Published in final edited form as:

Psychon Bull Rev. 2015 April ; 22(2): 586–593. doi:10.3758/s13423-014-0694-2.

Implicit learning mediates base rate acquisition in perceptual categorization

Corey J. Bohil and Andrew J. Wismer

Department of Psychology, University of Central Florida, Orlando, FL 32816, USA

Corey J. Bohil: corey.bohil@ucf.edu

Abstract

We explored the possibility, suggested by Koehler (*Behavioral and Brain Sciences*, 19, 1–53, 1996; also Spellman *Behavioral and Brain Sciences*, 19, 38, 1996), that implicit learning mediates the influence of base-rates on category knowledge acquired through direct experience. In two experiments, participants learned simple perceptual categories with unequal base-rates (i.e., presentation frequency). In Experiment 1, participants received either response training or observational training. In Experiment 2, participants received response training with either immediate or delayed feedback. In previous studies, observational training and delayed feedback training have been shown to disrupt implicit learning. We found that base-rate influence was weaker in these conditions when category discriminability was low (i.e., when category membership was difficult to determine). This conclusion was based on signal detection β values as well as decision-bound modeling results. Because these disruptions to implicit learning attenuate the base-rate effect, we conclude that implicit learning does indeed underlie the influence of base-rates learned through direct experience. This suggests that the implicit learning system postulated by the COVIS theory of categorization (Ashby, Alfonso-Reese, Turken, & Waldron *Psychological Review*, 105, 442–481, 1998) may be involved in developing sensitivity to category base-rates.

Keywords

Base rate; Categorization; Implicit learning; COVIS

Introduction

Classification judgments are commonly based on observable stimulus features. Medical diagnosis, for example, relies on symptoms exhibited by a patient. But observable features are sometimes insufficient for accurate judgment. For instance, a set of symptoms (e.g., fever, aches, fatigue) may stem from a variety of causes. In such cases, choice may be guided by knowledge that is separate from outwardly apparent stimulus information. Base-rates—the relative prevalence of category alternatives—are a well-known example. Fever and fatigue may suggest allergies in July but imply the flu in January.

Several variables determine the influence of base-rate information, such as presentation format (e.g., probability vs. frequency summary statistics) and the diagnosticity of stimulus features (to name just a few; see Koehler, 1996 for a review). A useful task for studying these variables is perceptual classification learning. Many variables can be manipulated within this context, and decision-criterion location provides a basis for comparing results across studies. A widely-held conclusion has been that base-rates exert a strong influence over performance when directly experienced over a long series of classification trials (Bohil & Maddox, 2001; Estes, 1989; Gluck & Bower, 1988). Maddox and Bohil demonstrated that different levels of category discriminability and base-rate ratio moderate the influence that base-rates have on criterion placement when learned via direct experience (for a review see Maddox, 2002).

Although the effect of direct experience on base-rate learning has received much attention from researchers, one proposed explanation for this result has not. Koehler speculated that, while gaining direct experience with a set of categories, base-rate sensitivity results from implicit learning of relative prevalence over time (Koehler, 1996; see also Holyoak & Spellman, 1993; Spellman, 1996). To our knowledge, however, this possibility has not been directly tested. We hypothesize that manipulations that disrupt implicit learning should diminish the effect of base-rates during category learning, and this should be reflected in decision criterion placement.

Implicit learning in classification

Outside the base-rate literature, the role of implicit learning in classification has become a prominent research topic. Stemming largely from the work of Ashby and colleagues (e.g., Ashby & Maddox, 2011), implicit learning is thought to constitute a major system for learning some types of categorization rules. Specifically, when information from multiple stimulus dimensions must be integrated, excellent performance can be achieved despite participants' inability to verbalize their basis for responding. This result is embodied in the COmpetition between Verbal and Implicit Systems (COVIS) theory of categorization (Ashby, Alfonso-Reese, Turken, & Waldron, 1998).

COVIS is a neuropsychological theory that posits a role for separate brain regions in learning classification rules. An explicit learning system mediates verbalizable rule learning (via prefrontal cortex and other regions), which relies on conscious hypothesis testing by the learner. An implicit system learns classification rules incrementally via dopamine release in the striatum. Procedural learning (e.g., button pressing) seems to facilitate implicit category rule learning (Ashby, Ell, & Waldron, 2003).

