Research Article

Psychological Science
2014, Vol. 25(2) 447–457

DOI: 10.1177/0956797613509112
pss.sagepub.com

Deferred Feedback Sharply Dissociates Implicit and Explicit Category Learning

J. David Smith[1], Joseph Boomer[1], Alexandria C. Zakrzewski[1], Jessica L. Roeder[2], Barbara A. Church[1], and F. Gregory Ashby[2]
[1]Department of Psychology, University at Buffalo, The State University of New York, and [2]Department of Psychological and Brain Sciences, University of California, Santa Barbara

Abstract

The controversy over multiple category-learning systems is reminiscent of the controversy over multiple memory systems. Researchers continue to seek paradigms to sharply dissociate explicit category-learning processes (featuring category rules that can be verbalized) from implicit category-learning processes (featuring learned stimulus-response associations that lie outside declarative cognition). We contribute a new dissociative paradigm, adapting the technique of deferred-rearranged reinforcement from comparative psychology. Participants learned matched category tasks that had either a one-dimensional, rule-based solution or a multidimensional, information-integration solution. They received feedback either immediately or after each block of trials, with the feedback organized such that positive outcomes were grouped and negative outcomes were grouped (deferred-rearranged reinforcement). Deferred reinforcement qualitatively eliminated implicit, information-integration category learning. It left intact explicit, rule-based category learning. Moreover, implicit-category learners facing deferred-rearranged reinforcement turned by default and information-processing necessity to rule-based strategies that poorly suited their nominal category task. The results represent one of the strongest explicit-implicit dissociations yet seen in the categorization literature.

Keywords

category learning, implicit cognition, explicit cognition, associative learning, category rules, procedural learning, cognitive neuroscience

Received 6/17/13; Revision accepted 9/18/13

Categorization is an essential cognitive function and a focus of cognitive research (e.g., Ashby & Maddox, 2010; Brooks, 1978; Feldman, 2000; Knowlton & Squire, 1993; Medin & Schaffer, 1978; Murphy, 2003; Nosofsky, 1987; Smith, Redford, & Haas, 2008). A lasting issue is whether one or multiple category-learning systems are necessary to account for the diverse categorization abilities of humans. A similar issue has been debated in the memory literature. In fact, these debates are related because proposed category-learning systems might map onto proposed memory systems (Ashby & O'Brien, 2005). Learning is a process of laying down memory traces, and there seems no reason why memory systems should not also learn categories. In that case, there may be as many category-learning systems as memory systems. In the research reported here, we tested for multiple category-learning systems using a distinctively new paradigm.

Categorization researchers have described trade-offs that seem to support the hypothesis that there are multiple systems. For example, different processes seem to dominate categorization at early versus late stages of category learning (Cook & Smith, 2006; Smith, Chapman, & Redford, 2010; Smith & Minda, 1998; Wasserman, Kiedinger, & Bhatt, 1988), categorization of small versus large categories (Blair & Homa, 2003; Homa, Sterling, & Trepel, 1981; Minda & Smith, 2001), and categorization based on rules that are easy versus difficult to describe verbally (Ashby & Maddox, 2010). A growing consensus ascribes to humans multiple categorization capacities

Corresponding Author:
J. David Smith, 346 Park Hall, University at Buffalo, The State University of New York, Buffalo, NY 14260
E-mail: psysmith@buffalo.edu

(e.g., rule learning, prototype abstraction) that specialize in different aspects of learning and rely on different forms of memory (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby & Ell, 2001; Erickson & Kruschke, 1998; Homa et al., 1981; Maddox & Ashby, 2004; Rosseel, 2002; Smith et al., 2012; Smith & Minda, 1998). Even so, some researchers question multiple-system interpretations of some results (e.g., Nosofsky & Johansen, 2000), and others argue that all categorization phenomena can be explained using a unitary, exemplar-based process in which every previously seen exemplar from all relevant categories is accessed and compared with the current stimulus before a category judgment is made (e.g., Newell, Dunn, & Kalish, 2010; Nosofsky, Stanton, & Zaki, 2005). Our research helps resolve this issue.

Rule-Based and Information-Integration Categorization

Our empirical approach draws on the cognitive neuroscience of categorization (Ashby & Ell, 2001; Ashby & Valentin, 2005; Maddox & Ashby, 2004). This area distinguishes an explicit categorization system that recruits declarative memory from an implicit system that recruits procedural memory. The explicit system learns by actively testing hypotheses using working memory and executive attention. It learns quickly, through sudden realizations of category rules that participants easily describe verbally. For example, people explicitly know a square's defining characteristics. In contrast, the implicit system learns associatively through procedural-learning processes akin to conditioning. It learns slowly, relying on temporally contiguous reinforcement signals. Participants generally cannot describe their implicit categorization strategies. For example, people correctly categorize wolves and German shepherds, but they do not easily explain how they do so.

Much of the evidence for these systems comes from rule-based (RB) and information-integration (II) category-learning tasks (Fig. 1). Each exemplar in these tasks is defined by its values on perceptually separable *X* and *Y* dimensions. For example, each stimulus might be a single line that varies across trials in length (Dimension *X*) and orientation (Dimension *Y*). In Figure 1, the gray and black symbols, respectively, denote the dimensional values of specific Category A and Category B members. Figure 1a shows possible stimuli for an RB task. Only variation in *X* carries valid category information; low and

