Exploration and Exploitation Reflect System-Switching in Learning

Li Xin Lim (lim226@purdue.edu)
Department of Psychological Sciences, 703 Third Street
West Lafayette, IN 47907 USA

Sebastien Hélie (shelie@purdue.edu)
Department of Psychological Sciences, 703 Third Street
West Lafayette, IN 47907 USA

Abstract

Mounting evidence suggests that human category learning is achieved by multiple qualitatively distinct biological and psychological systems. In an information-integration (II) categorization task, optimal performance requires switching away from rule and adopting a procedural response strategy. However, many participants perseverate with rules. This article attempts at understanding the difference between optimal and suboptimal participants in II categorization. To this end, we collected data in the Iowa Gambling Task (IGT) and an II categorization task. Performance in the IGT was used to estimate each participant's sensitivity to reward, punishment, and propensity to explore. The results show that optimal participants in the II task explored more in the IGT than suboptimal participants. However, optimal participants in the II task did not show higher sensitivity to punishment or lower sensitivity to reward. We conclude by discussing the implications of these findings on system-switching and theoretical work on multiple-systems model of perceptual category learning.

Keywords: perceptual categorization; decision-making; dual systems; exploration-exploitation

Introduction

Categorization is an important part of daily life. From categorizing objects as edible or not to categorizing people as friends or enemies, everyday life is filled with thousands of category decisions. Over the past 20 years, mounting evidence has been gathered that category learning is achieved using a number of different psychological and biological systems (e.g., Ashby et al., 1998; Ashby & Valentin, 2017; Erickson & Kruschke, 1998; Hélie et al., 2010; Nosofsky et al., 1994; Waldschmidt & Ashby, 2011). However, much less is known about the interactions between the multiple categorization systems (Hélie, 2017). For example, the COVIS theory of categorization (Ashby et al., 1998) assumes that participants begin by guessing or using simple rules generated by hypothesis testing. Only after these rules have failed will participants abandon rule-based strategies and proceed to using alternative, more intuitive and less verbal methods of categorization.

One task where the primacy of rule-based strategy is often observed is the information-integration (II) categorization task. In II categorization tasks, participants need to integrate information from more than one dimensions at a pre-decisional level in order to maximize accuracy. Example for the II category structures are shown in Figure 2B. In this figure, each symbol represents the coordinate of a stimulus in perceptual space and specify one specific rotation angle and frequency that allow for drawing a unique sine wave grating (see Figure 2A). In this example, participants need to learn to categorize the 'o' and '+' in separate categories. This can be achieved by drawing a line in Figure 2B, but notice that the line would not correspond to a meaningful verbal description. The verbal description would be: 'o' are stimuli where the rotation angle is larger than the frequency, which is not meaningful given that rotation angle and frequency are not commensurable.

In an II categorization task like the one presented in Figure 2, the most accurate verbal rules can produce an accuracy of about 75%. In order to perform optimally, participants need to abandon rules and rely on a non-verbal procedural strategy. Decision bound models (DBM) (Hélie et al.,2017; Maddox & Ashby, 1993) can be used to identify the type of strategy that participants are using, and a consistent finding over the past 30 years is that a substantial number of participants perseverate with rule-based strategy in II category tasks and as a result perform suboptimally.

Reward Processing

The goal of this study is to understand why certain participants fail to abandon rule-based strategies and adopt non-verbal procedural strategies. To generate predictions, we first used the COVIS model of categorization (Ashby et al., 1998; Hélie, Paul, & Ashby, 2012) to fit published II categorization data collected in our lab. The COVIS model implements a multiple-systems theory of category learning that includes an explicit hypothesis-testing system and an implicit procedural-learning system. The explicit system learns through declarative memory by choosing and testing simple verbally expressible rules, whereas the implicit system employs non-declarative memory whereby learning is mediated by reinforcement learning as the system gradually assigns motor responses to regions of perceptual space. On each trial, the model compares the confidence in both systems and produce one response, either from the explicit system or from the implicit system.

