frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 07 July 2015
doi: 10.3389/fnhum.2015.00389

Dissociation of category-learning systems via brain potentials

Robert G. Morrison[1], Paul J. Reber[2], Krishna L. Bharani[3] and Ken A. Paller[2]*

[1] Department of Psychology, Neuroscience Institute, Loyola University Chicago, Chicago, IL, USA, [2] Department of Psychology, Northwestern University, Evanston, IL, USA, [3] Department of Neuroscience, Medical University of South Carolina, Charleston, SC, USA

Behavioral, neuropsychological, and neuroimaging evidence has suggested that categories can often be learned via either an explicit rule-based (RB) mechanism critically dependent on medial temporal and prefrontal brain regions, or via an implicit information-integration (II) mechanism relying on the basal ganglia. In this study, participants viewed sine-wave gratings (Gabor patches) that varied on two dimensions and learned to categorize them via trial-by-trial feedback. Two different stimulus distributions were used; one was intended to encourage an explicit RB process and the other an implicit II process. We monitored brain activity with scalp electroencephalography (EEG) while each participant: (1) passively observed stimuli represented of both distributions; (2) categorized stimuli from one distribution, and, 1 week later; (3) categorized stimuli from the other distribution. Categorization accuracy was similar for the two distributions. Subtractions of Event-Related Potentials (ERPs) for correct and incorrect trials were used to identify neural differences in RB and II categorization processes. We identified an occipital brain potential that was differentially modulated by categorization condition accuracy at an early latency (150–250 ms), likely reflecting the degree of holistic processing. A stimulus-locked Late Positive Complex (LPC) associated with explicit memory updating was modulated by accuracy in the RB, but not the II task. Likewise, a feedback-locked P300 ERP associated with expectancy was correlated with performance only in the RB, but not the II condition. These results provide additional evidence for distinct brain mechanisms supporting RB vs. implicit II category learning and use.

Keywords: category learning, event-related potentials, explicit, implicit, EEG

OPEN ACCESS

Edited by:
Lynne E. Bernstein, George Washington University, USA

Reviewed by:
Carol Seger, Colorado State University, USA
Todd Maddox, University of Texas Austin, USA

****Correspondence:***
Robert G. Morrison, Department of Psychology, Neuroscience Institute, Loyola University Chicago, 1032 W Sheridan Road, Coffey Hall, Chicago, IL 60660, USA
rmorrison@luc.edu

Received: *05 April 2015*
Accepted: *19 June 2015*
Published: *07 July 2015*

Citation:
Morrison RG, Reber PJ, Bharani KL and Paller KA (2015) Dissociation of category-learning systems via brain potentials. Front. Hum. Neurosci. 9:389. doi: 10.3389/fnhum.2015.00389

Introduction

Categories, as conceptualized based on perceived regularities, allow us to make sense of, describe, and order our worlds (Rips et al., 2012). Categories come in many different forms—from categories based on a single feature (e.g., objects that are red) to much more complicated relational concepts (e.g., *chases* or *conduit*). Many have argued that human categorization is not a unitary process, but rather can engage different systems depending on the category structure or the conditions during category learning (e.g., Yamauchi and Markman, 1998; Nomura and Reber, 2008; Smith and Grossman, 2008; Seger and Miller, 2010; Ashby and Maddox, 2011). Behavioral, neuropsychological, and neuroimaging evidence suggests that these various systems can make differential demands on neural networks of

the brain (e.g., Kéri, 2003; Nomura and Reber, 2008; Smith and Grossman, 2008; Seger and Miller, 2010; Ashby and Maddox, 2011). However, describing the algorithm and neural implementation of category-learning systems, as well as the factors that determine when each system will be engaged and how these systems interact, is still a very active endeavor.

A prominent way to characterize category-learning systems postulates distinct rule-based (RB) and information-integration (II) categorization strategies that engage different neurocognitive networks (see Ashby and Maddox, 2011). Within this framework, Maddox et al. (2003) have developed a feedback category-learning paradigm which parametrically varies the perceptual properties of sine-wave gratings (Gabor patches) to create category distributions that encourage either RB or II category learning strategies (see **Figure 1**).

RB tasks are those where the categories can be learned via a reasoning process such as hypothesis testing (Ashby et al., 1998, 2005). In contrast, II category learning and use appears to occur implicitly, such that the rule for the category structure is difficult to learn consciously or to describe verbally. After learning, participants can explicitly describe the rule they use to categorize the stimuli. This RB mechanism would require maintaining and updating the rule and the boundary condition, requiring the use of both working memory, dependent on prefrontal cortex (PFC), and long-term memory, dependent on medial temporal lobe (MTL; Nomura and Reber, 2012).

In contrast, II learning appears to occur implicitly, such that the rule for the category structure is difficult to learn consciously or to describe verbally. II tasks appear to encourage participants to consider the stimuli holistically, integrating perceptual information from different stimulus features early during processing. II learning may depend on implicit learning supported by computations involving the caudate nucleus and visual processing areas in occipital cortex (Nomura and Reber, 2012). Dopaminergic reward circuits of the caudate may be responsible for associating specific categories with neuronal patterns in occipital cortex that code for relevant visual features (Ashby et al., 1998).

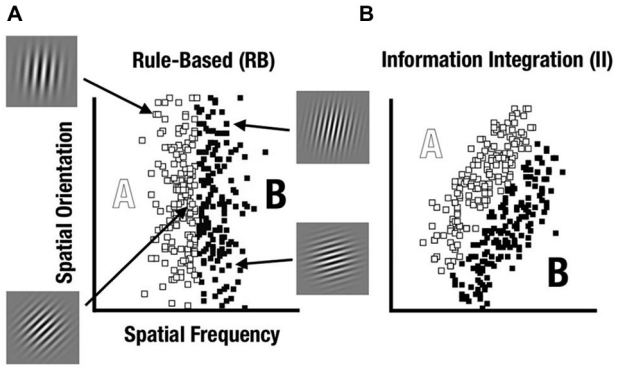

FIGURE 1 | Rule-based (RB) and information-integration (II) category distributions used in the study. Sine-wave gratings varied based on spatial frequency and spatial orientation. **(A)** The RB category was defined based on frequency whereas orientation varied unsystematically. **(B)** The II category was defined based on both frequency and orientation with a diagonal decision bound.

Numerous behavioral experiments comparing RB and II category learning have shown that they are employed using dissociable strategies. For example, working memory dual-task procedures interfered with RB much more than with II learning (e.g., Zeithamova and Maddox, 2006, 2007). Delaying feedback beyond an initial period did not interfere with RB learning but disrupted II learning (e.g., Maddox et al., 2003). Changing the response key associated with a particular category also interfered with II but not RB categorization, suggesting that II learning may require stimulus-response association learning with relatively immediate feedback, characteristics associated with implicit procedural learning (Ashby et al., 2003).