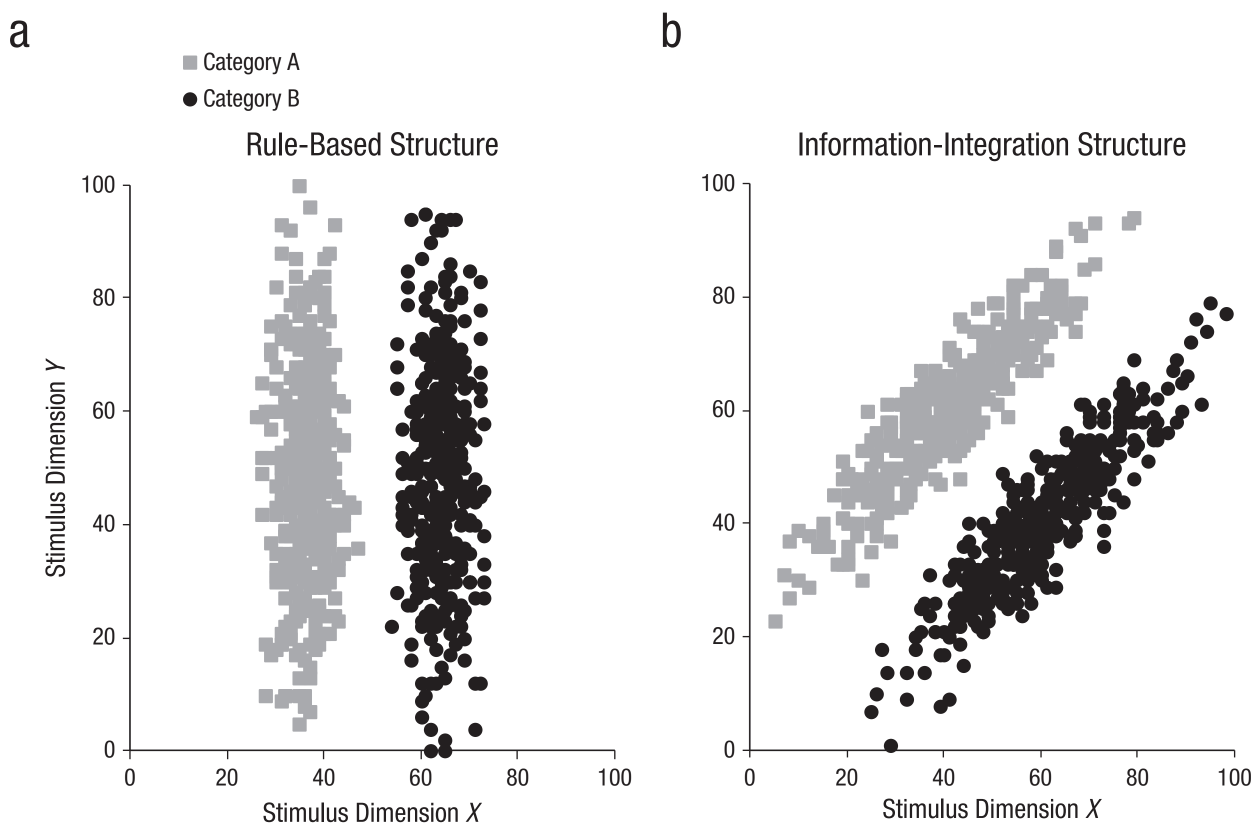

Fig. 1. A rule-based category structure (a) and an information-integration category structure (b), depicted within an abstract 101 × 101 stimulus space.

high values on Dimension *X* define Category A and B members, respectively. The participant must discover this rule from successive presentations of single category exemplars with feedback. This is an RB task because the solution is a one-dimensional rule. The rule is explicit because it can be verbalized and is discoverable through explicit hypothesis testing.

In Figure 1b, Dimensions *X* and *Y* carry partially valid category information. To categorize stimuli accurately, the participant must learn some principle of dimensional integration. This is an II task. One-dimensional rules are not optimal. A vertical or horizontal category boundary will not partition the categories sufficiently and will cause errors. The cognitive system accomplishes dimensional integration, but it does so implicitly and procedurally. Humans cannot explain their solution of an II task verbally, especially when the stimulus dimensions are in different units (e.g., length and orientation).

The RB and II tasks are elegant mutual controls. They are matched for category size, within-category exemplar similarity, between-category exemplar separation, class discriminability (e.g., d'), and the proportion correct that is achievable by an ideal observer. The category structures in Figure 1 are simply rotations of one another through stimulus space. Therefore, there is no objective, a priori difference in difficulty between RB and II tasks. Smith et al. (2011) confirmed this equivalence by showing that pigeons (*Columba livia*) learn RB and II tasks equally well and at the same rate. Pigeons may learn these tasks at the same rate because they lack an explicit category-learning system that selectively advantages RB category learning. Humans generally learn RB tasks faster than II tasks, perhaps because they have that explicit system. If humans' learning-rate difference arose because the II task is inherently difficult, then a less cognitively sophisticated species (pigeons) should be more challenged on the II task (relative to the RB task) than humans are. That pigeons learn the tasks at the same rate is strong evidence that humans' learning-rate difference arises because they learn the tasks differently, not because one task is difficult. In the same way, multiplication is faster than repeated addition, not because it is easier, but because it is a different process that unfolds differently.

Many dissociations between RB and II category learning have been demonstrated. For example, II learning is selectively impaired when reinforcement on categorization trials is delayed for several seconds (Maddox, Ashby, & Bohil, 2003; Maddox & Ing, 2005), when learning is unsupervised (Ashby, Queller, & Berretty, 1999), and when category knowledge is imparted observationally, not through trial-based reinforcement (Ashby, Maddox, & Bohil, 2002). II learning is apparently served by a cascade of temporally constrained events (perception, response selection, reinforcement—Maddox & Ashby, 2004).

In contrast, RB category learning may be more robust to reinforcement delays, unsupervised learning conditions, and so forth. RB learning may rely on hypotheses actively held in working memory. Other dissociations—for example, that RB learning is selectively hurt when working memory resources are occupied by a concurrent task—support this idea (Waldron & Ashby, 2001). Consequently, RB learning potentially has great flexibility in application. Because its category knowledge is held in declarative consciousness, it can be applied or adjusted before, during, or after a trial, and possibly even after the outcomes from several trials.