The COVIS model was fit to data from Experiment 2 in Hélie & Cousineau (2015) (Condition = 0.5) to understand the switching of learning system between explicit and implicit system in a perceptual categorization task. Decision bound models were fit to the data from each participant to separate participants using an optimal strategy from participants using a suboptimal strategy. The COVIS model was then fit to each group separately in order to identify which model parameters differed between simulations matching optimal participants and simulations matching suboptimal participants. Two hundred simulations were run for each subgroup of participants and the results are shown in Figure 1. The fit was excellent, with a RMSD of 1.5%. The model was able to differentiate optimal from suboptimal participants by changing the parameters odelta_e and odelta_c, which are the magnitude of the effect of the (negative and positive, respectively) feedback to adjust confidence in the hypothesis-testing system (Hélie, Paul, & Ashby, 2012). The simulations for optimal participants had a higher odelta_e value and a lower odelta_c value compared to the simulations of suboptimal participants, indicating that optimal participants are more sensitive to negative feedback while suboptimal participants are more sensitive to positive feedback. As a result, we hypothesize that optimal participants in II categorization tasks are more sensitive to negative feedback than participants who perseverate with rule-based strategies.

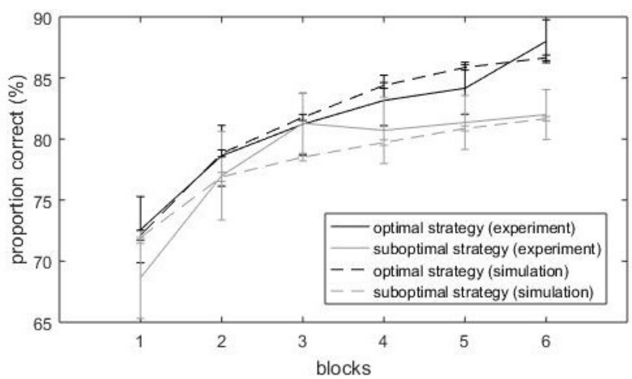

Figure 1: Average accuracy in Hélie & Cousineau (2015) and model results for each block of 100 trials. Black lines shows data for participants that use an optimal strategy, while grey lines indicate participants that use a suboptimal strategy. The participants' accuracy collected from the experiment are shown as solid lines, while the data from simulation are shown as dashed lines.

The Exploration-Exploitation Dilemma

One useful way to think about strategy switching and selection is to consider them in the context of the exploration-exploitation dilemma (Berger-Tal et al., 2014). Exploration and exploitation are seen as two opposing ways in the means of attention and resources allocation (Benner & Tushman, 2003; Gupta, Smith, & Shalley, 2006). Exploration entails risk taking, flexibility, discovery, and disengaging from the current task to allow for more room for experimentation, which is frequently associated with innovation. In contrast, exploitation is described with high-level engagement, choice-selection, efficiency and improvement (Laureiro-Martínez et al., 2015). The behavior of gathering information and exploiting are viewed as mutually exclusive events in many cases (Mettke-Hofmann, Winkler, & Leisler, 2002). When exploring, the agent seeks information about its environment as a way to improve performance, but in many situations it has to pay an opportunity cost (March, 1991). Agents that only exploit using current knowledge might be stuck in a suboptimal stable equilibrium, unable to adapt fully to the environment (March, 1991; Uotila et al., 2009). Thus, an optimal strategy in decision-making is to have balance between exploration and exploitation, allowing resource allocation between the two behaviors to yield the 'best' long-term rewards (March, 1991).

The exploration-exploitation dilemma to some extend resembles the results observed in the II categorization task. Assuming that participants begin by using a rule-based strategy, 'exploiters' may perseverate with a rule-based strategy since it allows for responding correctly in about 75% of the trials. Exploration is required to abandon rule-based strategies and try procedural strategies that are more optimal. As a result, we hypothesize that participants who explore more are more likely to perform optimally in an II categorization task.

Methods

To test for the hypotheses, we used the Iowa Gambling Task (IGT) (Bechara et al., 1994) to measure reward sensitivity and exploration tendencies. Each participant performed both an II categorization task and the IGT. Performance in the IGT was used to predict whether participants would use an optimal or suboptimal strategy in the II categorization task.

Participants

Fifty participants were recruited from the Purdue University undergraduate population. Each participant was given credit for participation as partial fulfillment of a course requirement. Participants gave written informed consent and all procedures were approved by the Purdue University Human Research Protection Program Institutional Review Board.

Materials and Procedure

Each participant did both the Iowa Gambling Task (IGT) and the perceptual categorization task (PCT) in random order of IGT-PCT ($n = 27$) and PCT-IGT ($n = 23$). The experiment was run on a Desktop PC equipped with a regular mouse and keyboard. Stimuli were displayed in a 21-inch monitor with $1{,}920 \times 1{,}080$ resolution. The experiment was controlled by in-house programs written using PsychoPy.