Mechanistically RB processing is thought to depend on hypothesis testing. For instance a participant trying to categorize line segments into two groups might hypothesize that length is what matters, with long segments being one category and short segments being the other. On each trial they test their theory with a response to each line segment. While they may find support for their theory quickly they gradually build a representation of the category threshold that allows them to improve their performance. After each test of their hypothesis they then need to update their memory with whether the test worked and with a candidate threshold value. This evaluation requires selective attention and working memory, likely implemented in PFC, as well as the ability to form enduring mental representations of the rule and boundary condition dependent on the hippocampus and MTL. In contrast, II learning is believed to require information integration of multiple stimulus attributes at a predecisional stage (Ashby et al., 1998). Unlike in RB learning, learners frequently cannot articulate what they have learned, but can show their learning through successful performance, a hallmark of nondeclarative memory (Squire, 2009). Thus, II learning may be likened to gaining category expertize with complex objects such as faces (Bentin et al., 1996) or Greebles (Rossion et al., 2002).

Working from this distinction, functional magnetic resonance imaging (fMRI) methods have been useful to spatially dissociate the brain networks responsible for categorization and use when participants learn either an RB or II category distribution. In a study by Nomura et al. (2007b), participants who learned the RB distribution showed greater activation in the MTL on correct than incorrect trials, while participants who learned the II distribution showed greater activation in the body of the caudate on correct than incorrect trials. Another category learning study using a different paradigm likewise found activity in the body and tail of the caudate and putamen to be active when learning stimulus-category associations (Cincotta and Seger, 2007). Nomura and Reber (2012) subsequently reanalyzed several sets of RB/II paradigm fMRI data (Nomura et al., 2007a) using PINNACLE (Parallel Interactive Neural Networks Active in Category Learning), a computational model that includes multiple competing categorization systems. Using a participant's behavioral decision data, PINNACLE employs principals of Decision-Bound Modeling Theory (Ashby and Maddox, 1993) to estimate which categorization system is likely engaged on a given trial. Thus, PINNACLE can be used to sort trials of

neuroimaging data to obtain estimates of the neural correlates for individual category-learning systems. This approach identified areas in PFC important for correct decisions during RB category learning, a finding consistent with another previous fMRI study of RB category learning (Filoteo et al., 2005). Posterior regions of occipital cortex were associated with correct decisions during II category learning, a finding consistent with previous fMRI studies of implicit category learning (Reber et al., 1998a,b; Waldschmidt and Ashby, 2011). In addition, this approach found evidence that regions of dorsolateral PFC were involved in the process of resolving competition between the two systems based on the model-identified moments of high levels of inter-system competition.

Further progress in understanding the neurocognitive mechanisms of category learning will depend on the ability to measure relevant processing. In particular, measures with high temporal resolution are needed to comprehensively distinguish RB and II mechanisms. In the present study we computed event-related potentials (ERPs) from scalp electroencephalographic (EEG) recordings to examine neural correlates of category learning during both categorization and feedback stages. Participants learned RB and II category distributions during separate testing sessions and their responses were analyzed using Decision-Bound Modeling Theory (Ashby and Maddox, 1993) to identify participants likely to be using RB and II category learning processes with corresponding distributions. Based on prior behavioral and neuroimaging results, we anticipated that RB and II category learning mechanisms would produce different ERPs, when comparing successful (correct) and unsuccessful (incorrect) trials. Specifically, we anticipated differences in an early occipital N1 ERP previously associated with visual category learning (Curran et al., 2002), and consistent with occipital activation found for II category learning in our previous work (Nomura and Reber, 2012). Secondly, given the previously demonstrated reliance of RB category learning on MTL (Seger and Cincotta, 2006; Nomura et al., 2007a; Seger et al., 2011) we predicted that a Late Positive Complex (LPC) ERP associated with explicit memory (Voss and Paller, 2008) would be modulated by accuracy in the RB condition but not the II condition. Lastly, to the extent that RB learning is more explicit than II learning (Huang-Pollock et al., 2011; Seger et al., 2011), we anticipated that the P300 to positive feedback would index participant's confidence in their learning (Hajcak et al., 2005).

Materials and Methods

Task Description

We used a visual category-learning paradigm (Maddox et al., 2003) in which subjects learned to categorize visual stimuli into two categories via feedback given at the conclusion of each trial. Stimuli were circular sine-wave gratings that varied in spatial frequency (number of lines per patch, also perceived as thickness of lines) and spatial orientation (tilt of lines). For the RB distribution, the stimuli were divided into two categories based on a vertical decision boundary such that category membership depended only on the spatial frequency of the sine-wave grating (**Figure 1A**). For the II group, the categories were defined by a diagonal decision boundary that required II of frequency and orientation information (**Figure 1B**). Trial timing was similar to that used by Nomura et al. (2007a) in their fMRI study (**Figure 2**).

Participants

Twenty-eight Northwestern University students served as participants in this experiment. Participants received US$15 per hour for two 2 to 3 hr testing sessions. Participants categorized the RB and II category distributions in separate sessions 1 week apart. Distribution order was counterbalanced across participants. Participants gave informed consent according to the oversight of the Northwestern University Institutional Review Board.

Procedure

Prelearning

In order to rule out differences in ERPs due to differences in the physical stimuli in the RB and II distributions, participants passively viewed 160 sine-wave gratings from both distributions over the course of two blocks prior to attempting to learn categories. Gratings were representative of the range of spatial frequency and orientation used during category learning. During prelearning participants received no instruction that categories of stimuli were present or that they should categorize. Prelearning trial timing was identical to that during category learning, but participants did not make a response during prelearning and thus received no feedback.

Category Learning

Participants categorized 320 sine-wave gratings presented in four blocks during each category-learning session. One session involved the RB distribution and the other session involved the II distribution. Distribution order was counterbalanced across participants. Prior to testing, subjects were familiarized with the procedures, including trial timing, button pressing, and feedback. Participants did not receive instructions about the nature of the categories; rather, they were asked to discover the categories with the aid of auditory feedback. Participants received auditory feedback 2.5 s after stimulus onset. For a correct decision the feedback was a bell sound. For incorrect decisions the feedback was a short buzzer, while participants heard a long buzzer of equal duration when no response was made in the allotted 2 s. Responses after 2 s were not considered in the analysis. Subjects were debriefed about their categorization strategies after the second testing session.

EEG

Continuous EEG recordings were made during prelearning and category-learning blocks from 59 evenly distributed scalp sites using tin electrodes embedded in an elastic cap (**Figure 3**). Four additional electrodes were used for monitoring horizontal and vertical eye movements, and two electrodes were placed over the left and right mastoid bones. Participants were instructed to attempt to refrain from blinking or moving their eye position

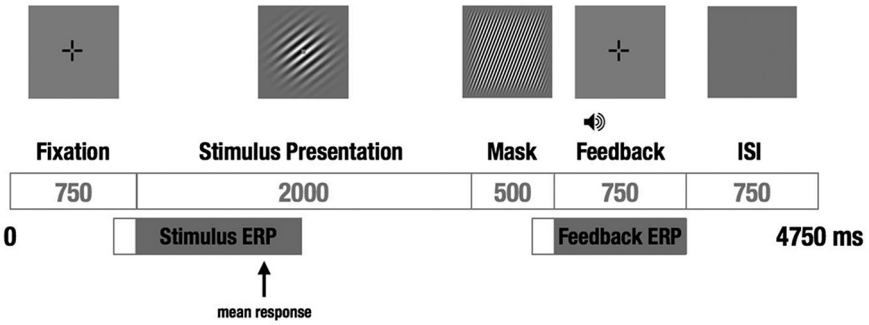

FIGURE 2 | Schematic of a single trial. A fixation cross was followed by the to-be-categorized-stimulus for a fixed duration, followed by a short visual mask, followed by auditory feedback and a brief ISI before the next trial. The subject responded "category A" or "category B" during the 2 s the stimulus was on the screen by pressing one of two buttons on a hand-held response box. EEG was recorded continuously, and stimulus- and feedback-locked ERPs were calculated from each trial.