No single-system model and no simple difficulty hypothesis that has been proposed can account for even a few of the dissociations between RB and II category learning that have been reported. In contrast, the multiple-systems framework described here essentially predicts all of these dissociations a priori.

The evidence for multiple systems notwithstanding, the theoretical stakes are high, just as when the possibility of multiple memory systems emerged. Some categorization researchers apply a strict parsimony standard to defend a single-system account of categorization, and work continues to definitively evaluate the multiple-systems framework. Here, we introduce a new empirical dissociation that may qualitatively distinguish implicit and explicit categorization. The temporal constraints and flexibility of II and RB category learning, respectively, are crucial to the present test of the implicit-explicit framework in categorization.

A New Empirical Approach

Our paradigm incorporates a technique from recent cross-species studies (e.g., Smith, Beran, Redford, & Washburn, 2006). Smith et al. sought to prevent monkeys' associative learning, to keep them from using trial-by-trial reinforcement to form stimulus-response linkages. The idea was to see if macaques, like humans, could supply instead their own cognitive construal of a task. Yet the researchers also had to include enough reinforcement to sustain the animals' participation.

The technique of deferred-rearranged reinforcement (hereafter, deferred reinforcement) met these requirements. Monkeys completed trial blocks with no feedback. At each block's end, they received together the reinforcements from all correct trials and then together the timeouts from all error trials. The processes of conditioning were defeated. The monkeys could not know which stimuli and responses they had gotten wrong or right. They could not learn stimulus-response pairs associatively.

Smith et al. (2006) showed that this technique did make the task's associative structure invisible for at least one macaque who clearly supplied his own cognitive construal instead.

Predictions

This technique is ideally suited for studying and possibly dissociating RB and II category learning. Deferred reinforcement should defeat the reinforcement-based processes underlying II learning. One prediction of our study was that II learning would collapse under deferred reinforcement. In contrast, RB learners—by current theory—would have in mind their hypothesis and could evaluate its success at block's end just as at trial's end. Thus, a second prediction was that RB learning would flourish under deferred reinforcement.

Finally, by current theory, the explicit system emphasizes one-dimensional rules. If deferred reinforcement disables II but not RB category learning, II participants facing deferred reinforcement might turn—by information-processing necessity—to one-dimensional rules instead. Thus, our third prediction was that participants in the II condition would supply their own RB task construal because that was what they still could do—much as the macaque in the study by Smith et al. (2006) supplied his own construal of a task with an associative structure that was made invisible.

The confirmation of these predictions would provide one of the clearest dissociations yet seen between RB and II category learning.

Method

This experiment included four between-participants conditions created by crossing two category structures to be learned (RB, II) with two reinforcement conditions (immediate, deferred).

Participants

University at Buffalo undergraduates with normal or corrected-to-normal vision participated for course credit. Participants' data were excluded if they completed fewer than 300 trials (1 and 4 participants excluded, respectively, from the RB-immediate and II-immediate conditions) or if they showed significantly lower performance during their last 100 trials than during their first 100 trials (1, 2, 3, and 2 participants excluded, respectively, from the RB-immediate, RB-deferred, II-immediate, and II-deferred conditions). The final sample included 84 participants divided equally among the four conditions (i.e., 21 participants in each condition).

Stimuli

The stimuli were unframed rectangles containing green lit pixels, presented on a black background in the center at the top of a 17-in. computer monitor (resolution of 800 × 600 pixels). They were viewed from a distance of about 24 in.

The stimulus rectangles varied in size and pixel density. Both dimensions had 101 levels (Levels 0–100). Rectangle width and height (in screen pixels) were calculated as 100 + *level* and 50 + *level*/2, respectively. Thus, rectangle size ranged from 100 × 50 (Level 0) to 200 × 100 (Level 100). Pixel density—that is, the proportion of pixel positions that were illuminated—was calculated as $0.05 \times 1.018^{level}$. Thus, density varied from .0500 (Level 0) to .2977 (Level 100). Figure 2 shows the stimuli in the four corners of the stimulus space.

Category structures

Figure 1 shows the category structures used: a vertical RB structure with size (Dimension *X*) relevant and a major-diagonal II structure with size and density (Dimension *Y*) relevant. The categories were defined by bivariate normal distributions along the stimulus dimensions, as specified in Table 1. Each exemplar was selected as a coordinate pair in the 101 × 101 space, and the abstract values were transformed into concrete size and density values. Each participant received his or her own sample of randomly selected category exemplars appropriate to the assigned task. To control for statistical outliers, we did not present exemplars whose Mahalanobis distance (e.g., Fukunaga, 1972) from the category mean exceeded 3.0.

Procedure

Participants were assigned randomly to the RB or II task and the immediate- or deferred-reinforcement condition. Trials continued until the 55-min session ended or the participant completed 600 trials.

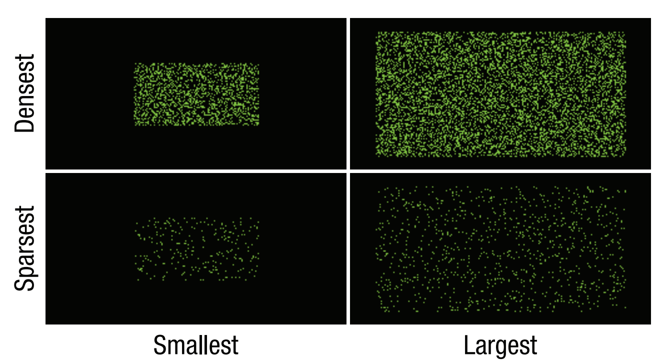

Fig. 2. Illustration of the 101 × 101 stimulus space.