Several studies have tested COVIS predictions contrasting explicit and implicit rule learning (Ashby & Maddox, 2011). Participants often are tasked with learning a set of two-dimensional category structures—those that can be separated by a unidimensional rule along a single critical dimension (rule-based [RB] categories) and those that require a rule that integrates information from both dimensions (information-integration or II categories). RB categories appear to be learned via the explicit system, whereas II category learning seems dependent on implicit learning.

In one study, participants learned categories either via response-training (a classification response followed by corrective feedback on each trial) or by observational-training (Ashby, Maddox, & Bohil, 2002). In observational-training, participants watched each stimulus presented along with its category label without making a classification response. Training condition made little difference for RB category structures. For II category structures, though, response-training led to good performance while observational-training severely disrupted performance. These results support the contention that acquiring II-category rules depends on implicit learning, which recruits the procedural process of making classification responses (via button presses on a keyboard) with feedback. Observational training removes the procedural element, disrupting implicit learning. In another study, participants made classification responses followed by corrective feedback on each trial but with either immediate or delayed (2.5, 5, or 10 seconds) feedback (Maddox, Ashby, & Bohil, 2003). Delayed feedback had no effect on RB category learning, but II category performance was much worse under delayed feedback conditions. This suggests that delayed feedback disrupted implicit learning, which depends on the timing of dopamine release in the basal ganglia.

Current research

The current research tests the idea that base-rate knowledge acquired through direct experience with categories relies on implicit learning. *Critically, as the studies described above show, direct experience with category exemplars doesn't always mean implicit learning* (i.e., RB categories learned through direct experience rely on explicit reasoning). We sought to test the hypothesis that implicit learning underlies base-rate acquisition through experience with categories.

We conducted two experiments using manipulations—observational and delayed feedback training—previously shown to disrupt implicit but not explicit rule learning. In both studies, participants learned univariate categories with unequal base-rates. To distinguish between explicit and implicit learning, we replicated the manipulations from the COVIS studies described above. In Experiment 1, participants received either response-training or observational-training. If implicit learning promotes base-rate acquisition, then base-rates should have less influence on decision-criterion placement after observational-training, because this disrupts implicit learning. In Experiment 2, all participants learned the categories through response-training, either with immediate or delayed feedback. Again, if implicit-learning underlies base-rate acquisition via direct experience, performance should be disrupted in the delayed feedback condition. As in our earlier work (described above), we tested base-rate effects at two levels of discriminability (d'), and (in Experiment 1) two different base-rate ratios. Also as in our previous work, some participants completed category pretraining before experiencing a base-rate manipulation, which might help to clarify the impact of base-rate knowledge.

Experiment 1

Experiment 1 compared response- and observational-training across ten between-participant conditions. Eight conditions resulted from factorial combination of two levels of

discriminability (d' = 1 or 2.2), two base-rate ratios (2:1 or 3:1), and two levels of training type (response or observational). In these eight conditions, participants first completed a baseline phase in which they learned the category structures with unbiased base-rates (details below). During this phase, category d' matched the ensuing experimental phase, and participants gave a categorization response followed by feedback on each trial. This procedure has been used in previous studies to isolate base-rate influence from category learning (Bohil & Maddox, 2003a, b). Two groups also completed conditions without first completing baseline category training. In these two conditions, participants completed either response or observational training with low discriminability categories ($d' = 1$) and the more pronounced (i.e., 3:1) base rate difference.

Methods

Participants—University of Central Florida undergraduates received course credit for participation. In the $d' 1$ conditions preceded by baseline training, 35, 44, 34, and 45 participants completed the 2:1-response, 3:1-response, 2:1-observational, and 3:1-observational conditions, respectively.¹ In the $d' 2.2$ conditions with baseline training, 31, 34, 30, and 29 participants completed the 2:1-response, 3:1-response, 2:1-observational, and 3:1-observational conditions, respectively. In the conditions without baseline training, 32 received response training, and 30 received observational training.

Stimuli—Stimuli were filled white rectangular bars on a black background displayed (one per trial) on a high resolution monitor. Each stimulus was 40 pixels wide and varied in height over trials (i.e., height determined category membership). The base of each bar graph was centered onscreen. The stimuli subtended between approximately 2–6 degrees of visual angle.