from fixation during the categorization and feedback portions of each trial. Electrode impedance was $\leq$5 kΩ. EEG signals were amplified with a band pass of 0.05–200 Hz and sampled at a rate of 1000 Hz. The online reference (left mastoid) was changed to average mastoids offline and a 59 to 60 Hz band-stop filter was applied. EMSE Software Suite (Source Signal Imaging, San Diego, CA, USA) was used to process raw EEG files and to compute ERPs. Electrooculograph (EOG) artifacts were corrected by using a blink-correction algorithm based on independent component analysis. Averaging epochs for stimulus and feedback lasted 1200 ms, including a 200 ms pre-stimulus baseline. Trials showing a greater than 100 μV deflection during the epoch were discarded. Fewer than 15% of trials were excluded for any given condition for any given participant.

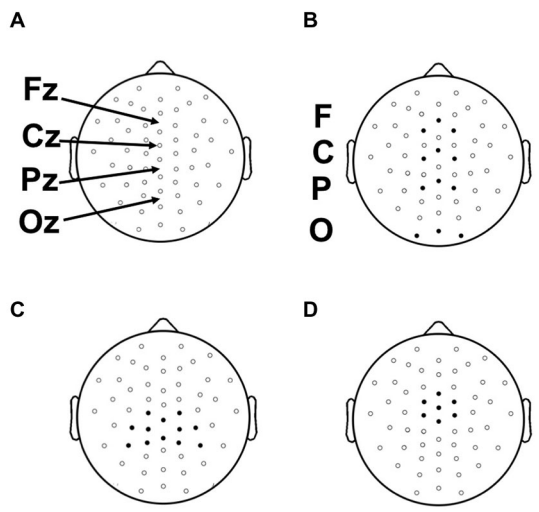

FIGURE 3 | (A) Placement of 52 scalp electrodes (of 58 total) used in this study with respect to 10–20 landmarks. **(B)** Frontal (F), Central (C), Parietal (P), and Occipital (O) electrode clusters used for the analysis of the N1. **(C)** Parietal electrode cluster surrounding location Pz used in the analysis of the Late positive complex (LPC). **(D)** Central electrode cluster surrounding and including electrode Cz used in the analysis of the feedback P300.

Decision-Bound Theory Modeling

Although participants received stimuli drawn from either the RB distribution or from the II distribution within each session, some participants would be expected to fail to adopt the optimal categorization strategy. As in prior work (Ashby and Maddox, 1993; Nomura and Reber, 2012), we used Decision-Bound Theory (DBT) models to classify behavioral patterns as consistent with either an RB strategy or II strategy. For each participant, the pattern of categorization responses across the stimulus space was compared to an RB-F model based on stimulus spatial frequency (thinness of the black/white strips reflected as a vertical boundary in stimulus space), an RB-O model based on spatial orientation (angle of the black/white strips reflected as a horizontal boundary in stimulus space) and an II model based on a diagonal partition of the stimulus space. The specific placement of the category boundary was optimized to the participant's behavior and the quality of the fit was contrasted across models. By this method, performance in each session can be identified as consistent with either an RB or II approach that either is relatively optimal for the administered stimulus set or reflects a suboptimal strategy. We fit each block of 80 trials using the DBT model. Participants whose performance was consistent with task demands (i.e., at least three of four blocks showed model-to-distribution agreement) were considered the Model-Conforming group and the remaining participants were designated as the Model-Nonconforming group. Using this technique to identify participants most clearly expressing the appropriate strategy strengthens the comparison of ERP differences between RB and II category learning.

Results

All 28 participants exhibited an RB distribution response best fit by an RB-F DBT model. For II, only 15 participants comprised the Model-Conforming Group because they exhibited an II distribution response profile best fit by an II DBT model. In contrast, 13 participants comprised the Model-Nonconforming

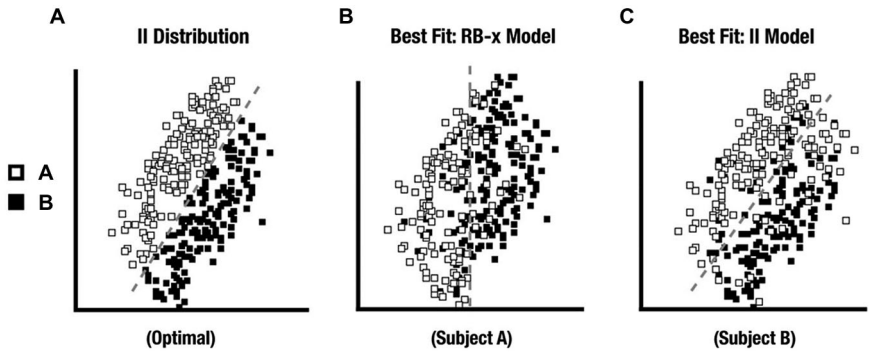

FIGURE 4 | (A) II distribution used in the experiment. **(B)** II distribution category responses from a participant whose responses were best fit by a RB decision-bound theory (DBT) model and who was excluded from further analysis. **(C)** II distribution category responses from a participant whose responses were best fit by an II DBT model and who was kept for further analysis.

Group because they exhibited an II distribution response profile best fit by an RB-F or RB-O DBT model (see **Figure 4** for distribution profiles from representative participants). Likewise, when the fits for these two groups were compared directly, the first group of participants exhibited better II model fits than did the second (t (26) = 2.7, p = 0.01). However, these two groups did not differ in the quality of their RB model fits with the RD distribution (t (26) = 0.02, *ns*). DBT model fitting thus allowed data from participants who were likely using a unidimensional RB strategy with the II category distributions to be excluded from subsequent analyses.

Behavioral Performance

Of the 15 participants whose DBT fits were consistent with II strategy use with II distributions, two did not have an adequate number of incorrect trials (<30) to allow for the correct/incorrect ERP analysis, so their results were excluded from further analysis. Data from one additional participant were eliminated because of poor EEG quality.

To evaluate potential differences in category-learning accuracy for the RB and II distributions, we ran a 2 (RB vs. II) by 4 (block) repeated-measures ANOVA. Accuracy for RB and II distributions (**Figure 5A**) did not reliably differ ($F_{(1,11)}$ = 1.6, p = 0.23, η_p^2 = 0.13). There was a main effect of block ($F_{(3,33)}$ = 24, $p < 0.001$, η_p^2 = 0.69), and category learning linearly increased over blocks ($F_{(1,11)}$ = 50, $p < 0.001$, η_p^2 = 0.81). However, RB and II distributions did not differ with respect to this pattern ($F_{(1,11)}$ = 0.4, p = 0.5, η_p^2 = 0.04). Thus, observed differences in correct/incorrect ERP subtractions (described below) cannot easily be attributed to differences in accuracy between RB and II learning.