Table 1. Distributional Characteristics for the Two Category Tasks

Task and category	Mean of *X*	Mean of *Y*	Variance of *X*	Variance of *Y*	*XY* covariance
Rule-based task					
Category A	35.86	50.00	16.33	355.55	0
Category B	64.14	50.00	16.33	355.55	0
Information-integration task					
Category A	40.00	60.00	185.94	185.94	169.61
Category B	60.00	40.00	185.94	185.94	169.61

Below each to-be-categorized stimulus were the letters "A" (on the left) and "B" (on the right), along with a central cursor. Participants depressed the "S" or "L" key on the computer keyboard to move the cursor across the screen until it reached the "A" or "B," to indicate which category they thought the stimulus was a member of. The response keys corresponded spatially to the "A" and "B" response icons on the screen.

In the immediate-reinforcement condition, after a correct response, participants heard a whoop sound, earned a point, and saw their accumulated points (correct responses – incorrect responses). After an incorrect response, they heard a buzz sound, lost a point, received a 4-s time-out, and saw their accumulated points. The next trial followed.

In the deferred-reinforcement condition, participants completed each of six trials in a block without feedback. After each response, the program simply presented the next trial. At block's end, participants received their positive outcomes grouped together (e.g., several whoops separated by 0.5 s for correct responses) and then their negative outcomes grouped together (e.g., several buzzes separated by 4 s for incorrect responses). Then they were updated on their accumulated points. The next trial block followed. (Fig. S1 in the Supplemental Material illustrates in more detail the progression of stimuli, responses, and reinforcements that characterized the immediate- and deferred-reinforcement conditions.)

Formal modeling

We used an *RB model* to specify the vertical or horizontal line through the stimulus space that would best partition a participant's "A" and "B" responses. We used an *II model* to specify the slope and intercept of the nonhorizontal, nonvertical line through stimulus space that would best partition a participant's "A" and "B" responses. The best-fitting values for the parameters in the models were estimated using maximum-likelihood methods. That is, we evaluated which model would have created with maximum likelihood the participant's distribution of responses in the stimulus space (details in Maddox & Ashby, 1993). The Bayesian information criterion (BIC; Schwarz, 1978) determined the best-fitting model ($BIC = r \times \ln N - 2\ln L$, with r = the number of free parameters, N = the sample size, and L = the model's likelihood given the data).

Results

Accuracy-based analyses

The proportion of correct responses on the last 100 trials was examined in a two-way analysis of variance (ANOVA) with task (RB, II) and reinforcement condition (immediate, deferred) as between-participant factors. The crucial result was a significant interaction between task and condition, $F(1, 80) = 4.03$, $p = .0481$, $\eta_p^2 = .0479$, reflecting that deferred reinforcement compromised II performance selectively.

Performance in the RB-immediate condition ($M = .82$ correct) was statistically indistinguishable from performance in the RB-deferred condition ($M = .84$ correct), $t(40) = -0.33$, $p = .744$, Cohen's $d = -0.10$. There were 14 and 15 strong learners (terminal performance ≥ .80) in these conditions, respectively. Thus, deferred reinforcement had no cost for RB category learning.

Performance in the RB-immediate condition ($M = .82$ correct) was also indistinguishable from performance in the II-immediate condition ($M = .77$ correct), $t(40) = 0.700$, $p = .488$, Cohen's $d = 0.22$. There were 14 and 11 strong learners in these conditions, respectively.

In contrast, performance in the II-immediate condition ($M = .77$ correct) was distinguishable from performance in the II-deferred condition ($M = .64$ correct), $t(40) = 3.76$, $p = .0005$, Cohen's $d = 1.16$. There were 11 and 0 strong learners in these conditions, respectively. Thus, deferred reinforcement had a high cost for II category learning. As we discuss shortly, the accuracy-based analyses sharply understated that cost.

Figure 3a shows a backward learning curve for the RB-deferred condition. To create this graph, we divided the trials into 20-trial blocks. From these blocks, we then excluded trials on which the *X* dimension (size) had a value greater than 40 or less than 60. In this way, we

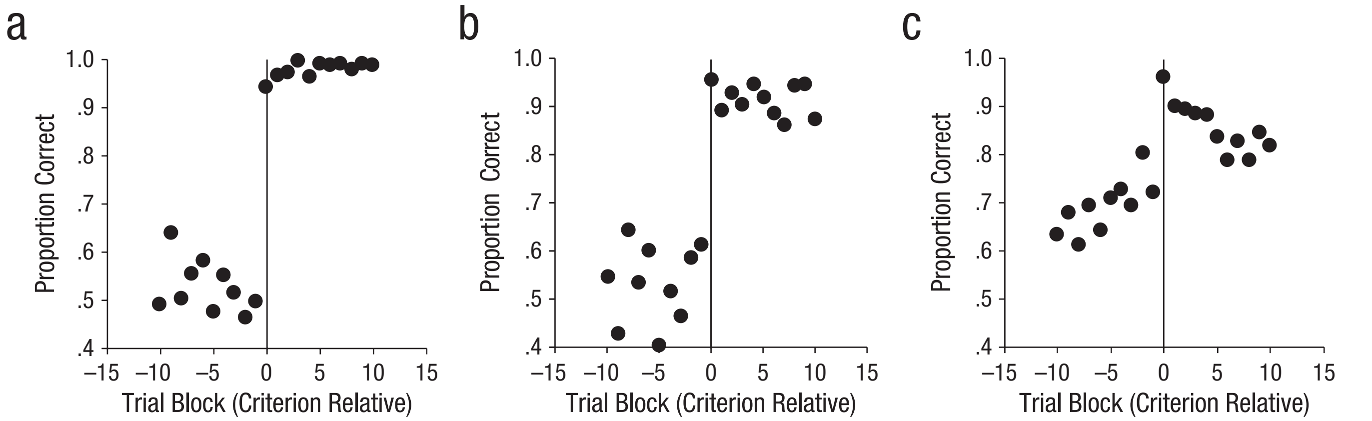

Fig. 3. Backward learning curves: proportion correct as a function of trial block (Block 0 = the block at which the learning criterion was met). The graph in (a) shows the path to rule-based (RB) learning among strong learners in the RB-deferred condition. Analogous graphs were created for performance on the information-integration task, assuming an RB standard of correct and incorrect performance. These graphs reflect the extent to which an RB strategy was used among strong learners who received (b) deferred reinforcement and (c) immediate reinforcement. (See the text for details on how these curves were created.)