Two categories—A and B—were sampled from overlapping univariate normal distributions ($d' 1$ condition Category A mean = 99.6, Category B mean = 120.5; $d' 2$ condition Category A mean = 87.4, Category B mean = 132.7; $SD = 21.0$ for all categories). Four sets of 60 bar-height values were sampled from these distributions, crossing the two levels of discriminability ($d' = 1$ or 2.2) and base-rate (2:1 or 3:1). In the 2:1 base-rate conditions, 40 stimuli were sampled from Category A and 20 from Category B. In the 3:1 base-rate conditions, 45 and 15 stimuli were sampled from Categories A and B, respectively. (The direction of base-rate bias was counterbalanced and found to have no impact on performance.)

Each experimental condition consisted of ten 60-trial blocks in which the appropriate stimuli were presented in random order. For baseline-training trials, 30 stimuli from each category were sampled and presented in random order.

Procedure—Participants performed a simulated medical diagnosis task. On each trial, they saw a bar graph indicating a patient's "test result" and were instructed to respond with the disease (A or B) that they thought was more likely given the bar height.

¹Analysis using the first 35 participants in the $d' 1$ 3:1 conditions (to roughly equate group sizes) led to the same conclusions as our analysis using all data files (reported below).

Each experimental condition consisted of ten blocks, alternating between 60-trial training and transfer phases, for a total of 600 trials. During training blocks in the response-training conditions, a stimulus was presented for 500 ms and then removed. The participant had 10 seconds to make a response by pressing either the 'z' key (labeled A) or the 'm' key (labeled B) on the keyboard. After responding, the correct response was shown for 500 ms. During training blocks in the observational-training conditions, the stimulus was presented first for 500ms, followed by the label for 500 ms (the stimulus was not shown during label presentation in any condition). Participants viewed the category labels and stimuli on each trial but never made a response during observational training blocks. The intertrial interval was 1,000 ms in all conditions.

During transfer phases, whether in a response- or observational-training condition, participants viewed a stimulus on each trial (for 1,000 ms), made a classification response by pressing either the key labeled A or B, and then viewed the message "response recorded" for 500 ms.

During the baseline training phase, participants completed up to 600 trials of category training with an unbiased (1:1) base-rate ratio. After completing the first 60 trials, accuracy was checked every 10 trials. If response accuracy was no less than 2% below optimal (optimal was 69% correct for $d' = 1$ and 86% for $d' = 2.2$ conditions) for the most recent 60 trials during baseline, the baseline phase ended and the participant was allowed to begin experimental trials. Participants failing to reach this training criterion in 600 trials were excluded from further analysis.

Results

In the conditions preceded by baseline training, 73% of participants in the $d' 1$ conditions reached the accuracy criterion by the end of training. In the $d' 2.2$ conditions, 85% reached this criterion. After removing participants who failed to reach this criterion, the remaining sample sizes were as follows. In the $d' 1$ conditions, 20, 36, 26, and 34 participants completed the 2:1-response, 3:1-response, 2:1-observational, and 3:1-observational conditions, respectively. In the $d' 2.2$ conditions, 24, 30, 27, and 25 participants completed the 2:1-response, 3:1-response, 2:1-observational, and 3:1-observational conditions, respectively. Only these participants—who had clearly learned the category structures before being exposed to a base-rate manipulation—were included in our analyses. In addition, all participants from the two conditions that did not include baseline training were included in our analyses. Our goal was to compare observational- vs. response-training under a variety of conditions, not to compare performance across conditions with or without category pretraining.

Table 1 displays average signal detection criterion, β , values (along with classification accuracy) for Experiment 1 conditions. Using β as a dependent variable, we performed a mixed-factor ANOVA comparing all 10 experimental conditions (between: 2 d' levels $\times$ 2 base-rate ratios $\times$ 2 training types and including presence/absence of baseline training; within: 5 test blocks). There was a main effect of discriminability, with larger β values in $d' 2.2$ ($M = 2.21$, $SE = 0.08$) than in $d' 1$ ($M = 1.28$, $SE = 0.05$) conditions, $F(1, 241) = 93.65$, $p < 0.001$, $\eta_p^2 = .28$. β values were reliably larger in 3:1 base-rate conditions ($M =$

