; Hickey & Peelen, 2015; Kaiser et al., 2016; Leong, Radulescu, Daniel, DeWoskin, & Niv, 2017). The lack of experiments exploring learning using naturalistic stimuli calls for reconsideration of existing findings based on abstract stimuli.

5. Conclusion

Here, we aimed to investigate learning about multi-dimensional naturalistic stimuli based on reward feedback. Crucially, our study is the first to compare response to multi-dimensional naturalistic stimuli and abstract stimuli in the context of learning. We demonstrate that learning about both types of stimuli involves transition from a feature-based to an object-based strategy, however, this transition is faster for naturalistic compared to abstract stimuli. Moreover, object-based learning is initially adopted more strongly for naturalistic than abstract stimuli, whereas the object-based strategy is adopted less for naturistic stimuli both overall and at the steady state. Overall, our results suggest that although naturalistic stimuli could be perceived as objects more strongly, leading participants to use the feature-based strategy less often initially and transition faster to object-based learning, the overall influence of individual features on learning was stronger for naturalistic stimuli.

Supplementary Materials

Supplementary data to this article, including parameters used for running the experiment and behavioral responses recorded for each subject can be download from http://ccnl.dartmouth.edu/DataShare/NatStiLer.zip.

Conflict of interest

The authors declare no competing interests.

Acknowledgements

We would like to thank Vivian Lee and Chanc VanWinkle Orzell for their helpful comments on the manuscript. This work is supported by the National Science Foundation (CAREER Award BCS1943767 to A.S.)

References

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*(2), 106–111.

Arntzen, E., & Lian, T. (2010). Trained and derived relations with pictures versus abstract stimuli as nodes. *The Psychological Record*, *60*(4), 659–678.

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, *18*(7), 622–628.

Barto, A. G., & Mahadevan, S. (2003). Recent advances in hierarchical reinforcement learning. *Discrete Event Dynamic Systems*, *13*(4), 341–379.

Battistoni, E., Kaiser, D., Hickey, C., & Peelen, M. V. (2018). The time course of spatial attention during naturalistic visual search. *Cortex*.

Bays, P. M., Catalao, R. F., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*(10), 7–7.

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*(5890), 851–854.

Boorman, E. D., Rajendran, V. G., O'Reilly, J. X., & Behrens, T. E. (2016). Two anatomically and computationally distinct learning signals predict changes to stimulus-outcome associations in hippocampus. *Neuron*, *89*(6), 1343–1354.

Bower, G. H., Karlin, M. B., & Dueck, A. (1975). Comprehension and memory for pictures. *Memory & Cognition*, *3*(2), 216–220.

Brady, T. F., Störmer, V. S., & Alvarez, G. A. (2016). Working memory is not fixed-capacity: More active storage capacity for real-world objects than for simple stimuli. *Proceedings of the National Academy of Sciences*, *113*(27), 7459–7464.

Brady, T. F., Störmer, V. S., Shafer-Skelton, A., Williams, J. R., Chapman, A. F., & Schill, H. M. (2019). *Scaling up visual attention and visual working memory to the real world*.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.

Braun, D. A., Mehring, C., & Wolpert, D. M. (2010). Structure learning in action. *Behavioural Brain Research*, *206*(2), 157–165.

Collins, A. G., Brown, J. K., Gold, J. M., Waltz, J. A., & Frank, M. J. (2014). Working memory contributions to reinforcement learning impairments in schizophrenia. *Journal of Neuroscience*, *34*(41), 13747–13756.

Collins, A. G., Ciullo, B., Frank, M. J., & Badre, D. (2017). Working memory load strengthens reward prediction errors. *Journal of Neuroscience*, *37*(16), 4332–4342.

Collins, A. G., & Frank, M. J. (2012). How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *European Journal of Neuroscience*, *35*(7), 1024–1035.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114.

Dayan, P., & Berridge, K. C. (2014). Model-based and model-free Pavlovian reward learning: revaluation, revision, and revelation. *Cognitive, Affective, & Behavioral Neuroscience*, *14*(2), 473–492.

Diuk, C., Tsai, K., Wallis, J., Botvinick, M., & Niv, Y. (2013). Hierarchical learning induces two simultaneous, but separable, prediction errors in human basal ganglia. *The Journal of Neuroscience*, *33*(13), 5797–5805.

Etkin, A., Büchel, C., & Gross, J. J. (2016). Emotion regulation involves both model-based and model-free processes. *Nature Reviews Neuroscience*, *17*(8), 532.

Farashahi, S., Rowe, K., Aslami, Z., Gobbini, M. I., & Soltani, A. (2018). Influence of learning strategy on response time during complex value-based learning and choice. *PloS One*, *13*(5), e0197263.

Farashahi, S., Donahue, C. H., Khorsand, P., Seo, H., Lee, D., & Soltani, A. (2017a). Metaplasticity as a neural substrate for adaptive learning and choice under uncertainty. *Neuron*, *94*(2), 401-414.

Farashahi, S., Rowe, K., Aslami, Z., Lee, D., & Soltani, A. (2017b). Feature-based learning improves adaptability without compromising precision. *Nature Communications*, *8*, 1768.