accommodated variation in participants' rule criterion near an *X* value of 50, the true category break point. Next, we counted how many of the remaining trials were responded to correctly, scoring as correct an "A" response if *X* was small (≤ 40) and a "B" response if *X* was large (≥ 60). We then found the participants (n = 15) who met the criterion of reaching and sustaining .85 accuracy over five consecutive blocks of 20 trials. A relatively strict definition of arrival at criterion was adopted in this case because this analysis excluded the most difficult Category A and Category B trials. We aligned the trial blocks at which participants reached this criterion (Block 0) to depict the path by which they solved the RB task. As the figure shows, RB performance transformed at Block 0 from near chance (.53 correct) to near ceiling (.98). Figure 3a is perfectly intuitive if—and probably only if—one assumes the sudden discovery of a category rule. This figure essentially defines the RB category-learning process that the literature has debated.

Figure 3b shows a backward learning curve for the II-deferred condition, but with a twist: We assumed an RB standard of correct and incorrect performance based on Dimension *Y* (density). Specifically, in each 20-trial block, we excluded trials on which *Y* had a value greater than 40 or less than 60. For the remaining trials, we counted as correct an "A" response if *Y* was large (≥ 60) and a "B" response if *Y* was small (≤ 40). We then found all the participants (n = 14) who met the learning criterion (.85 accuracy over five consecutive 20-trial blocks) in this rule-focused way and aligned the trial blocks at which they reached this criterion (Block 0) to depict the path by which they chose an RB strategy in their II task. These participants also jumped to rule use suddenly. This is a remarkable result because they were not reinforced for RB learning; they were given an II task with an II reinforcement contingency. Clearly, the reinforcement contingency was not controlling learning. Participants self-chose their RB strategy cognitively, facing the exigency of deferred reinforcement.

We repeated this analysis for participants in the II-immediate condition (Fig. 3c). There was strong learning prior to Block 0 and only a small jump in performance at criterion. Performance fell back down after the criterion blocks to a level continuous with that before criterion. Thus, the arrival at criterion was an artificial, statistical occurrence. The RB strategy was not evident in the II-immediate condition (Fig. 3c), but it emerged strongly in the II-deferred condition (Fig. 3b).

Model-based analyses

We modeled participants' last 100 trials to determine whether they adopted appropriate decision strategies and whether deferred reinforcement disrupted those strategies or altered them in a theoretically meaningful way. Figure 4a shows the modeling results for 15 RB-immediate participants. (The poor performance of the 6 other participants in this condition was not consistent with RB or II strategies; models for their performance indicated that they had a guessing strategy and no definite decision bound that could be drawn). The 15 decision bounds were primarily organized along the midline of the stimulus space's *X* dimension. Many participants found the RB task's adaptive solution—a one-dimensional size rule.

Figure 4b shows the modeling results for 17 RB-deferred participants (modeling indicated that 4 participants in this condition were guessers). The similarity to Figure 4a

