Multidimensional Rule, Unidimensional Rule, and Similarity Strategies in Categorization: Event-Related Brain Potential Correlates

Jonathan R. Folstein and Cyma Van Petten
University of Arizona

Forty participants assigned artificial creatures to categories after explicit rule instruction or feedback alone. Stimuli were typical and atypical exemplars of 2 categories with independent prototypes, conflicting exemplars sharing features of both categories, and "Others" with only 1 or 2 features of the well-defined categories. Ten feedback-only participants spontaneously adopted a unidimensional rule; 10 used a multidimensional similarity strategy. Event-related potentials (ERPs) recorded during the transfer phase showed a commonality between multidimensional rule and similarity strategies in late frontal brain activity that differentiated both from unidimensional rule use. Multidimensional rule users alone showed an earlier prefrontal ERP effect that may reflect inhibition of responses based on similarity. The authors also discuss the role of declarative memory for features and exemplars.

Theories of categorization can be broadly divided into those based on assessing the similarity between an item that must be classified and known members of a category (exemplar and prototype theories) and those based on applying rules to determine category membership (Hahn & Chater, 1998; Medin & Schaffer, 1978; E. E. Smith, Patalano, & Jonides, 1998; E. E. Smith & Sloman, 1994; J. D. Smith & Minda, 1998). The traditional hallmark of rule use is that application of a fixed set of criteria results in an all-or-none category boundary rather than in graded membership according to degree of similarity to a prototype or other category members. Most researchers agree that categorization performance can reflect a mixture of strategies that differ in their speed and utility for different sorts of materials and in their demands on various cognitive processes (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Nosofsky, Palmeri, & McKinley, 1994; E. E. Smith et al., 1998; E. E. Smith & Sloman, 1994). For natural categories, there is considerable evidence for both similarity- and rule-based strategies (Heinze, Münte, & Kutas, 1998; McCloskey & Glucksberg, 1979). Artificial stimuli have proved particularly useful for understanding different strategies, because their component features can be easily characterized, and because familiarity with features and their combinations can also be more easily controlled than for natural stimuli. For instance, Allen and Brooks (1991) were able to dissociate strategies in the categorization of artificial animals after instructing only some participants in a given rule but found that similarity to trained exemplars influenced behavior even when a rule strategy was dominant.

In this study, we asked participants to classify artificial humanoids (see Figure 1) and compared performance of participants trained by rule instruction or by feedback alone. In addition to accuracy and reaction-time measures, we recorded participants' brain electrical activity. We addressed two broad issues: (a) whether there is a pattern of brain activity specific to rule use, and whether such a pattern depends on the nature of the rule applied; and (b) the role of memory retrieval during categorization under different strategies.

What's Special About Rule Use?

In a recent review, E. E. Smith et al. (1998) described several cognitive differences assumed to differentiate similarity- and rule-based strategies. Among these is a larger burden on working memory during rule use, because the rule itself must be maintained in working memory. If the rule is multidimensional, a count of matching and/or mismatching features will also need to be held and updated as the stimulus is processed. According to E. E. Smith et al., rule use also entails decomposition of a stimulus into features or properties, with serial attention to each, leading to the typical result of slower responses from rule-instructed participants. An additional distinction offered by E. E. Smith et al. is that rule use is strategic, whereas similarity assessment is automatic. Our interpretation of this latter distinction is that successful rule application may require overriding rapidly available associations between a given response and a single feature (or collections of features) until the entire stimulus has been assessed in light of the rule.