Farashahi, S., Donahue, C. H., Hayden, B. Y., Lee, D. & Soltani, A. (2019). Flexible combination of reward information across primates. Nature Human Behaviour, 3(11), 1215-1224

Fukuda, K., Awh, E., & Vogel, E. K. (2010). Discrete capacity limits in visual working memory. *Current Opinion in Neurobiology*, *20*(2), 177–182.

Gershman, S. J., & Niv, Y. (2010). Learning latent structure: carving nature at its joints. *Current Opinion in Neurobiology*, *20*(2), 251–256.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological Review*, *103*(4), 650–669.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). The elements of statistical learning: data mining, inference and prediction. *Springer-Verlag*, *1*(8), 371–406.

Hickey, C., & Peelen, M. V. (2015). Neural mechanisms of incentive salience in naturalistic human vision. *Neuron*, *85*(3), 512–518.

Ito, M., & Doya, K. (2009). Validation of decision-making models and analysis of decision variables in the rat basal ganglia. *Journal of Neuroscience*, *29*(31), 9861–9874.

Jocham, G., Brodersen, K. H., Constantinescu, A. O., Kahn, M. C., Ianni, A. M., Walton, M. E., … Behrens, T. E. (2016). Reward-guided learning with and without causal attribution. *Neuron*, *90*(1), 177–190.

Kahnt, T., Park, S. Q., Burke, C. J., & Tobler, P. N. (2012). How glitter relates to gold: similarity-dependent reward prediction errors in the human striatum. *Journal of Neuroscience*, *32*(46), 16521–16529.

Kahnt, T., & Tobler, P. N. (2016). Dopamine regulates stimulus generalization in the human hippocampus. *ELife*, *5*, e12678.

Kaiser, D., Oosterhof, N. N., & Peelen, M. V. (2016). The neural dynamics of attentional selection in natural scenes. *Journal of Neuroscience*, *36*(41), 10522–10528.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Scene memory is more detailed than you think: The role of categories in visual long-term memory. *Psychological Science*, *21*(11), 1551–1556.

Leong, Y. C., Radulescu, A., Daniel, R., DeWoskin, V., & Niv, Y. (2017). Dynamic interaction between reinforcement learning and attention in multidimensional environments. *Neuron*, *93*(2), 451–463.

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, *17*(3), 347.

McWeeny, K. H., Young, A. W., Hay, D. C., & Ellis, A. W. (1987). Putting names to faces. *British Journal of Psychology*, *78*(2), 143–149.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81.

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*(21), 8145–8157.

Oemisch, M., Westendorff, S., Azimi, M., Hassani, S. A., Ardid, S., Tiesinga, P., & Womelsdorf, T. (2019). Feature-specific prediction errors and surprise across macaque fronto-striatal circuits. *Nature Communications*, *10*(1), 176.

Olson, I. R., Jiang, Y., & Moore, K. S. (2005). Associative learning improves visual working memory performance. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(5), 889.

Olsson, H., & Poom, L. (2005). Visual memory needs categories. *Proceedings of the National Academy of Sciences*, *102*(24), 8776–8780.

O'Reilly, R. C., & Frank, M. J. (2006). Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation*, *18*(2), 283–328.

Otto, A. R., Raio, C. M., Chiang, A., Phelps, E. A., & Daw, N. D. (2013). Working-memory capacity protects model-based learning from stress. *Proceedings of the National Academy of Sciences*, *110*(52), 20941–20946.

Rouder, J. N., Morey, R. D., Cowan, N., Zwilling, C. E., Morey, C. C., & Pratte, M. S. (2008). An assessment of fixed-capacity models of visual working memory. *Proceedings of the National Academy of Sciences*, *105*(16), 5975–5979.

Schad, D. J., Jünger, E., Sebold, M., Garbusow, M., Bernhardt, N., Javadi, A.-H., … others. (2014). Processing speed enhances model-based over model-free reinforcement learning in the presence of high working memory functioning. *Frontiers in Psychology*, *5*, 1450.

Soltani, A., & Izquierdo, A. (2019). Adaptive learning under expected and unexpected uncertainty. *Nature Reviews Neuroscience*, *20*(10), 635-644.

Spachtholz, P., & Kuhbandner, C. (2017). Visual long-term memory is not unitary: flexible storage of visual information as features or objects as a function of affect. *Cognitive, Affective, & Behavioral Neuroscience*, *17*(6), 1141–1150.

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J. & Friston, K. J. (2009). Bayesian model selection for group studies. *NeuroImage, 46*, 311–311.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: an introduction*. Cambridge, MA: MIT Press.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6582), 520.

Wills, A. J., Graham, S., Koh, Z., McLaren, I. P., & Rolland, M. D. (2011). Effects of concurrent load on feature-and rule-based generalization in human contingency learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *37*(3), 308.

Wilson, R. C., & Niv, Y. (2012). Inferring relevance in a changing world. *Frontiers in Human Neuroscience*, *5*, 189.

Wunderlich, K., Beierholm, U. R., Bossaerts, P., & O'Doherty, J. P. (2011). The human prefrontal cortex mediates integration of potential causes behind observed outcomes. *Journal of Neurophysiology*, *106*(3), 1558–1569.