Next we looked for potential differences in category-learning RT for the RB and II distributions by using a 2 (RB vs. II) by 2 (Correct vs. Incorrect) by 4 (block) repeated measures ANOVA (see **Figure 5B**). Participants were faster on correct than incorrect trials ($F_{(1,11)}$ = 27, $p < 0.001$, η_p^2 = 0.71). There was also a trend towards faster responses on RB trials compared to II trials ($F_{(1,11)}$ = 4.0, p = 0.07, η_p^2 = 0.27). Likewise, there was a trend suggesting an interaction between accuracy and distribution type ($F_{(1,11)}$ = 2.6, p = 0.14, η_p^2 = 0.19). Participants were faster on correct trials than on incorrect trials for both RB distributions ($F_{(1,11)}$ = 20, $p < 0.001$, η_p^2 = 0.65) and II distributions ($F_{(1,11)}$ = 14, p = 0.003, η_p^2 = 0.56). However, RB and II trials only differed for correct trials ($F_{(1,11)}$ = 6.6, p = 0.026, η_p^2 = 0.38) not incorrect trials ($F_{(1,11)}$ = 1.1, p = 0.31, η_p^2 = 0.09).

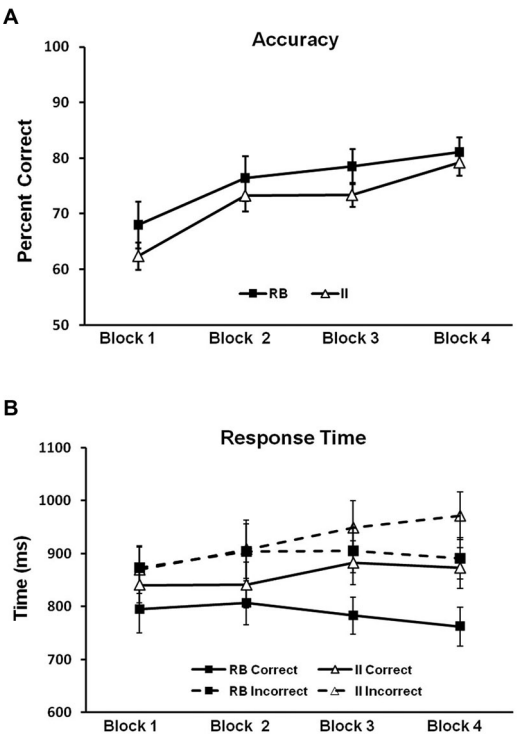

FIGURE 5 | Behavioral results for Model-Conforming Group. (A) Accuracy and **(B)** RTs for participants included based on DBT model fits and included in the analysis of brain potentials. Error bars represent ±1 standard error of the mean.

EEG Results

Categorization ERPs

Based on our predictions, stimulus-locked analyses were focused on an early occipital N1 ERP (**Figure 6**) and a later parietal LPC ERP (**Figure 7**) in the Model-Conforming Group.

To measure occipital N1 ERPs, we calculated mean amplitude from 150–250 ms for a cluster of inferior occipital electrodes (**Figure 6**). The same electrodes and time range were used for every participant. This time range included the occipital N1 peak for all participants. A 2 (RB vs. II) by 2 (Correct vs. Incorrect) ANOVA performed on mean amplitudes yielded a reliable interaction between distribution type and accuracy ($F_{(1,11)} = 6.1$, $p = 0.03$, $\eta_p^2 = 0.36$), but no main effect of distribution type ($F_{(1,11)} = 0.05$, $p = 0.8$, $\eta_p^2 = 0.004$) or accuracy ($F_{(1,11)} = 0.04$, $p = 0.9$, $\eta_p^2 = 0.003$). Amplitudes at this latency for correct and incorrect trials were reliably different for the II distribution ($F_{(1,11)} = 6.3$, $p = 0.03$, $\eta_p^2 = 0.37$) and showed a trend in the opposite direction for the RB distribution ($F_{(1,11)} = 2.6$, $p = 0.14$, $\eta_p^2 = 0.19$).

Also consistent with predictions, we found a stimulus-locked LPC ERP largest over the parietal electrodes (**Figure 7A,D**). To quantify LPC, we measured mean amplitude from 400–700 ms in a cluster of parietal electrodes (see **Figure 3C**). A 2 (RB vs. II) by 2 (Correct vs. Incorrect) ANOVA performed on mean amplitudes yielded a reliable interaction between distribution type and accuracy ($F_{(1,11)} = 9.6$, $p = 0.01$, $\eta_p^2 = 0.47$). The LPC was reliably larger for correct than incorrect trials in the RB condition ($F_{(1,11)} = 20$, $p = 0.001$, $\eta_p^2 = 0.65$), but not in the II condition ($F_{(1,11)} = 3.2$, $p = 0.1$, $\eta_p^2 = 0.23$). To uncover relationships between this ERP and performance (**Figure 7B,D**), we used a smaller parietal region and temporal window (500–600 ms) targeted for maximal mean amplitude differences as a function of accuracy. Magnitude of the Correct/Incorrect ERP differences were reliably correlated with RB performance (**Figure 7C**; $r_{(11)} = 0.68$, $p = 0.01$) but not with II performance (**Figure 7F**; $r_{(11)} = 0.05$, $p = 0.9$).

Feedback ERPs

In order to assess hypotheses about the extent to which categorization was based on explicit knowledge, we examined ERPs recorded during feedback (**Figure 8**). Participants interpret feedback signals as a function of their explicit expectations. P300 responses have been associated with confidence in learning with feedback (Hajcak et al., 2005). Accordingly, we expected P300 potentials to index learning in the RB but not in the II condition, given that explicit learning mechanisms are thought to dominate in the RB but not the II condition. Both Correct and Incorrect trials showed large positive potentials at approximately 300 ms with central-focused topographies (**Figures 8A,B,D**). A 2 (RB vs. II) by 2 (Correct vs. Incorrect) ANOVA was performed on post-feedback mean amplitudes at 200–400 ms from a cluster of seven central electrodes (**Figure 3D**). The analysis yielded a main effect of accuracy ($F_{(1,11)} = 43$, $p < 0.001$, $\eta_p^2 = 0.78$), but no effect of distribution type ($F_{(1,11)} = 0$, $p = 0.99$, $\eta_p^2 = 0$), and no interaction between distribution type and accuracy ($F_{(1,11)} = 0.25$, $p = 0.6$, $\eta_p^2 = 0.02$).

However, because the P300 is frequently associated with expectancy violations (Polich, 2007) and is larger when participants receive unexpected feedback (Hajcak et al., 2005), we hypothesized that participants who were better at RB categorization would show lower P300 response to correct feedback than would participants who had less-developed rules. To test this idea, we correlated categorization accuracy with P300 amplitude to correct feedback signals. Confirming our hypothesis, we found that accuracy was inversely correlated with P300 amplitude for the RB distribution (**Figure 8C**; $r_{(11)} = -0.71$, $p = 0.01$), but not for the II distribution (**Figure 8F**; $r_{(11)} = 0.07$, $p = 0.83$).