1.74, $SE = 0.05$) than in 2:1 conditions ($M = 1.52$, $SE = 0.07$), $F(1, 241) = 17.4$, $p < 0.001$, $\eta_p^2 = 0.07$. The base-rate difference was more pronounced in $d' 2.2$ conditions than in $d' 1$ conditions, $F(1, 241) = 7.88$, $p < 0.005$, $\eta_p^2 = 0.03$. Values of β increased toward optimal (optimal β was 3 in 3:1 base-rate conditions, 2 in 2:1 conditions) across blocks, $F(4, 964) = 15.47$, $p < 0.001$, $\eta_p^2 = 0.06$. Category discriminability also interacted with block. Criterion values increased more in the $d' 2.2$ conditions than in the $d' 1$ conditions, $F(4, 964) = 5.2$, $p < 0.001$, $\eta_p^2 = 0.02$. There was not a significant effect of training type, however, $F(1, 241) = 0.17$, $p = 0.69$, $\eta_p^2 = 0.001$. There also was no reliable effect of completing baseline category pretraining, $F(1, 241) = 0.13$, $p = 0.72$, $\eta_p^2 = 0.001$.

Based on previous work (e.g., Bohil & Maddox 2003a, b), we suspected that the influence of base-rates might most clearly be seen when category discriminability is low ($d' 1$ conditions) and when the base-rate differential was most pronounced (3:1 conditions). As a result, we conducted another ANOVA looking at only the $d' 1$, 3:1 base-rate conditions (including those with and without baseline training). In this case, we found a significant effect of training type (response or observational), $F(1, 124) = 11.8$, $p = 0.001$, $\eta_p^2 = 0.09$. There also was a significant effect of block [$F(4, 496) = 12.59$, $p < 0.001$, $\eta_p^2 = 0.09$] and an interaction between block and training [$F(4, 496) = 2.58$, $p = 0.037$, $\eta_p^2 = 0.02$]. Participants in the response-training conditions clearly adjusted their decision criteria further in the optimal direction than participants in the observational training conditions (Fig. 1).

Discussion

In conditions where category discriminability was low ($d' = 1$) and when the base-rate differential was large (3:1), participants clearly performed better with response training. Decision criterion (β) values were closer to optimal in these conditions than in the observational training conditions. When the procedural component of the task was removed—in the observational conditions—base-rates had a smaller influence over criterion placement. This suggests that implicit learning is indeed involved in acquiring base-rate knowledge from experience. Although the one-dimensional rule might be easily verbalized in this experiment, the participant learns where to position their decision cutoff incrementally over time (with corrective feedback providing a reward signal), which is very difficult to verbalize. This learning seems to have been disrupted by removal of the procedural response component, as in the earlier study by Ashby et al. (2002).

Experiment 2

Experiment 2 replicates the critical manipulation from (Maddox, Ashby, & Bohil, 2003) and provides an alternate method for assessing the importance of implicit learning in base-rate acquisition. In Experiment 2, participants made a response and received feedback on every trial, but feedback was either immediate or delayed.

Experiment 2 compared six base-rate learning conditions with either immediate or delayed feedback. Four conditions resulted from combining two levels of category discriminability ($d' = 1$ or 2.2) with two feedback types (immediate or delayed). As in Experiment 1, these four conditions were preceded by a baseline category learning phase. We also included conditions that excluded the baseline category learning in order to evaluate the effect of

base-rates and categories learned simultaneously ($d' = 1$, immediate or delayed feedback). In all conditions of Experiment 2, the base-rate ratio was 3:1.

Methods

Participants—University of Central Florida undergraduates received course credit for participation. In the $d' 1$ conditions preceded by baseline training, 36 and 34 participants completed the immediate and delayed feedback conditions, respectively. In the $d' 2.2$ conditions preceded by baseline training, 26 and 24 participants completed immediate and delayed conditions, respectively. In the conditions without baseline training, 23 and 21 participants completed the immediate and delayed conditions respectively.

Stimuli—The stimulus sets used in Experiment 1 were used in Experiment 2.