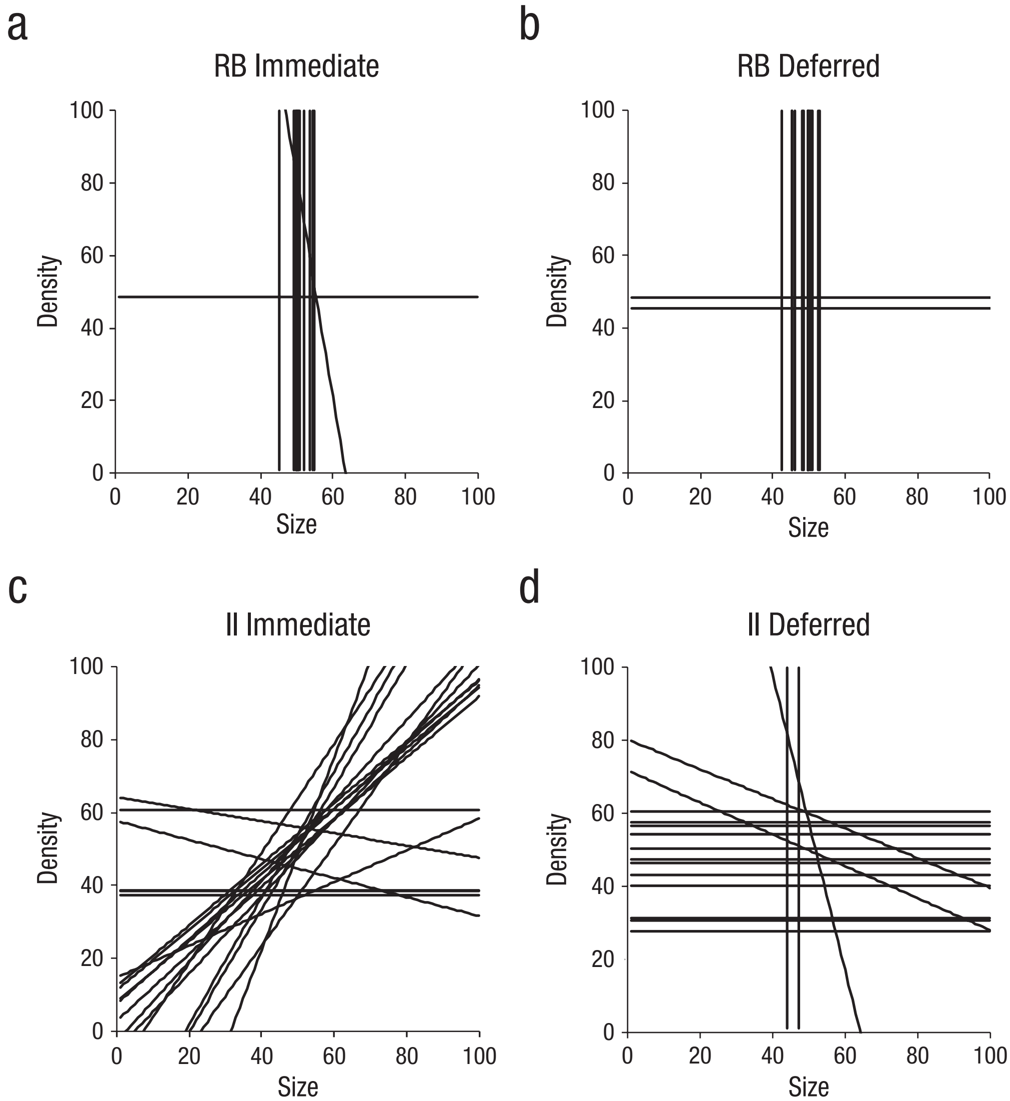

Fig. 4. The decision bounds that provided the best fits to the last 100 responses of participants in the (a) rule-based (RB) condition with immediate reinforcement, (b) RB condition with deferred reinforcement, (c) information-integration (II) condition with immediate reinforcement, and (d) II condition with deferred reinforcement.

is clear-cut. Many of these participants also found the one-dimensional size rule. The model-based and accuracy-based analyses converged to show that RB learning survived deferred reinforcement perfectly intact.

Figure 4c shows the modeling results for 19 II-immediate participants (modeling indicated that 2 participants in this condition were guessers). These decision bounds were generally organized along the stimulus space's major diagonal. Many participants found a way to integrate the informational signals provided by the two stimulus dimensions toward making appropriate category decisions.

Figure 4d shows the modeling results for 18 II-deferred participants (modeling indicated that 3 participants in this condition were guessers). The dissimilarity from Figure 4c is striking. II category learning failed disastrously in the face of deferred reinforcement. There were no decision bounds tracing the stimulus space's major diagonal. There was no appropriate information integration. Instead, participants defaulted to an RB strategy with decision bounds near the midline of Dimension *Y*. Possibly they defaulted to the only categorization process available under deferred reinforcement. They had to hold in mind a description of what they did over the whole trial block, so that the summary feedback—when it finally came—would be informative. As suggested by multiple-systems theory, this description took the form of a one-dimensional rule, not a principle for integrating information across dimensions. Figure 3b confirms these participants' process and point of rule discovery.

This result strengthens our interpretation of the accuracy-based analyses. The .64 accuracy achieved by II-deferred participants definitely does not signify 64% successful II learning. It signifies a heavy reliance on rules—that is, a reliance on the qualitatively wrong information-processing strategy for the II task. There was not 64% II learning in this condition. There was 0% II learning under deferred reinforcement.

General Discussion

Summary

The controversy over multiple category-learning utilities is reminiscent of the debate over multiple memory systems. Categorization researchers continue to seek more sharply distinguishing paradigms. We have contributed a new dissociative paradigm here by incorporating the technique of deferred reinforcement from comparative psychology.

We hypothesized that deferred reinforcement should disable associative learning and the II category learning that depends on it. Indeed, deferred reinforcement eliminated II category learning. There may be no comparably strong demonstration in the literature.

We also hypothesized that RB learners hold their category rule in working memory, still allowing its evaluation for adequacy at the end of the trial block when deferred reinforcement finally arrives. This hypothesis was also confirmed. RB learning was unscathed by deferred reinforcement. The demonstration that II learning is fully dependent on trial-by-trial reinforcement, but RB learning is fully independent, supports the multiple-systems view by confirming the operation of qualitatively different category-learning processes in different tasks.

Finally, we hypothesized that—with II category learning disabled by deferred reinforcement—participants might fall back, by necessity, to RB strategies. They would need to maintain in working memory a description of their blockwide strategy so they could interpret the summary feedback when it arrived. According to multiple-systems theory, this working description would be a one-dimensional rule and not an II principle. In fact, confirming multiple-systems theory in another way, II-deferred participants clearly adopted one-dimensional category rules that did not suit their task's II structure.

Dissociative frameworks in categorization

The confirmation of these three hypotheses constitutes one of the strongest RB-II dissociations yet seen. One cannot attribute this dissociation to differential difficulty. We have discussed how RB and II tasks are matched for every aspect of objective difficulty. Our tasks were learned equivalently under immediate reinforcement, which confirms that matching. And the result was not just that II learning worked more haltingly or with greater difficulty under deferred reinforcement. Instead, it did not engage at all—qualitatively—and the difficulty hypothesis cannot explain its complete absence. In addition, RB strategies showed a distinctive learning trajectory (Fig. 3a) that II strategies do not show. The difficulty hypothesis cannot explain that difference either.