Because the stimulus-locked LPC during categorization and the feedback-locked P300 both appear to index effective learning in the RB condition, but not in the II

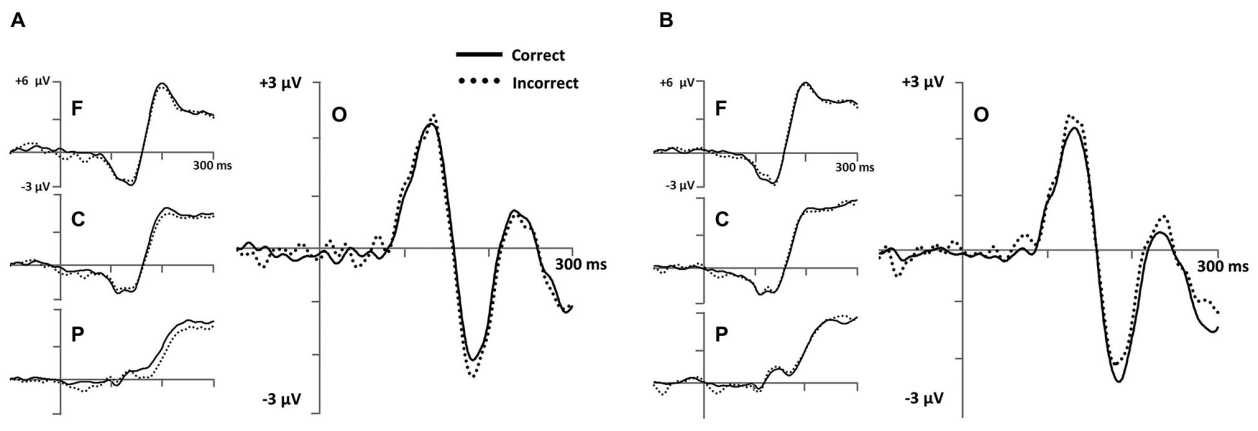

FIGURE 6 | Early stimulus-locked ERPs from frontal (F; three marked electrodes just posterior to Fz), central (C; three marked electrodes including Cz), parietal (P; three marked electrodes just posterior to Pz) and occipital (three marked electrodes just posterior to Oz including the Iniun) electrode clusters (see Figure 3B for precise electrode locations) for (A) RB and (B) II category-learning conditions.

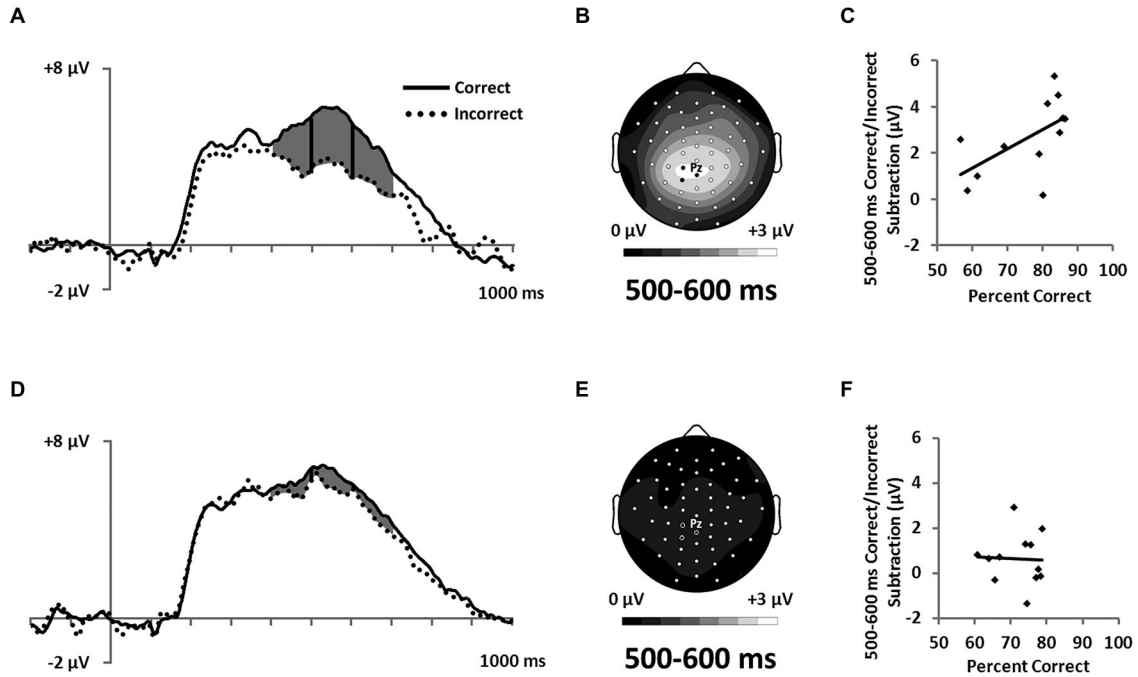

FIGURE 7 | ERPs showing the LPC ERP for both (A) RB and (D) II conditions in a cluster of parietal electrodes (12 marked electrodes surrounding Pz; see Figure 3C for precise electrode locations). Topographic maps representing correct minus incorrect subtractions from 500–600 ms for **(B)** RB and **(E)** II ERPs. Scatterplots showing the relationship of accuracy to the correct minus incorrect mean amplitude ERP subtractions from 500–600 ms for three parietal electrodes near Pz (indicated on the corresponding topographic maps) for **(C)** RB and **(F)** II conditions.

condition, we looked to see whether they were related across participants. The LPC correct/incorrect subtraction is negatively correlated with the feedback P300 correct/incorrect subtraction in the RB condition ($r = -0.59$, $p = 0.03$), but not in the II condition ($r = -0.08$, $p = 0.82$). We believe the negative correlation observed in RB trials indicates that better performers are generally more confident in their learning. Thus, they tend to update their memory more on correct than incorrect trials (larger Correct/Incorrect LPC difference) and also tend to be less surprised when they receive positive feedback, (smaller Correct/Incorrect Feedback P300 difference). This correlation is dramatically absent in II learning suggesting that even good II learners do not have explicit awareness of their learning.

Prelearning ERPs

Our critical comparisons during category learning were between correct and incorrect trials within either RB or II distributions, not across the two distributions. Yet, we took steps to ensure that differences were not due to the nature of the stimuli in the RB vs. II distributions. Accordingly, we analyzed ERPs from prelearning at the same latencies and scalp locations used in the categorization analyses for N1 and LPC. Neither N1 ($t_{(10)} = 1.0$, $p = 0.34$) nor LPC ($t_{(10)} = 0.11$, $p = 0.91$) differed between the two distributions, confirming that effects can be ascribed to learning rather than physical stimulus differences.

Discussion

ERP measures differentiated RB and II category-learning processes from each other. During categorization, differences in neural activity were observed in an early, occipital N1 ERP component in the form of differential correct/incorrect activity patterns for RB and II conditions (**Figure 6**). N1 amplitudes in the II condition were more negative for correct than for incorrect trials, while a trend toward the opposite pattern was observed in the RB condition. At a later latency, LPC amplitudes during RB learning were larger for correct than for incorrect trials, whereas LPC amplitudes during II learning categorization were not modulated by success (**Figure 7**). In addition, a central P300 ERP to positive feedback was correlated with accuracy for the RB but not the II condition (**Figure 8**). Together, these differences in brain waves associated with category learning expand on related results from neuropsychological and fMRI studies. In addition, the current findings add neurocognitive information about the temporal order of processing, as discussed further below. Moreover, the lack of ERP differences for stimuli prior to learning makes it possible to rule out trivial physical stimulus factors. Accordingly, we attribute these ERP differences to the distinctive neurocognitive computations engaged during category learning and use.