Iowa Gambling Task Participants were presented with four blue rectangles. The blue rectangles were labeled as "Deck A", "Deck B", "Deck C", and "Deck D". The task required participants to repeatedly draw 'cards' from the four decks, by clicking on the blue rectangle on the screen with a mouse. Participants were required to select a deck on each trial within four seconds. If a participant failed to select a deck before the deadline, the program randomly selected a deck. The participant could only select one deck for each trial.

The expected values of the decks differed so that two decks were associated with high immediate rewards but long-term overall loss (disadvantageous decks A and B), and two other decks yielded lower immediate rewards but long-term overall gains (advantageous decks C and D). The experiment was designed to record the participant's affinity towards each deck given the rewards and penalties presented in each trial upon selection of a particular deck. The reward and penalty from the selected deck in the particular trial, as well as the total accumulated gain from the rewards and penalty gathered thus far was presented to the participant at the end of each trial. The rewards and penalties were generated to meet the requirements listed in Table 1. Each deck contained 10 different cards and was re-shuffled after all 10 cards had been drawn. Each participant performed 120 trials grouped into six blocks of 20 trials each. Completing the IGT took about 10 minutes.

Table 1: Deck properties (Bechara et al., 1994)

Card	Deck A	Deck B	Deck C	Deck D
P(penalty)	0.5	0.1	0.5	0.1
Penalty	-150 to -350	-1250	-25 to -75	-250
Reward	100	100	50	50
Expectation	-250	-250	250	250

Perceptual Categorization Task (PCT) The stimuli used in the PCT were circular sine-wave gratings of fixed contrast and size, as shown in Figure 2. The stimuli differed in terms of bar width and orientation. The bar width was derived as the frequency of lines in a 2D space in cycles per degree, while the orientation is the counterclockwise rotation of the lines from horizontal in radian. The stimuli were categorized as A and B, with a diagonal line as a category bound as shown in Figure 2. Perfect accuracy was possible and optimal performance required responding to the A-B stimuli using a procedural strategy.

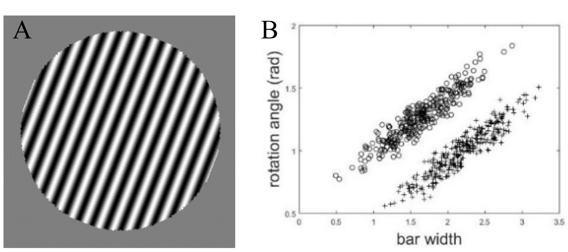

Figure 2: (A) Example stimulus shown to the participants for PCT, (B) Category structures used in PCT.

The participants were informed that they were taking part in a categorization experiment and that they needed to learn to categorize the stimuli presented into either category A or B with trial-and-error. In each trial of this task, a "crosshair" was presented on the screen for one second, followed by a single stimulus presented in the center of the screen. Participants were required to choose a category for the stimulus. Responses were given on a standard keyboard: "s" key for category A and "k" key for category B. After each trial, visual feedback showing "Correct", "Incorrect", or "Wrong Key" was given to the participant according to the response they pushed. The response for stimulus on each trial was recorded, as well as the response time. The participants did 600 trials grouped into six blocks of 100 trials each. The PCT took about 35 minutes to complete.

Decision bound models

The objective of the study was to explore the difference in sensitivity to reward and punishment between participants that used an optimal strategy and participants that did not use an optimal strategy. To allow for the classification of participants into optimal strategy users and suboptimal strategy users, Decision Bound Models (DBM) were applied to the perceptual categorization task to identify how participants learned to assign responses to regions of perceptual space. In DBM, it is assumed that participants determine the region of the percept and give the associated response. The decision bound is described as a partition segregating competing response regions. Three general classes of decision bound models were fit to response data of the PCT (Hélie et al., 2017; Maddox & Ashby, 1993): (1) guessing models, (2) explicit rule-reasoning models, and (3) procedural learning models.