The difficulty hypothesis also raises more general concerns. If one defines difficulty by any objective standard, our tasks were matched for difficulty. If one defines difficulty by humans' speed of learning and then uses difficulty to explain humans' speed of learning, the reasoning is circular. If one explains humans' speed of RB learning by an additional process or system that makes RB learning "easier" than II learning, one endorses multiple systems or processes. In the same way, one would not say that procedural learning—as compared with declarative memory—is preserved in amnesia because it is easier. There are more precise and theoretically illuminating things to say.

Nor can one claim that deferred reinforcement simply weakened the reinforcement signal by making it sporadic. In that case, if RB and II learning depended on that signal equivalently, they would have been impaired equivalently. They were not. The demonstration that one learning process is reliant on trial-by-trial reinforcement, and the other is not, supports the multiple-systems view. Moreover, no possible explanation based on the strength of the reinforcement signal can explain why II-deferred participants qualitatively shifted to RB processing. That shift is definitely not what the reinforcement signal was communicating, no matter its strength.

Another feature of the data disconfirms a single-system, exemplar-based explanation. The ability of exemplar models to capture learning trajectories in RB and II tasks has been debated (e.g., Ashby & Ell, 2002; Nosofsky & Kruschke, 2002). Exemplar models fit learning curves through gradual parameter changes. But the change shown in Figure 3a is qualitatively sudden, not gradual. Single-system models cannot fit this qualitative shift, or explain why there was no learning (i.e., no parameter changes) until Block 0, or why learning suddenly exploded at Block 0 (large parameter changes in one block). In contrast, all aspects of Figure 3a flow from assuming the sudden discovery of a category rule.

Our results do support the multiple-systems framework of categorization. Indeed, they instantiate perfectly the current theory of II learning. II category learning is presumably organized around a series of time-critical events that may surround the reinforcement-mediated strengthening of dopamine-related synapses (Ashby, Ennis, & Spiering, 2007). II learning cannot survive deferred reinforcement because the time-critical arrival of reinforcement is disrupted.

Our results also instantiate perfectly the current theory of RB learning. RB learning is presumably organized around rules actively held in working memory. RB learning survives deferred reinforcement because it does not depend on any temporal sequence of shaping or conditioning events. RB category knowledge is timeless in a sense because it is constantly available to consciousness. It can be flexibly applied and flexibly updated after a block of trials as well as after a single trial.

The present findings strengthen the functional MRI evidence, the neuropsychological evidence from patients, and results from other RB-II dissociative paradigms. Possibly the present data will bring the debate about RB, II, and multiple category-learning systems nearer to a consensual and collegial conclusion that would further the categorization literature's theoretical development.

Adaptive complementarity in categorization

The multiple-systems framework in categorization points to an elegant division of cognitive labor that is insufficiently appreciated. Through II learning, cognition creates stimulus-response bonds in a sense. Consequences (reinforcements) are the glue for associating adaptive behaviors to stimuli. This system has considerable strengths. It produces stable behavior. It produces the behavior with the highest probability of reinforcement. It slowly, conservatively commits to behavioral solutions. It slowly, conservatively lets behavioral solutions go (through the class of extinction phenomena). It operates preattentionally, out of awareness, which potentially grants it great phylogenetic breadth.

But this system has constraints. It depends on immediate reinforcement, time-critical event sequences, and persistent event repetition. Learning cannot occur off-line or with displacement in time or space from the task's trials. New approaches cannot be chosen instantly at need. Old approaches cannot be replaced instantly at need.

RB learning is a perfect complement to II learning. It is not rigidly time locked. It does not depend on immediate reinforcement or event repetition. Learning can occur off-line and with displacement. Learning and unlearning can occur suddenly at need.

The phylogenetic depth of adaptive complementarity

The adaptiveness of these complementary categorization utilities raises a question concerning their phylogenetic depth that recent research has addressed. Smith, Beran, Crossley, Boomer, and Ashby (2010) found that rhesus macaques (*Macaca mulatta*), like humans, sometimes learn RB tasks much faster than II tasks. Thus, nonhuman primates have some structural components of humans' capacity for explicit categorization and glean some of the benefits of the RB category-learning system, though they may not have all the components of humans' explicit system (e.g., full declarative awareness). Smith et al. (2012) generalized this finding to another large primate group, the New World monkeys.

In contrast, Smith et al. (2011) found that pigeons learned RB and II tasks to the same level at the same speed. In pigeons, the cognitive system may not be strongly committed to dimensional analysis and category rules. Pigeons may lack the explicit-implicit complementarity in categorization that primates possess. Pigeons' performance could shed light on the ancestral vertebrate categorization system from which the categorization system of primates and humans emerged.

Conclusion

The dissociative framework describing explicit and implicit systems of categorization continues to illuminate and enrich the cognitive and comparative literatures on categorization. It guides productive empirical research, generates testable predictions, and expresses important adaptive complementarities among the categorization utilities possessed by humans and some nonhuman species.

Author Contributions

J. D. Smith conceived the project and was the article's principal author. F. G. Ashby contributed critical revisions and theoretical insights framing the article. B. A. Church supervised data collection, analysis, and interpretation and contributed critical revisions. J. L. Roeder conducted the article's formal modeling. J. Boomer and A. C. Zakrzewski conducted data collection and statistical analyses and contributed to writing the manuscript. All authors approved the final version of the manuscript for submission.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

The preparation of this article was supported by Grant HD-060563 from the National Institute of Child Health and Human Development (NICHD), by Grant HD-061455 from NICHD, by Grant BCS-0956993 from the National Science Foundation, by Grant P01 NS044393 from the National Institute of Neurological Disorders and Stroke, and by the U.S. Army

Research Office through the Institute for Collaborative Biotechnologies under Grant W911NF-07-1-0072.

Supplemental Material

Additional supporting information may be found at http://pss.sagepub.com/content/by/supplemental-data

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*, 442–481.