Several of these component processes—working memory, directing attention, and inhibition of prepotent responses—have been attributed to prefrontal cortex (D'Esposito, Postle, & Rypma, 2000; Knight, Staines, Swick, & Chao, 1999; Kopp, Rist, & Mattler, 1996). Prefrontal cortex has also been associated with the processes of rule discovery and shifting from one simple rule to another, via studies of brain-damaged patients (Lezak, 1995; Milner, 1963; Stuss et al., 2000) and hemodynamic imaging studies in healthy individuals (Nagahama et al., 1999; Rogers, Andrews,
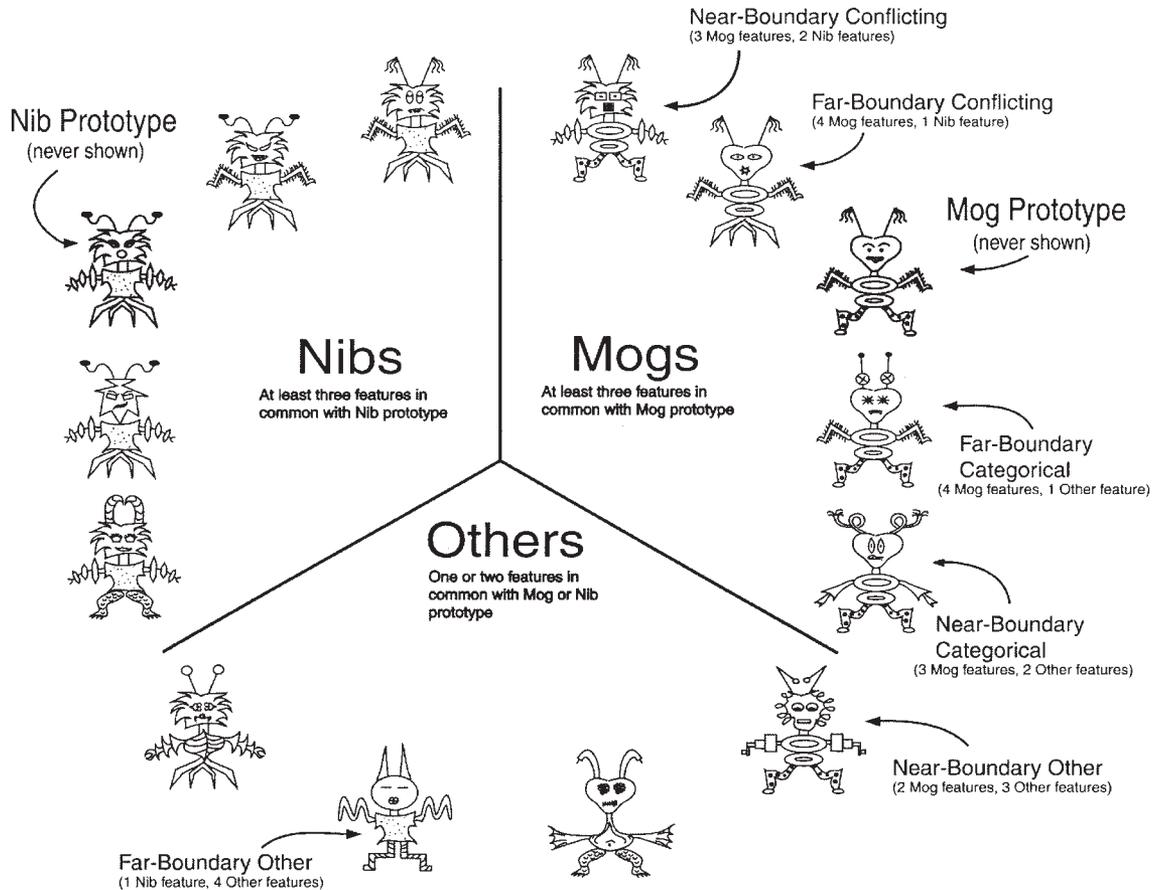
Figure 1. Examples of the Categorical (includes New and Familiar), Conflicting, and Other stimulus types, arranged by distance from category boundary.

Grasby, Brooks, & Robbins, 2000; Strange, Henson, Friston, & Dolan, 2001).

Other work suggests that prefrontal cortex can be critical for rule application as well, but indicates a need for careful characterization of the rules themselves. Robin and Holyoak (1995) reviewed evidence from animal studies that prefrontal lesions impair performance on problems of high relational complexity as defined by the number of elements and the number of relationships between elements that must be considered simultaneously. Problems that include temporal delays between the presentation of individual relationships, or problems for which a response based on a simple relationship is opposed to a correct response based on higher-order relationships, are particularly vulnerable to prefrontal damage. Humans with damage to frontal cortex perform normally on Raven matrices problems that are nonrelational or that require consideration of one dimension (e.g., color) of the elements of a matrix but perform poorly when consideration of two dimensions is required (Waltz et al., 1999). In a functional magnetic resonance imaging (fMRI) study of normal participants, solving two-relation problems led to increased blood flow in the inferior and middle frontal gyri and the caudate nucleus, as compared with the nonrelational and one-relation problems (which did not differ from one another; Christoff et al., 2001). In categorization tasks closer to our

interests in this article, assigning artificial creatures to categories via a multidimensional rule led to greater prefrontal blood flow than baseline conditions (Goel & Dolan, 2000; Patalano, Smith, Jonides, & Koeppe, 2001). However, neither included a contrast with unidimensional rule use.

Waldron and Ashby (2001) have compared unidimensional to multidimensional rule use and, contrary to the suggestions above, concluded that unidimensional rule use was more dependent on prefrontal cortex. Their theoretical claim is that unidimensional rules are *explicit,* or easily verbalized, and verbalizable rules preferentially engage prefrontal cortex (Ashby et al., 1998). The empirical conclusion rested on greater interference from a concurrent task of number comparison when participants were learning and applying multidimensional rules as compared with when they were learning and applying unidimensional rules, combined with the assumption that the number-comparison task relied on prefrontal cortex. No measures of neural activity were recorded, so the inference of greater prefrontal involvement for unidimensional rule use was indirect.