The guessing models assumes that participants do not use the information on the screen and randomly responded "A" or "B" in each trial. The explicit models set a boundary to segregate participant's responses with a vertical line or horizontal line (or the combination of both vertical and horizontal lines). An adjusted diagonal line is used as the boundary instead in the procedural learning models. For each participant's data set, the best model is selected using the Bayes information criterion (BIC). Participants whose data were best-fit by the optimal models, which is the procedural learning model in this case, are labelled as "optimal strategy" and all other participants are labelled as "suboptimal strategy".

Rescorla-Wagner Model

The data recorded in the IGT were fitted with the Rescorla-Wagner (1972) model (RW). The RW was used to calculate a value for each deck and estimate a participant's sensitivity towards reward and punishment.

Data for each participant was fed into the RW model. For each trial, t in a particular task block, the parameter for sensitivity to reward, b_{rew} was multiplied with the magnitude of reward, R received following the participant's response in each trial, while the parameter for sensitivity to punishment, b_{pun} was multiplied with the magnitude of punishment, P received following the participant's response in each trial.

The key equations to update reward and punishment sensitivity were:

$$B_{rew} = \frac{b_{rew} \times (R(t) - P(t))}{\max(R)} \quad (1)$$
$$B_{pun} = \frac{b_{pun} \times (P(t) - R(t))}{\max(P)}$$

where, B_{rew} and B_{pun} are the sensitivity to reward and punishment, for the perceived net gain and loss in each trial. The key equations to update reward and punishment sensitivity were:

$$\mathrm{Qdeck(t)} = \mathrm{Qdeck(t-1)} + \alpha\left(B_{rew} - \mathrm{Qdeck(t-1)}\right) \quad (2)$$
$$\mathrm{Qdeck(t)} = \mathrm{Qdeck(t-1)} + \alpha(B_{pun} - \mathrm{Qdeck(t-1)})$$

where, Qdeck is the Q-value for each deck and α is the learning rate. The equation on top in Eq. 2 updates the deck value with B_{rew}, while the equation below updates with B_{pun}. In trials where an overall reward was received, the equation with B_{rew} was used to update the deck value; in trials where overall punishment was received, the equation with B_{pun} was applied to update the deck value. All parameters were estimated using Maximum A Posteriori (MAP).

The sensitivity towards reward and punishment b_{rew} and b_{pun} for each participant were then normalized. The weighted proportion of b_{rew} and b_{pun} with respect to the summation of b_{rew} and b_{pun} were computed with Equation 3.

$$W_{rew} = \frac{b_{rew}}{b_{rew} + b_{pun}} \quad (3)$$
$$W_{pun} = \frac{b_{pun}}{b_{rew} + b_{pun}}$$

where, W_{rew} and W_{pun} are the weighted proportion of b_{rew} and b_{pun}, respectively.

Results

Effects of sensitivity to punishment and rewards

Participants in the PCT were categorized into participants who found the optimal strategy and participants who did not using DBM. The sensitivity to punishment (b_{pun}) and reward (b_{rew}) were computed with the RW. W_{rew} and W_{pun} were computed as the weighted proportion of sensitivity to reward and punishment, respectively, and the mean estimates of the proportion of W_{rew} and W_{pun} for participants that used an optimal strategy and a suboptimal strategy are shown in Figure 3. Confirming our hypothesis, W_{rew} [t(48) = 1.901, p = 0.032] of participants that used an optimal strategy was lower than that of participants that used a suboptimal strategy, whereas W_{pun} [t(48) = -1.901, p = 0.032] of participants that used an optimal strategy was higher than that of participants that used a suboptimal strategy. These results show that participants using an optimal strategy in the PCT have a greater sensitivity to punishment, while participants using a suboptimal strategy in the PCT have a higher sensitivity to reward.

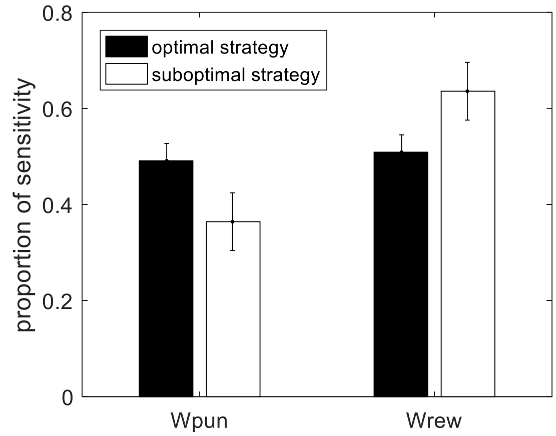

Figure 3: Estimated sensitivity to reward and punishment (IGT) for optimal and suboptimal participants (PCT). Error bars are standard error of the mean.