Ashby, F. G., & Ell, S. W. (2001). The neurobiology of human category learning. *Trends in Cognitive Sciences*, *5*, 204–210.

Ashby, F. G., & Ell, S. W. (2002). Single versus multiple systems of category learning: Reply to Nosofsky and Kruschke (2001). *Psychonomic Bulletin & Review*, *9*, 175–180.

Ashby, F. G., Ennis, J. M., & Spiering, B. J. (2007). A neurobiological theory of automaticity in perceptual categorization. *Psychological Review*, *114*, 632–656.

Ashby, F. G., & Maddox, W. T. (2010). Human category learning 2.0. *Annals of the New York Academy of Sciences*, *1224*, 147–161.

Ashby, F. G., Maddox, W. T., & Bohil, C. J. (2002). Observational versus feedback training in rule-based and information-integration category learning. *Memory & Cognition*, *30*, 666–677.

Ashby, F. G., & O'Brien, J. B. (2005). Category learning and multiple memory systems. *Trends in Cognitive Sciences*, *2*, 83–89.

Ashby, F. G., Queller, S., & Berretty, P. M. (1999). On the dominance of unidimensional rules in unsupervised categorization. *Perception & Psychophysics*, *61*, 1178–1199.

Ashby, F. G., & Valentin, V. V. (2005). Multiple systems of perceptual category learning: Theory and cognitive tests. In H. Cohen & C. Lefebvre (Eds.), *Handbook of categorization in cognitive science* (pp. 547–572). New York, NY: Elsevier.

Blair, M., & Homa, D. (2003). As easy to memorize as they are to classify: The 5-4 categories and the category advantage. *Memory & Cognition*, *31*, 1293–1301.

Brooks, L. R. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 169–211). Hillsdale, NJ: Erlbaum.

Cook, R. G., & Smith, J. D. (2006). Stages of abstraction and exemplar memorization in pigeons' category learning. *Psychological Science*, *17*, 1059–1067.

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

Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature*, *407*, 630–633.

Fukunaga, K. (1972). *Introduction to statistical pattern recognition*. New York, NY: Academic Press.

Homa, D., Sterling, S., & Trepel, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory*, *7*, 418–439.

Knowlton, B. J., & Squire, L. R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science*, *262*, 1747–1749.

Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*, *53*, 49–70.

Maddox, W. T., & Ashby, F. G. (2004). Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behavioural Processes*, *66*, 309–332.

Maddox, W. T., Ashby, F. G., & Bohil, C. J. (2003). Delayed feedback effects on rule-based and information-integration category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 650–662.

Maddox, W. T., & Ing, A. D. (2005). Delayed feedback disrupts the procedural-learning system but not the hypothesis-testing system in perceptual category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 100–107.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207–238.

Minda, J. P., & Smith, J. D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 775–799.

Murphy, G. L. (2003). *The big book of concepts*. Cambridge, MA: MIT Press.

Newell, B. R., Dunn, J. C., & Kalish, M. (2010). The dimensionality of perceptual category learning: A state-trace analysis. *Memory & Cognition*, *38*, 563–581.

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 87–108.

Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar-based accounts of multiple-system phenomena in perceptual categorization. *Psychonomic Bulletin & Review*, *7*, 375–402.

Nosofsky, R. M., & Kruschke, J. K. (2002). Single-system models and interference in category learning: Commentary on Waldron and Ashby (2001). *Psychonomic Bulletin & Review*, *9*, 169–174.

Nosofsky, R. M., Stanton, R. D., & Zaki, S. R. (2005). Procedural interference in perceptual classification: Implicit learning or cognitive complexity? *Memory & Cognition*, *33*, 1256–1271.

Rosseel, Y. (2002). Mixture models of categorization. *Journal of Mathematical Psychology*, *46*, 178–210.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464.

Smith, J. D., Ashby, F. G., Berg, M. E., Murphy, M. S., Spiering, B., Cook, R. G., & Grace, R. C. (2011). Pigeons' categorization may be exclusively nonanalytic. *Psychonomic Bulletin & Review*, *18*, 422–428.

Smith, J. D., Beran, M. J., Crossley, M. J., Boomer, J., & Ashby, F. G. (2010). Implicit and explicit category learning by macaques (*Macaca mulatta*) and humans (*Homo sapiens*). *Journal of Experimental Psychology: Animal Behavior Processes*, *36*, 54–65.

Smith, J. D., Beran, M. J., Redford, J. S., & Washburn, D. A. (2006). Dissociating uncertainty states and reinforcement signals in the comparative study of metacognition. *Journal of Experimental Psychology: General*, *135*, 282–297.

Smith, J. D., Chapman, W. P., & Redford, J. S. (2010). Stages of category learning in monkeys (*Macaca mulatta*) and humans (*Homo sapiens*). *Journal of Experimental Psychology: Animal Behavior Processes*, *36*, 39–53.

Smith, J. D., Crossley, M. J., Boomer, J., Church, B. A., Beran, M. J., & Ashby, F. G. (2012). Implicit and explicit category learning by capuchin monkeys (*Cebus apella*). *Journal of Comparative Psychology*, *126*, 294–304.

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1411–1436.

Smith, J. D., Redford, J. S., & Haas, S. M. (2008). Prototype abstraction by monkeys (*Macaca mulatta*). *Journal of Experimental Psychology: General*, *137*, 390–401.

Waldron, E. M., & Ashby, F. G. (2001). The effects of concurrent task interference on category learning: Evidence for multiple category learning systems. *Psychonomic Bulletin & Review*, *8*, 168–176.

Wasserman, E. A., Kiedinger, R. E., & Bhatt, R. S. (1988). Conceptual behavior in pigeons: Categories, subcategories, and pseudocategories. *Journal of Experimental Psychology: Animal Behavior Processes*, *14*, 235–246.