In this study, we compared brain electrical activity from participants explicitly instructed to use a multidimensional and deterministic rule with activity from those trained by feedback only. Our most general prediction was that rule-instructed participants

would exhibit larger ERPs at scalp sites overlying prefrontal cortex. However, we also subjected the strategies used by individuals to Bayesian analyses to determine the number of dimensions governing their responses. Although not instructed in a rule, feedback-only participants may or may not construct such rules on their own, and such spontaneous rules may be unidimensional or multidimensional. We thus examined whether the magnitude of prefrontal activity is contingent on individual strategy in addition to experimental instruction.

Memory Retrieval During Rule- and Similarity-Based Categorization

Theories of similarity-based categorization posit a central role for memory retrieval during classification of new items: retrieval of either a category prototype, or one or more previously studied exemplars for comparison to the new item (Medin & Schaffer, 1978; Nosofsky, 1986; J. D. Smith & Minda, 1998). Theoretical descriptions of rule application place much less emphasis on memory retrieval. In some accounts (E. E. Smith et al., 1998), long term memory demands are described only as storage of the rule itself. However, further thought suggests that even pure rule application should rely on memory for the features themselves. This is particularly true of paradigms using artificial creatures, in which the rule might specify "a member of Category A must have at least three Type A body parts": Participants must consult visual memories of Type A legs, torsos, arms, and so on in addition to the verbal rule. Feature memories may coexist alongside memories for prototypical combinations of features and/or individual exemplars. One question we address in the present study is whether rule-instructed and feedback-only participants rely differentially on memories for individual features and/or exemplars.

A small but persuasive body of literature indicates that some varieties of category learning can occur without access to the form of conscious memory that underlies performance in direct recognition tests for exemplars (*declarative* memory, Squire, 1987). Knowlton and Squire showed that amnesic patients can classify dot patterns as similar or dissimilar to those studied earlier, although they performed very poorly in old–new recognition tests of studied and unstudied dot patterns (Knowlton & Squire, 1993; Squire & Knowlton, 1995). These dot-pattern studies differ in a critical way from other categorization paradigms in that an old category was compared with an unlearned category or with random patterns. It is much less clear whether the learning process employed by amnesics affords distinctions among two or more equally familiar categories (see Filoteo, Maddox, & Davis, 2001; Kolodny, 1994; Zaki, Nosofsky, Jessup, & Unverzagt, 2003, for variable results on this issue).

In recent studies, researchers have investigated the abilities of amnesic patients to form multiple categories of artificial animals, for which we suggest that some form of memory for individual features, if not memory for whole exemplars, is necessary. Kitchener and Squire (2000) concluded that when descriptions of the animals are presented verbally, amnesics are unable to distinguish between two categories because "each spoken item may need to be retained in memory long enough for the common elements to be identified across trials and for a common theme to be constructed" (p. 910). In contrast, with visual presentation of animals and an overlapping group of patients, Reed, Squire, Patalano, Smith, and

Jonides (1999) concluded that categorization performance was intact in the patient group, despite absence of declarative memory for the individual features. However, the tests of categorization and declarative memory in this latter study were poorly matched in a number of ways. The stimuli consisted of nine-dimensional animals, with two possible values for each dimension (e.g., body markings are stripes or spots). After studying exemplars possessing seven or eight features with a "Peggle" value, participants were asked to classify new exemplars as Peggles or not-Peggles. During this categorization test, one could thus achieve perfect accuracy based on memory for Peggle features, and only a signal of unfamiliarity for nonPeggle features. In contrast, the declarative memory test consisted of presentation of the name of each dimension (e.g., "body markings"), and verbal responses were scored as completely correct only if they indicated memory for both values (e.g., "body markings can be spots or stripes"); a participant reporting Peggle values alone would thus achieve only 50% correct. In addition to the differential weight accorded to nonPeggle features in the two tests, the declarative memory test was one of cued recall after a modality change between the study and test phases, whereas the categorization test could be construed as a (typically easier) recognition test. Zaki and Nosofsky (2001) have similarly criticized the match between the two tests by noting that attention to (and memory for) only a small number of features could result in high accuracy in the categorization test but not in the declarative memory test. It is possible that the two tests in Reed et al. (1999) tapped the same variety of memory for features, but with differential difficulty.