RB processing is thought to depend on hypothesis testing, whereby a candidate rule is evaluated by comparing the representation of the stimulus in the current trial to that of a

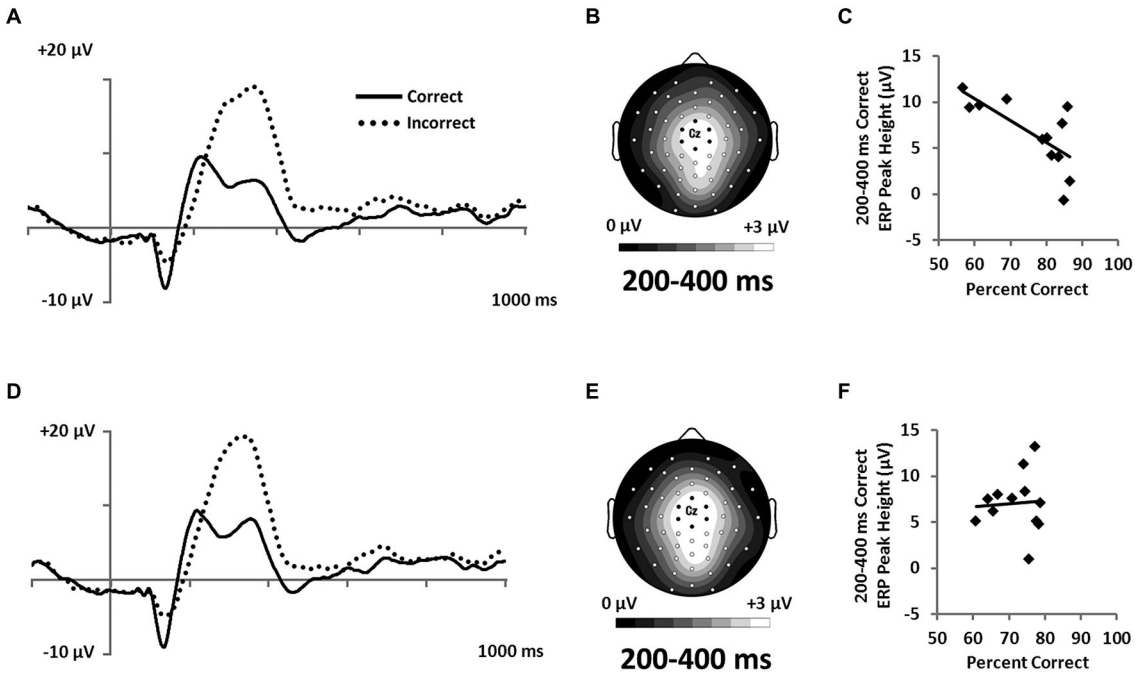

FIGURE 8 | Feedback-locked ERPs from a central cluster of electrodes (seven marked electrodes surrounding Cz; see Figure 3D for precise electrode locations) for (A) RB and (D) II category-learning conditions. Topographic maps representing mean amplitude from correct responses from 200–400 ms for **(B)** RB and **(E)** II ERPs. Scatterplots showing the relationship of accuracy to mean amplitude for correct trials for **(C)** RB and **(F)** II conditions.

representation of a category threshold. This evaluation requires selective attention and working memory, likely implemented in PFC, as well as the ability to form enduring mental representations of the rule and the threshold, dependent on the hippocampus and MTL. In contrast, II learning may be likened to gaining category expertize with complex objects such as faces (Bentin et al., 1996) or Greebles (Rossion et al., 2002).

ERP results were consistent with both of these descriptions. The more positive potential for correct compared to incorrect RB trials late during each trial (**Figure 7**) is similar to positive potentials that have been found in many different tasks and variously referred to as the P3b, P600, or LPC. These positive potentials with broad parietal topographies have been associated with working memory (Kok, 2001; Polich, 2007) and episodic memory retrieval (Paller et al., 1988, 2009; Halgren et al., 1994; Fernández et al., 1999; Guillem et al., 1999). The LPC found here may reflect retrieval/updating of the categorization rule and some mental representation of the boundary condition, two functions consistent with the function of anatomical regions previously associated with the RB category-learning system (Filoteo et al., 2005; Seger and Cincotta, 2006; Nomura et al., 2007a; Seger et al., 2011; Nomura and Reber, 2012). Likewise, we only found these LPC differences when participants' categorization response patterns suggested they are using a simple rule based on a single feature. Similarly, the magnitude of the Correct/Incorrect difference was positively correlated with individual participant categorization success.

LPC potentials were also apparent in the II condition, but there were no reliable differences between Correct and Incorrect trials, and the magnitude of the Correct/Incorrect difference was unrelated to individual participant categorization success. One possible explanation for the elevation of the LPC here is that the neural machinery responsible for the LPC is engaged during the II condition; however it is not responsible for successful categorization. This interpretation of the LPC is consistent with context-updating theory whereby information from an incoming stimulus results in revision of a maintained mental representation (Donchin, 1981). Given the gradual nature of feedback learning it is likely that participants are updating the mental representation of the boundary condition throughout successful RB learning. In contrast, when participants are relatively confident of the rule they are using, but uncertain about whether a given stimulus is an A or B they may not update (lower LPC). In the II condition they are constantly trying to update their rule and/or boundary condition, but this does not result in successful learning. In this interpretation the neural systems responsible for the LPC is engaged during II learning, but it's output is likely inhibited (Ashby and Maddox, 2011) and thus not responsible for the final behavioral decisions. Nomura and Reber (2012) proposed that RB and II systems are both active and interact competitively during categorization with the DLPFC resolving this competition based on appraising confidence in both systems. Our LPC ERP is consistent with this proposal that the explicit category-learning system is engaged in both the RB and II tasks, but it is only effective

in guiding optimal categorization performance in the RB condition.

We also observed an early occipital Correct/Incorrect difference wave (**Figure 6**). A prior visual category learning study also as associated with implicit category learning N1 ERP (Curran et al., 2002). The authors speculated that this ERP could be related to the N170 ERP frequently observed in studies of face processing (e.g., Bentin et al., 1996) and expert categorization (e.g., Tanaka and Curran, 2001; Rossion et al., 2002). This type of processing frequently engages extrastriate visual cortex (e.g., Kanwisher et al., 1997; Gauthier et al., 1999), an area found to be more active in the II condition of this task (Nomura and Reber, 2012) and previously implicated in several other category-learning tasks (Reber et al., 1998a,b). The early time-course of our effect suggests a shaping of visual perception that occurs as part of the category learning process in tasks like II categorization.

One hypothesis is that the observed N1 may reflect the degree to which a participant uses holistic processing to process the sine-wave gratings. Ashby and Maddox (2011) have argued that II tasks encourage participants to integrate perceptual information from different stimulus features at a predecisional level. In contrast, RB tasks encourage participants to consider single features and judge them against a rule.[1] Thus, holistic processing is advantageous with the II distribution, while it may be detrimental with the RB distribution where attention to spatial orientation could distract the participant from focusing on the spatial frequency information necessary to appraise the rule used to define the RB categories in this study. The presence of the N1 effect in both RB and II conditions is also consistent with the idea that both processes are regularly active during categorization, but that the results of the earlier II process may be inhibited to allow the RB to respond (Ashby and Maddox, 2011).