Procedure

Event timing for the immediate feedback conditions matched the response-training conditions of Experiment 1. In the delayed feedback conditions, the duration between stimulus offset and feedback onset was 5 seconds. As in Maddox et al. (2003), a random noise mask replaced the stimulus after each response and remained there until feedback was presented (i.e., the mask was onscreen for 5 seconds in the delayed feedback condition). Because participants made a response on each trial in Experiment 2, there were only 5 blocks of trials (rather than 10 as in Experiment 1). Responses were followed by feedback (the correct category label). Finally, because most Experiment 1 participants who met the baseline accuracy criterion did so rather quickly, participants in Experiment 2 completed only 300 baseline trials.

Results

In conditions preceded by baseline training, 69% of participants in the $d' 1$ conditions and 84% in the $d' 2.2$ conditions reached the accuracy criterion (not less than 2% below optimal) by the end of training. Participants failing to reach this criterion were omitted from further analysis. Of those remaining, 24 and 24 were left in the $d' 1$ immediate and delayed conditions, respectively. In the $d' 2.2$ immediate and delayed conditions, 20 and 22 participants, respectively, remained for further analysis. All participants from the two conditions without baseline training were included in our analysis.

As in Experiment 1, we compared signal detection β values across experimental conditions (Table 2 shows average β values and accuracy). We performed a mixed-factor ANOVA (between: 2 d' levels X 2 training delays and including presence/absence of baseline training; within: 5 blocks of trials).

Values of β were larger in $d' 2.2$ ($M = 2.07$, $SE = 0.07$) than in $d' 1$ ($M = 1.28$, $SE = 0.04$) conditions, $F(1,137) = 81.81$, $p < 0.001$, $\eta_p^2 = 0.37$. There also was a main effect of block, with β values increasing toward optimal over blocks, $F(4,548) = 41.52$, $p < 0.001$, $\eta_p^2 = 0.23$. The increase toward optimal β over blocks was more pronounced in the $d' 2.2$ conditions than in $d' 1$ conditions, $F(4,548) = 19.8$, $p < 0.001$, $\eta_p^2 = 0.13$. The effect of feedback delay was not significant when all conditions were included in the analysis,

$F(1,137) = 0.32, p = 0.575, \eta_p^2 = 0.002$, nor was there a significant effect of baseline training, $F(1,137) = 0.39, p = 0.534, \eta_p^2 = 0.003$.

As in Experiment 1, we conducted an analysis of β values in the $d' = 1$ conditions (with and without baseline training). For these low-discriminability conditions, we did find a significant effect of feedback delay, $F(1,101) = 5.09, p = 0.026, \eta_p^2 = 0.05$, as well as a significant effect of block, $F(4,404) = 16.76, p < 0.001, \eta_p^2 = 0.14$. β values were closer to optimal in immediate than in delayed feedback conditions and increased over blocks (Fig. 2).

Discussion

When category discriminability was low (when $d' = 1$), performance was closer to optimal in the immediate feedback conditions than in the delayed conditions. This again conforms to the prediction that implicit learning is involved in acquisition and use of base-rate knowledge. Performance was disrupted when a categorization response was not immediately met with corrective feedback. This is predicted by COVIS.

To help relate our results to previous studies of base-rate learning through direct experience, we sought corroboration of our signal-detection results by fitting computational models to the data. The modeling results are briefly summarized next.

Modeling results

We fit two decision-bound models to each data set (separately for each participant, condition, and block). The models use maximum likelihood estimation to find a boundary value that best predicts the participant's classification responses to the stimuli presented during training (modeling details can be found elsewhere; e.g., Maddox, 2002). The location of the bound in stimulus space reflects the participant's response bias toward either category. For example, a decision bound equidistant between the category means would reflect complete insensitivity to base-rates. Decision-bound models have informed our interpretation of base-rate learning from experience in several previous studies and allow us to instantiate and test specific hypotheses (Bohil & Maddox, 2003a, b; Bohil & Maddox, 2001).

We compared a free-boundary model and an unbiased-boundary model. The free-boundary model included two free parameters: a decision-bound parameter and noise parameter (which estimates a combination of perceptual and criterial noise). The model is free to find the criterion value that best describes the participant's responses. The unbiased-boundary model has only a single free parameter: the noise parameter. In the unbiased-boundary model, the decision criterion value is fixed at an unbiased value (i.e., where $\beta = 1$, as would be optimal if the base-rate ratio were 1:1). This reflects the assumption that the participant was not influenced by the base-rate manipulation at all. Model fits were compared based on Akaike Weights (Wagenmakers & Farrell, 2004).