Categorization of artificial stimuli has also been studied with hemodynamic imaging methods in normal participants, in which one metric of access to declarative memory is activity in medial temporal lobe structures. These have and have not shown medial temporal lobe activity during dot-pattern classification (Aizenstein et al., 2000, vs. Reber, Stark, & Squire, 1998a, 1998b; Reber, Wong, & Buxton, 2002). During categorization of artificial animals, Patalano et al. (2001) observed no medial temporal activity in either rule-instructed or feedback-only participants, providing no support for the suggestion that similarity-based strategies rely more strongly on memory retrieval than do rule-based strategies. In contrast, Goel and Dolan (2000) did observe such activity in participants who successfully used feedback to induce the category rule, as compared with participants given the rule in advance. It is difficult to draw conclusions from positive versus null effects in fMRI and positron emission tomography studies at present, as hemodynamic signals from the medial temporal lobe are not always observed during the retrieval phase of canonical declarative-memory tasks (see Table 8 in Cabeza & Nyberg, 2000), and studied versus unstudied items frequently do not produce different hemodynamic signals in this region (and sometimes in no region reported; e.g., Buckner et al., 1998; Henson et al., 2003; Henson, Rugg, Shallice, Josephs, & Dolan, 1999; Lepage, Brodeur, & Bourgouin, 2003; Nyberg, 2000; Rugg, Fletcher, Chua, & Dolan, 1999).

In this study, we compared brain electrical activity elicited by familiar exemplars presented during the training and test phases to novel exemplars presented only at test as an assay of memory for whole exemplars. Although less clearly localized to particular brain structures, event-related potentials (ERPs) robustly differentiate new from incidentally repeated items in a variety of para-

digms (see Friedman & Johnson, 2000; Rugg, 1995 for reviews). When participants offer a recognition judgment, this *old–new effect* is contingent on successful retrieval, in that recognition hits are distinct from correct rejections (CR), misses, and false alarms (Neville, Kutas, Chesney, & Schmidt, 1986; Rubin, Van Petten, Glisky, & Newberg, 1999; Van Petten & Senkfor, 1996). Because the old–new effect during incidental repetition has been largely indistinguishable from the hit–CR difference during direct recognition tests and has proven sensitive to the same depth-of-encoding manipulations, it has been attributed to the same process—conscious recognition of studied items (Paller & Kutas, 1992; Paller, Kutas, & McIsaac, 1995). The ERP old–new effect is reduced or absent in amnesic patients proportional to the severity of memory impairment (Olichney et al., 2000).

If participants retrieve exemplars from the training phase during the categorization test, we expect to observe a difference between familiar and novel exemplars that closely resembles the difference previously observed for studied and unstudied items in direct-recognition and incidental-repetition paradigms. To the extent that participants who have not been instructed in the categorization rule rely more heavily on retrieval of studied exemplars, there will be a larger difference between familiar and novel exemplars in the feedback-only group as compared with the rule group.

We used a similar metric to examine memory for features. Across both training and test, individual features occurred very frequently (>100 occurrences) or rather infrequently (no more than 12 occurrences), and the summed frequency of features in a given exemplar follows a graded pattern from very high (a combination of mostly high-frequency features) to very low (a combination of mostly low-frequency features). No previous study has examined memory for parts of a single stimulus using a measure of neural activity, but our working hypothesis was that if categorization is based on retrieval of individual features, we would observe a graded impact of the frequency of feature occurrence. Because we postulate that memory for features is necessary for both rule-based and similarity-based categorization, we predicted a feature frequency effect in both groups of participants.

Method

Participants

We paid 22 men and 18 women $8 an hour for their participation. Age range was 19–39 years (mean age was 25.8 years, with no difference between rule and feedback-only groups). Thirty-four were right-handed, and 6 were left-handed. An additional 15 individuals participated but did not generate usable data: 6 because of excessive blink and movement artifacts in their electroencephalograms, 3 because of very low categorization accuracy, 2 because they fell asleep, 2 because of equipment failure, and 2 because of experimenter error.

Stimuli

Stimuli were 330 line drawings of anthropomorphic "aliens," divided into three categories illustrated in Figure 1. Ninety were presented during the training phase, and 240 at test. *Mogs* and *Nibs* were derived from separate prototypes, whereas *Others* were too distant from either the Mog or Nib prototype to be considered members of either of these narrowly defined categories. Each alien had five *diagnostic dimensions*—antennae, head, body, arms, and legs—each instantiated by 12 possible *features*. For each diagnostic dimension, 1 feature is associated with Mogs, another with

Nibs, and 10 with all categories equally (Other feature). In addition to the diagnostic dimensions, each alien had one of 15 varieties of eyes and one of 15 varieties of mouth. Eyes and mouth were nondiagnostic dimensions because each variety was equally associated with Mogs, Nibs, and Others.

The rule specifying category membership was multidimensional and additive: An alien was a Mog if it possessed at least three (of five) Mog features, a Nib if it possessed at least three Nib features, and an Other if it possessed at least three Other features (or, equivalently, too few Mog or Nib features). Note that this rule has both a verbal and a visual component: "Three of five" is easily expressed in words, but the features to be counted—the exact nature of Mog, Nib, and Other features—would be quite difficult to describe verbally. According to the rule, 90 of the stimuli presented in the test phase were Mogs, 90 were Nibs, and 60 were Others. As described below, test stimuli varied along three other dimensions: categorical typicality, summed frequency of their composite features, and similarity of test exemplars to those presented during the training phase.