The electrophysiological methods used in this study also allowed us to separate neural correlates of categorization accuracy from neural signals accompanying feedback. We observed a differential Correct/Incorrect P300 response during feedback that did not differ in amplitude between RB and II conditions (**Figure 8**). However, feedback-related P300 amplitude on correct trials negatively correlated with RB accuracy but not with II accuracy (**Figures 8C,F**). P300 responses to feedback may be sensitive to expectancies, as in prior studies with very different tasks (e.g., Courchesne et al., 1977; Duncan-Johnson and Donchin, 1977; Johnson and Donchin, 1980), and when participants receive unexpected feedback (Hajcak et al., 2005). In the present case, the observed correlations may reflect an explicit/implicit distinction between RB and II category-learning strategies. Specifically, over trials participants in the RB condition are developing a hypothesized categorization rule including a representation for the boundary condition for that rule. Each new stimulus is considered with respect to this context. When those expectations are confirmed by positive feedback, participants are less surprised the more confident they are in their rule and boundary condition representation. In contrast, while participants perform similarly with respect to accuracy in the II condition, they do not become confident in their rule because an explicit RB rule is not driving their performance. This result is consistent with participants' self-reports, which indicate confidence in their rule description after RB learning and little to no confidence after II learning. Thus, these results provide further evidence for an explicit/implicit distinction between RB and II learning.

The majority of our ERP analyses in this study are based on correct/incorrect subtractions that seek to isolate what is unique about successful RB and II categorization. The advantage of this subtractive approach (see also Nomura et al., 2007a) is that aspects of the two tasks that may be common such as seeing the stimulus, making a response, and hearing feedback are subtracted away leaving us with what is unique. However, this means by definition that our descriptions of RB and II category learning are incomplete because these common processes are certainly part of the whole mechanism and may be important to achieve a full understanding of category learning. Likewise, it is difficult for us to use this approach to look at how the category-learning processes changes over time as so does the balance of correct and incorrect trials. Given successful learning, correct trials are more abundant at the end of the experiment than at the beginning when their neural correlates are likely more affected by guessing with either RB or II distributions. These factors are both important, particularly when we consider categories that may be learned and used frequently over the course of a lifetime. Recently, in their ambitious study of expertise in category learning (participants performed 10,000 trials over the course of the experiment compared to our 320 trials), Waldschmidt and Ashby (2011) demonstrated that even when considering just a single distribution type the neural correlates responsible for category use can change as participants approach expertise in categorization.

In summary, the present ERP findings illustrate two distinct neurocognitive processes responsible for successful category learning. These processes appear to compete on each categorization trial. The II process utilizes a network including, but not limited to the occipital cortex likely reflecting changes in perceptual processing as a result of implicit category learning. In contrast the more deliberative RB process occurs later during processing of a stimulus and employs more anterior cortical regions associated with working and long-term memory, most likely in association with MTL networks. In addition, neural activity measured during feedback suggests participants are aware of their learning when using an RB process to make their categorization decisions, but not when they are using the II process. Our findings do not appear to arise from differences in stimuli, but rather stem from differences in the neurocognitive processes which can be engaged while learning different types of categories. This experimental approach provides new perspectives on these category-learning mechanisms as well as a new way to investigate their interaction and competition during learning.

[1]While not used in this study, RB tasks can also require use of a conjunctive rule whereby information about more than one feature is evaluated against a more complex rule at a later stage of processing.

Human Research Statement

Humans participated in this experiment according to procedures approved by the Northwestern University Institutional Review Board. Before beginning the experiment, participants were required to read and sign the informed consent form. They were encouraged to ask any questions and had the option of leaving at any time with no adverse consequences. The informed consent forms are kept on record in the lab.

Acknowledgments

We thank Emi Nomura for programing DBT Models, Joel Voss and John Rudoy for technical assistance, Richard Greenblatt, Mark Pfielger and Demetrios Voreades from Source Signal Imaging, and Courtney Clark and Ilya Bendich for assistance in data collection. We are also grateful to our two reviewers, Todd Maddox and Carol Seger, for their comments to an earlier version of the manuscript. Generous support for the authors was provided by the American Federation of Aging Research and Rosalinde and Arthur Gilbert Foundation (RGM, KLB), the Illinois Department of Public Health (RGM, KLB), the Loyola University Chicago Dean of Arts and Sciences and the Graduate School (RGM), and the Northwestern University Mechanisms of Aging and Dementia Training Grant funded by the National Institute on Aging (2T32AG020506, RGM). Preliminary versions of these results were presented at the 31st Annual Conference of the Cognitive Science Society, Amsterdam, Netherlands, and the 2009 Cognitive Neuroscience Society annual meeting.

References

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., and Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychol. Rev.* 105, 442–481. doi: 10.1037//0033-295x.105.3.442

Ashby, F. G., Ell, S. W., Valentin, V. V., and Casale, M. B. (2005). FROST: a distributed neurocomputational model of working memory maintenance. *J. Cogn. Neurosci.* 17, 1728–1743. doi: 10.1162/089892905774589271

Ashby, F. G., Ell, S. W., and Waldron, E. M. (2003). Procedural learning in perceptual categorization. *Mem. Cognit.* 31, 1114–1125. doi: 10.3758/bf03196132

Ashby, F. G., and Maddox, W. T. (1993). Relations between prototype, exemplar and decision bound models of categorization. *J. Math. Psychol.* 37, 372–400. doi: 10.1006/jmps.1993.1023

Ashby, F. G., and Maddox, W. T. (2011). Human category learning 2.0. *Ann. N Y Acad. Sci.* 1224, 147–161. doi: 10.1111/j.1749-6632.2010.05874.x

Bentin, S., Allison, T., Puce, A., Perez, E., and McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *J. Cogn. Neurosci.* 8, 551–565. doi: 10.1162/jocn.1996.8.6.551

Cincotta, C. M., and Seger, C. A. (2007). Dissociation between striatal regions while learning to categorize via feedback and via observation. *J. Cogn. Neurosci.* 19, 249–265. doi: 10.1162/jocn.2007.19.2.249

Courchesne, E., Hillyard, S. A., and Courchesne, R. Y. (1977). P3 waves to the discrimination of targets in homogeneous and heterogeneous stimulus sequences. *Psychophysiology* 14, 590–597. doi: 10.1111/j.1469-8986.1977.tb01206.x

Curran, T., Tanaka, J. W., and Weiskopf, D. M. (2002). An electrophysiological comparison of visual categorization and recognition memory. *Cogn. Affect. Behav. Neurosci.* 2, 1–18. doi: 10.3758/cabn.2.1.1

Donchin, E. (1981). Presidential address, 1980. surprise!. . .surprise? *Psychophysiology* 18, 493–513. doi: 10.1111/j.1469-8986.1981.tb01815.x

Duncan-Johnson, C. C., and Donchin, E. (1977). On quantifying surprise: the variation of event-related potentials with subjective probability. *Psychophysiology* 14, 456–467. doi: 10.1111/j.1469-8986.1977.tb01312.x

Fernández, G., Effern, A., Grunwald, T., Pezer, N., Lehnertz, K., Dümpelmann, M., et al. (1999). Real-time tracking of memory formation in the human rhinal cortex and hippocampus. *Science* 285, 1582–1585. doi: 10.1126/science.285.5433.1582