Table 3 displays the proportion of cases for which the unbiased-boundary model provided the most parsimonious account. The free-boundary model often was more parsimonious than the unbiased-boundary model. However, the unbiased-boundary model fared better in the

observational ($z = 2.52, p < 0.05$) and delayed feedback ($z = 5.86, p < 0.001$) conditions (collapsed over blocks). There were more cases of base-rate neglect in these conditions than in conditions supporting implicit-learning performance.

This conclusion is further supported by examining β values from the free-boundary model (Fig. 3). Mirroring the signal-detection β values, the modeling analyses show that base-rate influence was smaller in conditions where implicit learning was disrupted. Criterion values differed significantly between observational and response training, $F(1,98) = 8.58, p = 0.004, \eta_p^2 = 0.08$, and nearly so between immediate and delayed feedback conditions, $F(1,79) = 3.36, p = 0.07, \eta_p^2 = 0.04$.

General discussion

We examined base-rate influence on decision-criterion placement during classification learning. Our goal was to distinguish between direct experience and the role of explicit or implicit learning. We replicated manipulations from previous work showing implicit and explicit learning differences in categorization. In accord with the explanation proposed by Koehler (1996; see also Spellman, 1996), we found that base-rate influence was smaller when the performance of the implicit learning system was disrupted.

One advantage of examining the role of implicit learning using two different experimental designs was that we gain more certainty that implicit learning underlies performance. For example, perhaps performance was worse in the observational conditions of Experiment 1 due to lower task engagement than in response conditions. In Experiment 2, participants responded on every trial, so we presume engagement was roughly equivalent for immediate and delayed feedback. Furthermore, in Experiment 1, we disrupt implicit learning by removing the procedural learning component of the task. In Experiment 2, we delayed feedback beyond the window of effectiveness for dopamine to reinforce learning in the basal ganglia. Both of these effects are consistent with the explanation offered by COVIS.

The current studies were intended to expand upon prior work testing COVIS predictions by disrupting implicit learning during base-rate acquisition. However, truly dissociating implicit and explicit system contributions during base-rate acquisition will require a study that includes a manipulation to disrupt explicit learning as well (e.g., a dual task that uses working memory resources). In previous COVIS studies, occupying working memory resources with a secondary task has been found to disrupt learning of a verbalizable (explicitly-learned) rule more than learning of a nonverbalizable (implicitly-learned) rule (Zeithamova & Maddox, 2006). A secondary working-memory task should not diminish the effect of unequal category base-rates if learning is implicit.

Indeed, a potential alternative explanation for our results is related to availability of working memory and the rate of drift in memory for the location of the decision criterion. Previous work has shown that delaying feedback in a rule-based task—when working memory also is taxed—reduces classification accuracy (Ell, Ing, & Maddox, 2009). Perhaps the performance decrements we report are due to increased working memory demands. If the performance decrement in our delayed feedback condition were due to an increase in

working memory demands, then this should result in greater variability in criterion placement (i.e., criterial noise). To test this possibility, we examined the noise parameters from the free-boundary model for each experiment. In Experiment 2, we found no reliable difference in noise estimates between immediate and delayed feedback conditions. This corresponds to Ell et al.'s findings for a single-criterion learning condition and further underscores our interpretation that performance decrement in the delayed condition is due to disruption of implicit learning. However, in Experiment 1, we did find some evidence that criterial noise was higher in the early stages of observational training. Although this reached significance only for the third block of test trials [$t(116) = 2.4$, $p = 0.018$], it also was near significance for blocks 1 and 2. As participants approached asymptote in the task (by blocks 4 and 5 of testing), there was no reliable difference in noise estimates between response and observational conditions.