Category typicality. One third of the individual Mogs possessed four diagnostic features and were thus close to the Mog prototype, distant from the category boundaries separating Mogs from both Nibs and Others (*far boundary*). Two thirds of the Mogs possessed only three diagnostic features and were thus *near boundary*. Individual Nibs were similarly more or less prototypical of their category. Near-boundary Others were those possessing three Other features and two features characteristic of either Mogs or Nibs (but not both); far-boundary Others were those with four Other features and one Mog (or Nib) feature.

Feature frequency. Each alien can be described according to the summed frequency of the features comprising it. Because there were 10 possible Other features for each dimension and only 1 Mog and 1 Nib feature, Mog and Nib features occurred much more frequently than did Other features across the stimulus set. Aliens in the Other category thus had lower overall feature frequencies than Mogs and Nibs, but near-boundary Others had higher feature frequencies than far-boundary Others. Similarly, for Mogs composed of Mog features plus Other features, summed feature frequency was correlated with categorical typicality: A far-boundary Mog had higher feature frequency than a near-boundary Mog (and likewise for Nibs). To deconfound categorical typicality and feature frequency, an additional stimulus type was introduced. *Conflicting aliens* were Mogs and Nibs composed solely of high-frequency Mog and Nib features. A conflicting Mog was one possessing 4 Mog features and 1 Nib feature, or 3 Mog features and 2 Nib features; these differed in distance from the category boundary but were equal in feature frequency. Table 1 lists the mean feature frequency for each stimulus type and shows a gradation in five equal steps. Each Mog and Nib feature occurred with equal (high) frequency, and each Other feature with equal (low) frequency. Exemplars were evenly distributed around the two prototypes in feature space.

Whole-exemplar familiarity. At test, 180 of the 240 stimuli were novel combinations of diagnostic features. Such *New* Mogs and Nibs at test might

Table 1
Mean Feature Frequency

Condition	Feature frequency (%)
Conflictings	30.4–31.4
Far-boundary Mogs and Nibs	25.1–25.9
Near-boundary Mogs and Nibs	19.8–20.3
Near-boundary Others	14.5–14.8
Far-boundary Others	9.2

Note. The frequency of a feature (e.g., a particular head) is the percentage of total trials (training plus test phase) in which that feature occurs: roughly 31% for each Mog and Nib feature and 3.8% for each Other feature. Mean feature frequency of an alien stimulus is the average frequency of the five diagnostic features comprising it. These are described as ranges because of the adaptive training procedure: Individual participants viewed training sets from two to four times.

include the same critical features (i.e., Mog antennae, arms, and legs) as those studied in the training phase but combine with different Other features (head and body) to yield a rather different alien. The other 60 test stimuli consisted of 30 Mogs and 30 Nibs that were very similar to those presented during the training phase, differing only in the nondiagnostic eyes and mouth. Figure 2 shows examples of aliens that were *New* versus *Familiar* at test.

Procedure

The experiment was conducted in one 3.5–4-hr session. After electrode application, each participant rested in a reclining chair in an electrically and acoustically shielded chamber ~100 cm from a 20 in. (50.8 cm) video monitor under the control of a PC. Accuracy, reaction time, and ERPs were recorded during the test phase. In the training phase, accuracy was recorded only to determine whether a participant was ready to proceed to the next step in the graduated training procedure and was not subjected to formal analysis.

Training phase. During the training phase, participants categorized 90 stimuli at least twice, with feedback on each trial. Training included Mog, Nib, and Other aliens, but neither rule nor feedback-only participants were exposed to Conflicting aliens during training. Training for the rule participants began with instruction on the "three out of five" rule; Mog and Nib features were identified by the experimenter during the first block of training trials. Training for both groups was graded, beginning with long

Figure 2. Relationship of Familiar and New aliens in the transfer phase to stimuli in the training phase. During training, participants viewed stimuli composed of Categorical body parts (in this example, body, legs, and antennae) plus one or two Other features (in this example, head and arms). At test, both Familiar and New aliens are composed of the same combinations of categorical body parts, but Familiar aliens additionally include the same Other features as during the training phase. Note that Familiar aliens are not completely identical to trained items but differ in the nondiagnostic features of eyes and mouth. New aliens include the same categorical features experienced in training, combined with different Other features.

exposure durations, verbal responses, and verbal feedback from the experimenter, and progressing toward faster exposure durations, keypress responses, and feedback on the computer screen. All participants received five sets of practice trials, presented twice in a different order. When more than two errors in a category occurred in a block of 18 trials, that practice set was presented a third time. If accuracy did not reach criterion, the set was presented a fourth time. After four presentations of a set, the experiment progressed regardless of accuracy. In the long run, 2 participants were replaced because their overall accuracy was more than 2 standard deviations below the mean.