Filoteo, J. V., Maddox, W. T., Simmons, A. N., Ing, A. D., Cagigas, X. E., Matthews, S., et al. (2005). Cortical and subcortical brain regions involved in rule-based category learning. *Neuroreport* 16, 111–115. doi: 10.1097/00001756-200502080-00007

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., and Gore, J. C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nat. Neurosci.* 2, 568–573. doi: 10.1038/9224

Guillem, F., Rougier, A., and Claverie, B. (1999). Short- and long-delay intracranial ERP repetition effects dissociate memory systems in the human brain. *J. Cogn. Neurosci.* 11, 437–458. doi: 10.1162/089892999563526

Hajcak, G., Holroyd, C. B., Moser, J. S., and Simons, R. F. (2005). Brain potentials associated with expected and unexpected good and bad outcomes. *Psychophysiology* 42, 161–170. doi: 10.1111/j.1469-8986.2005.00278.x

Halgren, E., Baudena, P., Heit, G., Clarke, J. M., Marinkovic, K., Chauvel, P., et al. (1994). Spatio-temporal stages in face and word processing. 2. depth-recorded potentials in the human frontal and rolandic cortices. *J. Physiol. Paris* 88, 51–80. doi: 10.1016/0928-4257(94)90093-0

Huang-Pollock, C. L., Maddox, W. T., and Karalunas, S. L. (2011). Development of implicit and explicit category learning. *J. Exp. Child Psychol.* 109, 321–335. doi: 10.1016/j.jecp.2011.02.002

Johnson, R. Jr., and Donchin, E. (1980). P300 and stimulus categorization: two plus one is not so different from one plus one. *Psychophysiology* 17, 167–178. doi: 10.1111/j.1469-8986.1980.tb00131.x

Kanwisher, N., McDermott, J., and Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17, 4302–4311.

Kéri, S. (2003). The cognitive neuroscience of category learning. *Brain Res. Brain Res. Rev.* 43, 85–109. doi: 10.1016/S0165-0173(03)00204-2

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology* 38, 557–577. doi: 10.1017/s0048577201990559

Maddox, W. T., Ashby, F. G., and Bohil, C. J. (2003). Delayed feedback effects on rule-based and information-integration category learning. *J. Exp. Psychol. Learn. Mem. Cogn.* 29, 650–662. doi: 10.1037/0278-7393.29.4.650

Nomura, E. M., Maddox, W. T., Filoteo, J. V., Ing, A. D., Gitelman, D. R., Parrish, T. B., et al. (2007a). Neural correlates of rule-based and information-integration visual category learning. *Cereb. Cortex* 17, 37–43. doi: 10.1093/cercor/bhj122

Nomura, E. M., and Reber, P. J. (2008). A review of medial temporal lobe and caudate contributions to visual category learning. *Neurosci. Biobehav. Rev.* 32, 279–291. doi: 10.1016/j.neubiorev.2007.07.006

Nomura, E. M., Maddox, W. T., and Reber, P. J. (2007b). "Mathematical models of visual category learning enhance fMRI data analysis," in *Proceedings of the 29th Annual Meeting of the Cognitive Science Society*, eds D. McNamara and G. Trafton (Austin, TX: Cognitive Science Society), 539–544.

Nomura, E. M., and Reber, P. J. (2012). Combining computational modeling and neuroimaging to examine multiple category learning systems in the brain. *Brain Sci.* 2, 176–202. doi: 10.3390/brainsci2020176

Paller, K. A., Voss, J. L., and Westerberg, C. E. (2009). Investigating the awareness of remembering. *Perspect. Psychol. Sci.* 4, 185–199. doi: 10.1111/j.1745-6924.2009.01118.x

Paller, K. A., Zola-Morgan, S., Squire, L. R., and Hillyard, S. A. (1988). P3-like brain waves in normal monkeys and in monkeys with medial

temporal lesions. *Behav. Neurosci.* 102, 714–725. doi: 10.1037//0735-7044.102.5.714

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* 118, 2128–2148. doi: 10.1016/j.clinph.2007.04.019

Reber, P. J., Stark, C. E., and Squire, L. R. (1998a). Contrasting cortical activity associated with category memory and recognition memory. *Learn. Mem.* 5, 420–428.

Reber, P. J., Stark, C. E., and Squire, L. R. (1998b). Cortical areas supporting category learning identified using functional MRI. *Proc. Natl. Acad. Sci. U S A* 95, 747–750. doi: 10.1073/pnas.95.2.747

Rips, L. J., Smith, E. E., and Medin, D. L. (2012). "Concepts and categories: memory, meaning and metaphysics," in *Oxford Handbook of Thinking and Reasoning*, eds K. J. Holyoak and R. G. Morrison (New York, NY: Oxford University Press), 177–209.

Rossion, B., Gauthier, I., Goffaux, V., Tarr, M. J., and Crommelinck, M. (2002). Expertise training with novel objects leads to left-lateralized facelike electrophysiological responses. *Psychol. Sci.* 13, 250–257. doi: 10.1111/1467-9280.00446

Seger, C. A., and Cincotta, C. M. (2006). Dynamics of frontal, striatal and hippocampal systems during rule learning. *Cereb. Cortex* 16, 1546–1555. doi: 10.1093/cercor/bhj092

Seger, C. A., Dennison, C. S., Lopez-Paniagua, D., Peterson, E. J., and Roark, A. A. (2011). Dissociating hippocampal and basal ganglia contributions to category learning using stimulus novelty and subjective judgments. *Neuroimage* 55, 1739–1753. doi: 10.1016/j.neuroimage.2011.01.026

Seger, C. A., and Miller, E. K. (2010). Category learning in the brain. *Annu. Rev. Neurosci.* 33, 203–219. doi: 10.1146/annurev.neuro.051508.135546

Smith, E. E., and Grossman, M. (2008). Multiple systems of category learning. *Neurosci. Biobehav. Rev.* 32, 249–264. doi: 10.1016/j.neubiorev.2007.07.009

Squire, L. R. (2009). Memory and brain systems: 1969–2009. *J. Neurosci.* 29, 12711–12716. doi: 10.1523/JNEUROSCI.3575-09.2009

Tanaka, J. W., and Curran, T. (2001). A neural basis for expert object recognition. *Psychol. Sci.* 12, 43–47. doi: 10.1111/1467-9280.00308

Voss, J. L., and Paller, K. A. (2008). Brain substrates of implicit and explicit memory: the importance of concurrently acquired neural signals of both memory types. *Neuropsychologia* 46, 3021–3029. doi: 10.1016/j.neuropsychologia.2008.07.010

Waldschmidt, J. G., and Ashby, G. (2011). Cortical and striatal contributions to automaticity in information-integration categorization. *Neuroimage* 56, 1791–1802. doi: 10.1016/j.neuroimage.2011.02.011

Yamauchi, T., and Markman, A. B. (1998). Category learning by inference and classification. *J. Mem. Lang.* 39, 124–148. doi: 10.1006/jmla.1998.2566

Zeithamova, D., and Maddox, W. T. (2006). Dual-task interference in perceptual category learning. *Mem. Cognit.* 34, 387–398. doi: 10.3758/bf03193416

Zeithamova, D., and Maddox, W. T. (2007). The role of visuospatial and verbal working memory in perceptual category learning. *Mem. Cognit.* 35, 1380–1398. doi: 10.3758/bf03193609

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.