Conclusions

This research sheds light on the role of implicit learning in acquiring base-rate knowledge through direct experience. The results would seem to underscore the difficulty in using base-rates reported in studies that simply present this information in summary form. Beyond base-rate effects, these results add to our understanding of decision criterion learning in general, which has been the focus of many studies (e.g., Kubovy & Healy, 1977; Treisman & Williams, 1984; Erev, 1998). Our results suggest a role for the implicit learning system in categorization that goes beyond associating directly-observable stimulus features to categories (as has been the focus of prior tests of COVIS). And although COVIS predicts the discovery of relevant stimulus dimensions in rule-based tasks via explicit hypothesis testing, there is still much to learn about how the decision criterion changes over time and how a criterion is ultimately selected by learners.

Acknowledgments

The authors thank Todd Maddox for helpful comments regarding this research. We also thank Harry Mohan, Matthew Murray, and Troy Schiebel for their help in collecting data.

References

1. Ashby FG, Maddox WT. Human category learning 2.0. *Annals of the New York Academy of Sciences*. 2011; 1224:147–161. [PubMed: 21182535]
2. Ashby FG, Alfonso-Reese LA, Turken AU, Waldron EM. A neuropsychological theory of multiple systems in category learning. *Psychological Review*. 1998; 105:442–481. [PubMed: 9697427]
3. Ashby FG, Ell SW, Waldron EM. Procedural learning in perceptual categorization. *Memory & Cognition*. 2003; 31(7):1114–1125. [PubMed: 14704026]
4. Ashby FG, Maddox WT, Bohil CJ. Observational versus feedback training in rule-based and information-integration category learning. *Memory & Cognition*. 2002; 30:665–676.
5. Bohil CJ, Maddox WT. Category discriminability, baserate, and payoff effects on perceptual categorization. *Perception & Psychophysics*. 2001; 63:361–376. [PubMed: 11281110]
6. Bohil CJ, Maddox WT. A test of the optimal classifier's independence assumption in perceptual categorization. *Perception & Psychophysics*. 2003a; 65(3):478–493. [PubMed: 12785076]
7. Bohil CJ, Maddox WT. On the generality of optimal versus objective classifier feedback effects on decision criterion learning in perceptual categorization. *Memory & Cognition*. 2003b; 31(2):181–198. [PubMed: 12749461]

8. Ell SW, Ing AD, Maddox WT. Criterial noise effects on rule-based category learning: The impact of delayed feedback. *Attention, Perception, & Psychophysics*. 2009; 71(6):1263–1275.
9. Erev I. Signal detection by human observers: A cutoff reinforcement learning model of categorization decisions under uncertainty. *Psychological Review*. 1998; 105(2):280–298. [PubMed: 9669925]
10. Estes WK. Base-rate effects in category learning: A comparison of parallel network and memory storage-retrieval models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1989; 15:556–571.
11. Gluck MA, Bower G. From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*. 1988; 117:227–247. [PubMed: 2971760]
12. Holyoak KJ, Spellman BA. Thinking. *Annual Review of Psychology*. 1993; 44:265–315.
13. Koehler JJ. The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences*. 1996; 19:1–53.
14. Kubovy M, Healy AF. The decision rule in probabilistic categorization: What it is and how it is learned. *Journal of Experimental Psychology: General*. 1977; 106(4):427–446.
15. Maddox WT. Toward a unified theory of decision criterion learning in perceptual categorization. *Journal of the Experimental Analysis of Behavior*. 2002; 78:567–595. [PubMed: 12507020]
16. Maddox WT, Ashby FG, Bohil CJ. Delayed feedback effects on rule-based and information-integration category learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*. 2003; 29:650–662.
17. Spellman BA. The implicit use of base rates in experiential and ecologically valid tasks. *Behavioral and Brain Sciences*. 1996; 19:38.
18. Treisman M, Williams TC. A theory of criterion setting with an application to sequential dependencies. *Psychological Review*. 1984; 91(1):68–111.
19. Wagenmakers E-J, Farrell S. AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*. 2004; 11(1):192–196. [PubMed: 15117008]
20. Zeithamova D, Maddox WT. Dual-task interference in perceptual category learning. *Memory & Cognition*. 2006; 34(2):387–398. [PubMed: 16752602]

Fig. 1. Average signal detection β values for $d'1$, 3:1 base-rate conditions in Experiment 1. Optimal β value = 3

Fig. 2.
Average signal detection β values for $d'1$ conditions in Experiment 2. Optimal β value = 3

Fig. 3.
Average free- β model β values for Experiment 1 (top) and Experiment 2 (bottom)