Test phase. The composition of stimuli in the test phase is shown in Table 2; the various conditions were randomly intermixed. The test phase was broken into 12 blocks of 20 trials each, with rest breaks as desired between blocks. Prior to the first block, participants were forewarned of the Conflicting aliens as follows: "Some of the aliens might look a little like Mogs and a little like Nibs." Rule-trained participants were told to apply the rule, and feedback-trained participants were told to categorize these as they saw fit.

Stimuli were presented in white on a black background for 450 ms with a 9-s intertrial interval. A fixation cross appeared on the screen between stimuli and brightened 1 s before stimulus onset as a warning cue. Keypresses with the right and left index fingers were used to make "Mog" and "Nib" responses (counterbalanced across participants within group). "Other" responses were made with the middle finger of the right or left hand (also counterbalanced).

Electrophysiological Methods

The electroencephalogram (EEG) was recorded from tin electrodes in an elastic cap (Electrocap International, Eaton, OH). Seven electrodes spanned the midline of the scalp from prefrontal to occipital (Fpz, Fz, Fcz, Cz, Cpz, Pz, Oz). Six dorsal lateral pairs were used (Fp3/4, F3/4, Fc3/4, C3/4, P3/4, and O1/2), and five more lateral pairs spanned inferior frontal and temporal sites (ventral: F7/8, Ft7/8, T4/4, Tp7/8, T5/6). The vertical electrooculogram (EOG) was recorded from an electrode below the right eye. Scalp and vertical EOG electrodes were referenced to the left mastoid during the recording, and digitally rereferenced to the mean of the right and left mastoids offline. Electrodes lateral to the external canthi of the two eyes were used to record the horizontal EOG. The EEG was amplified by a Grass Model 12 polygraph (Grass, West Warwick, RI) with half-amplitude cutoffs of 0.01 and 100 Hz, digitized online at a sampling rate of 250 Hz, and stored on CD along with stimulus codes for subsequent averaging. Trials contaminated by eye movements and blinks (EOG artifact), amplifier saturation, or movement artifact were rejected prior to averaging. Averaging epoch began 200 ms prior to stimulus presentation and continued for 1,300 ms poststimulus. Mean amplitudes in selected latency windows were measured with respect to the 200 ms prestimulus baseline.

Behavioral Results

Mean accuracies split by instructional group (rule, feedback-only), condition (Familiar, New, Other, Conflicting), and distance from the category boundaries (boundary) are shown in Figure 3. Figure 4 shows reaction times for correct trials. Both figures show nearly identical results for New and Familiar stimuli. Preliminary analyses of variance (ANOVAs) comparing these two conditions did not show main effects in accuracy or speed and there were no interactions with instructional group, so that New and Familiar are collapsed into a single *Categorical* condition for subsequent analyses.

Figures 3 and 4 show that stimuli close to the category boundaries generally received less accurate and slower responses than

Table 2
Test Phase Stimuli

Condition	Category	Boundary distance	Diagnostic features	N
New				60
	Mog	Near	3 Mog, 2 Other	20
	Mog	Far	4 Mog, 1 Other	10
	Nib	Near	3 Nib, 2 Other	20
	Nib	Far	4 Nib, 1 Other	10
Familiar				60
	Mog	Near	3 Mog, 2 Other	20
	Mog	Far	4 Mog, 1 Other	10
	Nib	Near	3 Nib, 2 Other	20
	Nib	Far	4 Nib, 1 Other	10
Other				60
	Other	Near	2 Mog, 3 Other	20
	Other	Far	1 Mog, 4 Other	10
	Other	Near	2 Nib, 3 Other	20
	Other	Far	1 Nib, 4 Other	10
Conflicting				60
	Mog	Near	3 Mog, 2 Nib	20
	Mog	Far	4 Mog, 1 Nib	10
	Nib	Near	3 Nib, 2 Mog	20
	Nib	Far	4 Nib, 1 Mog	10

Figure 4. Reaction times by stimulus type and boundary distance for the two instructional groups ($n = 20$ for each group). Error bars represent standard errors.

did those far from the boundaries. A general effect of instruction is also evident: Rule-instructed participants were generally more accurate, but slower, than feedback-only participants. Finally, the figures suggest differences in the pattern of results contingent on instruction: The rule group tended to be most accurate in the Other condition, whereas the feedback-only group was particularly disadvantaged in the Conflicting condition.