Exploration affects category learning

Exploration was quantified as the number of deck switches in the IGT and was compared between the two groups of participants. The number of deck switches was subjected to an independent samples t-test to test the effect of using optimal or suboptimal strategy in PCT. The average number of deck switches for the two groups is shown in Figure 4. The main effect of the strategy used [t(48) = 1.684, p = 0.049] was significant. The number of deck switches for participants that used an optimal strategy in the PCT (mean = 71.91) was higher than the number of deck switches for participants that used a suboptimal strategy in the PCT (mean = 60.38). The results suggest that, as predicted, participants who conduct more exploration are more likely to perform optimally in an II categorization task.

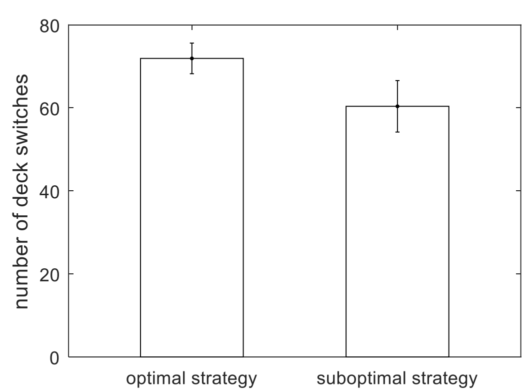

Figure 4: Average number deck switches in the IGT for optimal and suboptimal participants in the PCT. Error bars are standard error of the mean.

To further assess the role of exploration in strategy selection, we measured the entropy of choosing different decks: A, B, C and D. Entropy gives a sense of disorder and uncertainty. Hence higher entropy means that all decks were sampled equally often, whereas an entropy of 0 means that participants always selected the same deck. We first calculated the correlation between entropy and number of deck switches in the IGT (Figure 5A). This analysis informs whether participants always switch between a subset of the decks or if all decks are sampled. The correlation was 0.513, which is statistically significant [$t(48) = 4.136$, $p < 0.001$]. This result suggests that participants with more deck switches sample from all decks.

Next, we computed the linear relationship between entropy and sensitivity to feedback in the IGT ($b_{pun} + b_{rew}$). Here, b_{pun} is negative, so a negative number means higher sensitivity to punishment while a positive number means a higher sensitivity to reward (0 means equally sensitive to both types of feedback). This analysis informs about the relationship between the breadth of exploration (sampling from some or all the decks) and feedback sensitivity in the IGT. The correlation was -0.665, which reached statistical significance [$t(48) = -6.168$, $p < 0.001$] (Figure 5B). This suggests that higher sensitivity to punishment leads to sampling from more decks, which is consistent with our hypothesis that greater sensitivity to punishment leads to more exploration.

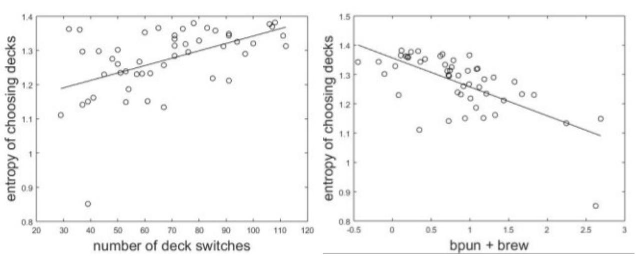

Figure 5: (A) Relationship between entropy and number of deck switches in the IGT. (B) Relationship between feedback sensitivity and entropy in the IGT.

Discussion

This article presents the results of an experiment aimed at understanding why some participants fail to select an optimal procedural strategy in II categorization and instead perseverate with using suboptimal rule-based strategies. By fitting the COVIS model to published II categorization data, we hypothesized that participants using an optimal strategy in II categorization would be more sensitive to punishment whereas participants using a suboptimal strategy would be more sensitive to reward. We further hypothesized that participants with a tendency to explore would be more likely to use an optimal strategy in the II task.

We tested these predictions by running participants in an II categorization task and the IGT. Decision-bound models were fit to II categorization data to classify each participant as optimal or suboptimal. The RW model was fit to the IGT to estimate each participant's sensitivity to reward and punishment. The number of deck switches and entropy of choice in the IGT were used to estimate the propensity of each participant to explore. The results partially supported our hypotheses. As predicted, exploration was related to the selection of an optimal strategy in II categorization. Sensitivity to punishment was also related to propensity to explore, but only in the IGT. The hypothesis that sensitivity to punishment would be related to the selection of an optimal strategy in II categorization was not supported in the experiment.