Table 1
Average proportion correct (and β values) observed in each condition and block of Experiment 1

Condition	Block 1	Block 2	Block 3	Block 4	Block 5
<i>d</i> 1 2:1 Obs	0.64 (0.99)	0.64 (1.14)	0.65 (1.16)	0.66 (1.19)	0.63 (1.15)
<i>d</i> 1 2:1 Rsp	0.63 (0.99)	0.64 (1.13)	0.64 (1.21)	0.67 (1.33)	0.66 (1.46)
<i>d</i> 1 3:1 Obs	0.65 (1.05)	0.67 (1.25)	0.68 (1.35)	0.68 (1.34)	0.67 (1.25)
<i>d</i> 1 3:1 Rsp	0.67 (1.12)	0.70 (1.29)	0.70 (1.43)	0.70 (1.42)	0.70 (1.62)
<i>d</i> 2.2 2:1 Obs	0.81 (1.10)	0.84 (1.79)	0.84 (1.87)	0.84 (2.26)	0.83 (2.42)
<i>d</i> 2.2 2:1 Rsp	0.80 (0.96)	0.85 (2.37)	0.84 (1.66)	0.86 (2.18)	0.85 (1.98)
<i>d</i> 2.2 3:1 Obs	0.84 (2.26)	0.85 (2.71)	0.86 (2.84)	0.85 (2.62)	0.85 (2.94)
<i>d</i> 2.2 3:1 Rsp	0.86 (1.61)	0.88 (2.42)	0.88 (2.87)	0.88 (2.64)	0.89 (2.63)
<i>d</i> 1 3:1 Obs NB	0.64 (1.14)	0.64 (1.20)	0.66 (1.28)	0.65 (1.31)	0.63 (1.26)
<i>d</i> 1 3:1 Rsp NB	0.66 (1.22)	0.68 (1.41)	0.67 (1.53)	0.68 (1.57)	0.68 (1.64)

Obs, observational-training; Rsp, response training; NB, no baseline training.
Average β values in parentheses. Optimal $\beta = 2$ in 2:1 base-rate conditions, 3 in 3:1 base-rate conditions.

Table 2
Average proportion correct (and β values) observed in each condition and block of Experiment 2

Condition	Block 1	Block 2	Block 3	Block 4	Block 5
<i>d</i> 1 3:1 Del	0.66 (1.00)	0.69 (1.23)	0.68 (1.23)	0.69 (1.23)	0.69 (1.29)
<i>d</i> 1 3:1 Imm	0.68 (1.17)	0.69 (1.27)	0.70 (1.37)	0.70 (1.38)	0.70 (1.34)
<i>d</i> 2.2 3:1 Del	0.84 (1.06)	0.86 (1.90)	0.85 (1.95)	0.87 (2.50)	0.87 (3.06)
<i>d</i> 2.2 3:1 Imm	0.83 (1.13)	0.85 (1.95)	0.87 (2.29)	0.87 (2.16)	0.88 (2.76)
<i>d</i> 1 3:1 Del NB	0.63 (1.07)	0.68 (1.26)	0.67 (1.34)	0.65 (1.23)	0.68 (1.40)
<i>d</i> 1 3:1 Imm NB	0.63 (1.13)	0.70 (1.31)	0.70 (1.39)	0.71 (1.30)	0.72 (1.60)

Del, delayed feedback; Imm, immediate feedback; NB, no baseline training.
Average β values in parentheses. Optimal $\beta = 2$ in 2:1 base-rate conditions, 3 in 3:1 base-rate conditions.

Table 3
Proportion of participants whose data were best fit by the unbiased boundary model

Training type	Block 1	Block 2	Block 3	Block 4	Block 5
Experiment 1					
Response	0.31	0.22	0.12	0.12	0.07
Observational	0.38	0.28	0.23	0.17	0.17
	#	#	$p = 0.077$	#	$p = 0.084$
Experiment 2					
Immediate	0.32	0.09	0.05	0.05	0
Delayed	0.63	0.22	0.24	0.2	0.2
	***	*	**	*	***

Significance of comparisons across training conditions are reported below proportions.

- * $p < 0.05$.
- ** $p < 0.01$.
- *** $p < 0.001$.
- # $p > 1$.