We postpone formal analyses of the accuracy and reaction time (RT) data because inspection of the data from individual participants revealed substantial individual variability. In particular, the
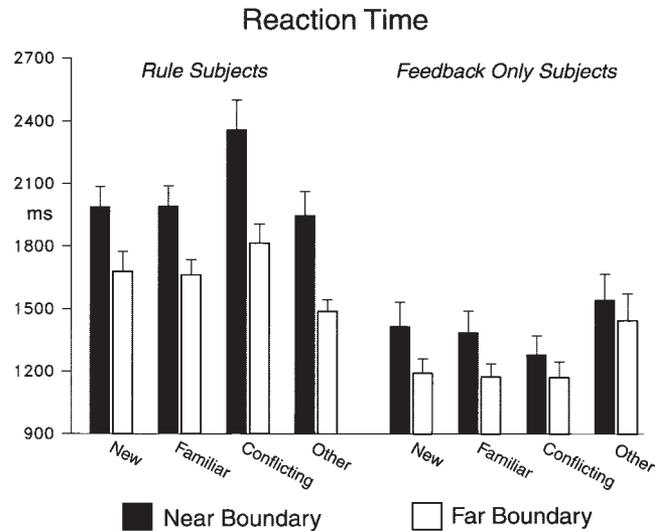
feedback-only participants did not form a uniform group. Below, we describe the methodology for elucidating the strategies of individual participants and then return to statistical analyses of the behavioral data.

Characterization of Individual Strategies

As in previous studies of rule- versus nonrule-based categorization, the instructions to the rule participants were quite specific, whereas the feedback participants were left to develop their own strategies. One such strategy might be global similarity comparisons between test items and category prototypes (or exemplars from the training phase), but other strategies might involve formulation of rules, including simple rules based on single features of the stimuli (Ward & Scott, 1987).

To describe the strategy of an individual participant, we began by calculating the Bayesian probability that he or she would classify an alien in Category A given the presence of a feature diagnostic of Category A, for example, the probability an alien would be categorized as a Mog if it had a Mog head, no matter what the alien's category was according to the experimenter-defined rule. We label this probability the feature's *endorsement rate*: p(respond "Mog") if (Mog head present). We then rank ordered the endorsement rates across the five Mog and five Nib features for each participant. The logic here is that a given participant's classifications might be driven most strongly by the heads of the aliens, less strongly by bodies, and so on, whereas a different participant might base his or her decisions most strongly on the type of arms in a stimulus, and so on. Note that if a person's classifications were completely driven by the multidimensional rule, no such biases toward specific features would be observed, as the rule accords equal weight to each of the five dimensions of each stimulus.

Figure 5 shows the mean rank-ordered endorsement rates in the rule and feedback groups. As predicted, the rule participants
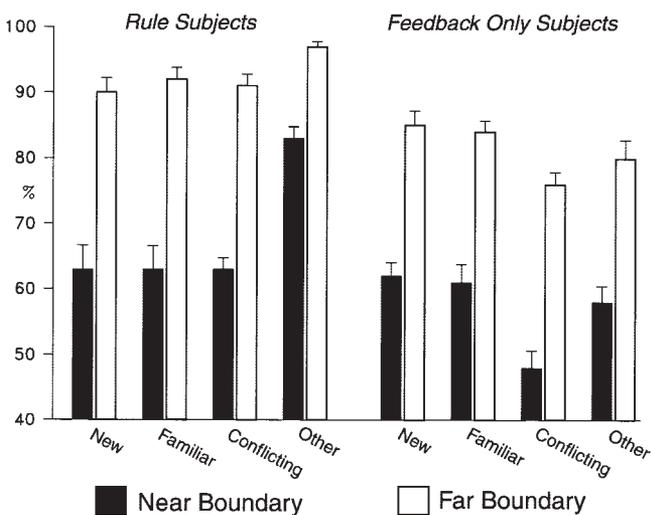

Figure 3. Accuracies by stimulus type and boundary distance for the two instructional groups ($n = 20$ for each group). Error bars represent standard errors.
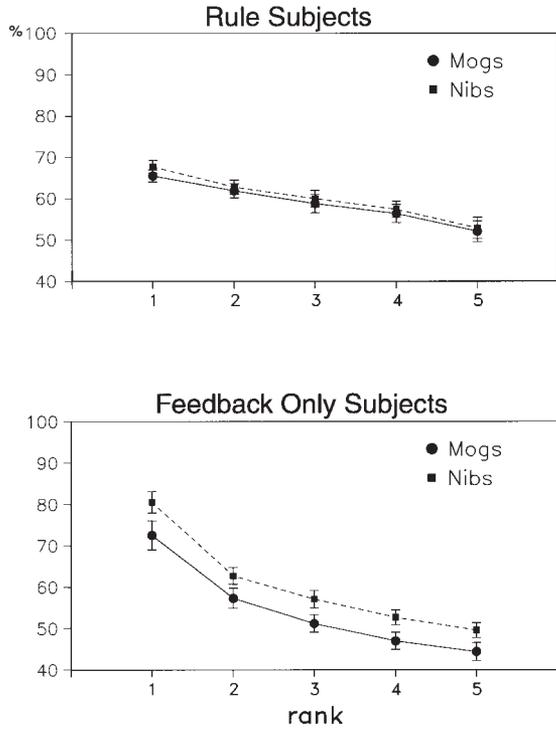
Figure 5. Mean likelihoods of endorsing a stimulus as a Mog or Nib, contingent on the stimulus containing a specific Mog or Nib feature, across all stimuli in the test phase. For each participant, features were ranked according to that eliciting the highest endorsement rate (Rank 1) through that eliciting the lowest endorsement rate (Rank 5). A 100% endorsement would mean that every stimulus containing a given feature (e.g., Nib body) was assigned to the category probabilistically associated with that feature (e.g., a response of "Nib"), whereas a low endorsement rate would mean that a given feature was ineffective in eliciting the categorical response probabilistically associated with that feature.