System-Switching vs. Rule-switching

Individuals vary considerably in terms of their sensitivity to reward and punishment. Sensitivity to reward can be described as how an individual's behavior is driven by reward-related stimuli, while sensitivity to punishment is described as how an individual's behavior is subdued by punishment-related stimuli. Studies suggest that individuals with greater sensitivity to reward are more reactive to rewarding outcomes but are less sensitive to monitoring loss, while greater sensitivity to punishment are linked to avoidance and giving up actions in absence of immediate reward (Kim et al., 2015).

As predicted by COVIS, the selection of certain strategies and the abandonment of others depends on the evaluation of how rewarding the strategy is. The implementation of one system over the other depends on the confidence and trust in the system. The trust is a function of the effect of received feedback when using a particular system. Thus, the switching of strategies from a rule-based to aa procedural strategy depends on the feedback received when using the particular strategy, which can be explained in terms of the reward and punishment the system or strategy gets when providing a response.

Our study confirms the finding that the selection of an optimal strategy in II categorization task is associated with greater sensitivity to punishment and perseveration with suboptimal strategies in II tasks is associated with greater sensitivity to reward. If the participant focuses more on the effect of punishment, losses in the task leads to giving up and avoidance of certain strategies, which leads to the possibility of adopting a strategy that leads to the optimal outcome. If the participant has greater sensitivity towards reward, s/he is less sensitive to immediate loss and tend to perseverate with strategies that bring a certain degree of rewards. In the II categorization task, participants tend to perseverate with a rule-base strategy since the strategy allows for a certain degree of accuracy (typically about 70%). However, additional research is required to determine how participants change from attending to specific stimulus dimension(s) to an integrated procedural-based strategy.

Exploration in the IGT

Deck switch is used as a measure of exploration. A larger number of deck switches indicates that participants were willing to disengage from the current strategy or deck, and were willing to explore other possible options, despite the uncertainty and risk of getting punished. As seen in Figure 4, both optimal and suboptimal participants explored greatly in the IGT. The difference in exploration between optimal and suboptimal participants may have been caused by several factors, which includes willingness to take risk to maximize gain. The tendency for exploration appears to be robust and may be predictive of both rule-switching and system-switching. This was shown by the number of switches being both related to entropy in the IGT and the selection of an optimal strategy in II categorization. COVIS does not explicitly model exploration but a tendency to explore would be characterize by noise in system selection.

One observation is that many participants tend to choose Deck B in the IGT. With an expected value of a net loss of $250 and a relatively large loss of $1250 as compared to the other decks in the task, it would normally inhibit participants from selecting Deck B. The basic assumption is that the largest loss would trigger an alarming signal from the intact somatic system, thus inhibiting further selection of deck B as it guides the process of decision-making (Lin et al., 2007). However, Deck B has a low loss-frequency owing to a small number of trials with large losses (or can be seen as a high gain-frequency), which may explain why participants choose the deck despite great immediate loss when a penalty card is drawn from the deck. Most participants' behavior are driven by the high gain-frequency, instead of inhibited by the great loss while choosing Deck B (Dunn et al., 2006; Lin et al., 2007).

Participants that used suboptimal strategies tend to fixate on specific deck(s) and were not willing to explore for more reward, which might cause them to be stuck in a local minimum, and lose the chance to seek out strategies that are more efficient. The fixation can be due to contentment, unwillingness to take risks, or pros-to-cons weighing. Additional research is needed to determine why certain participants are reluctant to explore.

Future Work

This experiment came with a few limitations. Some of the advantage and disadvantage decks used in the IGT were difficult to identify through limited interactions with the decks, which might misguide participants while performing the tasks. For example, exploiting Deck B in IGT results in an overall loss, the frequency of loss is small. Hence, participants may consider Deck B to be an advantageous deck and continue choosing the deck along with other advantage decks. Questionnaires could be given to participants to ask for the decks the participants believed to be advantageous. This would allow for better understanding whether participants considered each deck as "risky" or not and disentangle risk taking from bad estimation of deck expectation.

Finally, a task needs to be designed that shares properties with the IGT but requires system-switching instead of rule-switching (or deck switching). This new task would allow to more directly estimated sensitivity to reward and punishment between-system and would provide a more definitive test of the hypothesis that optimal participants, who switch system in an II categorization task, are more sensitive to punishment then suboptimal participants, who are more sensitive to reward.

Acknowledgement

This research was supported, in part, by National Science Foundation award #1662230 and by NIMH grant #2R01MH063760.

References

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*(3), 442–481.

Ashby, F. G., & Valentin, V. V. (2017). Multiple systems of perceptual category learning: Theory and Cognitive Tests. *Handbook of Categorization in Cognitive Science*, 157–188.

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, *50*, 7–15.

Benner, M. J., & Tushman, M. L. (2003). Exploitation, exploration, and process management: the productivity dilemma revisited. *Academy of Management Review*, *28*, 238–256.

Berger-Tal, O., Nathan, J., Meron, E., & Saltz, D. (2014). The exploration-exploitation dilemma: A multidisciplinary framework. *PLOS ONE*, *10*, e0119116.

Crossley, M., Roeder, J., Hélie, S., & Ashby, F. (2018). Trial–by–trial switching between procedural and declarative categorization systems. *Psychological Research*, *82*, 371–384.

Dunn, B. D., Dalgleish, T., & Lawrence, A. D. (2006). The somatic marker hypothesis: A critical evaluation. *Neuroscience Biobehaviour*, *30*(2), 239–271.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and Exemplars in Category Learning. *Journal of Experimental Psychology: General*, *127*(2), 107–140.

Gupta, A. K., Smith, K. G., & Shalley, C. E. (2006). The interplay between exploration and exploitation. *Academy of Management Journal*, *49*, 693–70.

Hélie, S. (2017). Practice and preparation time facilitate system-switching in perceptual categorization. *Frontiers in Psychology*, *8*, 1964.

Hélie, S., & Cousineau, D. (2015). Differential effect of visual masking in perceptual categorization. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(3), 816–825. https://doi.org/10.1037/xhp0000063

Hélie, S., Paul, E. J., & Ashby, F. G. (2012). Simulating the Effects of Dopamine Imbalance on Cognition: From Positive Affect to Parkinson's Disease. *Neural Networks*, *32*, 74–85.

Hélie, S., Turner, B. O., Crossley, M. J., & Ell, S. W. (2017). Trial-by-trial identification of categorization strategy using

iterative decision bound modeling. *Behaviour Research Method*, *49*, 1146–1162.
Hélie, S., Waldschmidt, J. G., & Ashby, F. G. (2010). Automaticity in rule-based and information-integration categorization. *Attention, Perception, & Psychophysics*, *72*, 1013–1031.
Kim, S. H., Yoon, H., Kim, H., & Hamann, S. (2015). Individual differences in sensitivity to reward and punishment and neural activity during reward and avoidance learning. *Social Cognitive and Affective Neuroscience*, *10*(9), 1219–1227.
Laureiro-Martínez, D., Brusoni, S., Canessa, N., & Zollo, M. (2015). Understanding the exploration-exploitation dilemma: an fMRI study of attention control and decision-making performance. *Strategic Management Journal*, *36*, 319–338.
Lin, C. H., Chiu, Y. C., Lee, P. L., & Hsieh, J. C. (2007). Is deck B a disadvantageous deck in the Iowa Gambling Task? *Behavioral and Brain Functions*, *3*(16).
Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception and Psychophysics*, *53*(1), 49–70.
March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, *2*, 71–87.
Mettke-Hofmann, C., Winkler, H., & Leisler, B. (2002). The significance of ecological factors for exploration and neophobia in parrots. *Ethology*, *108*, 249–272.
Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-Plus-Exception Model of Classification Learning. *Pyschological Review*, *101*(1), 53–79.
Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In H. A. Prokasy & W. F. Black (Eds.), *Classical conditioning II: Current research and theory*. New York: Appleton- Century-Crofts.
Uotila, J., Maula, ., Keil, T., & Zahra, S. A. (2009). Exploration, exploitation, and financial performance: analysis of S&P 500 corporations. *Strategic Management Journal*, *30*, 221–231.
Waldschmidt, J. G., & Ashby, F. G. (2011). Cortical and striatal contributions to automaticity in information-integration categorization. *NeuroImage*, *56*(3), 1791–1802.