showed a relatively flat profile, indicating little bias toward specific features. The feedback participants showed a distinct slope, indicating that their categorization responses did not accord equal weight to all features. A regression line was fit to the average profile for each participant (Mogs and Nibs combined). Although the slopes of rule participants deviated significantly from zero (indicating incomplete adherence to the instructed rule), significantly steeper slopes were observed in the feedback group, $t(38) =$ 6.75, $p < .0005$.

Inspection of the endorsement-rate profiles for each participant indicated little individual variability within the rule group. Within the feedback group, however, inspection of individual data revealed two qualitatively distinct patterns. Ten of the feedback-only participants showed a relatively linear change in endorsement rate across features. Sample profiles for these individuals, who are referred to as the *distributed-profile* subgroup, are shown in the left column of Figure 6. This sort of profile indicates that categorizations were sensitive to multiple features of the stimuli, although some were accorded greater weight than others. For the other 10 feedback participants, "Mog" responses were strongly driven by the presence of one specific Mog feature, and "Nib" responses were strongly driven by the presence of one specific Nib feature,

whereas other features had relatively weak and relatively equal influences on response selection. Sample profiles for these individuals are shown in the right column of Figure 6 and are referred to as the *single-feature* subgroup. For 9 of 10 of these participants, the preferred Mog and preferred Nib feature were from the same dimension (e.g., both heads or both arms). This sort of profile suggests the spontaneous adoption of a unidimensional rule. A statistical analysis confirmed the visible difference between the two subgroups of feedback participants: When endorsement rates for first-rank and second-rank features were entered into an ANOVA with subgroup as a between-subjects factor, the Subgroup × Feature Rank interaction was significant, $F(1, 18) =$ 137.1, $p < .0001$.

Accuracy

An omnibus ANOVA taking group (rule, single-feature, distributed-profile), condition (Categorical, Conflicting, Other),
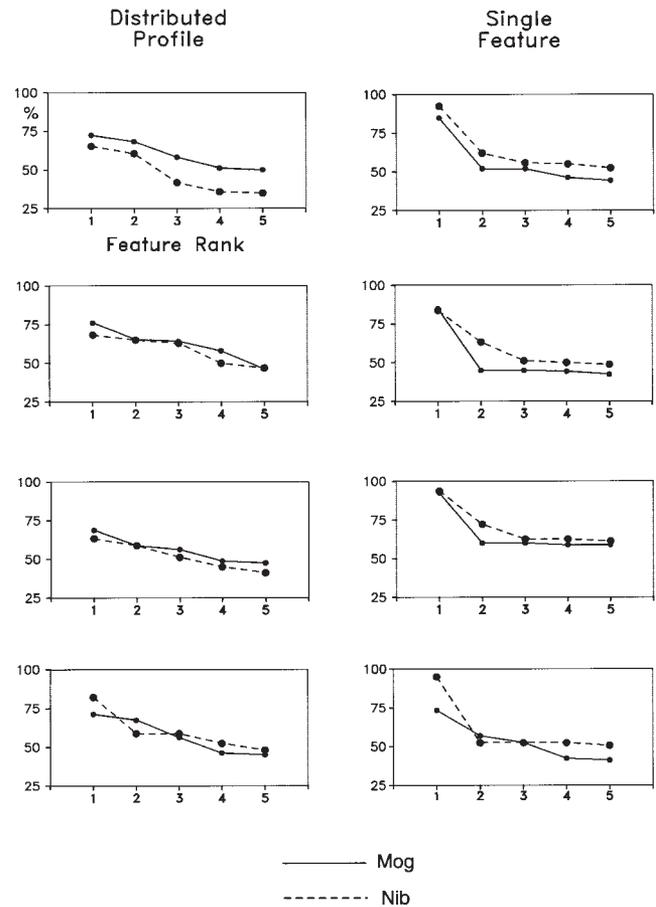

Figure 6. As in Figure 5, including response profiles from 8 of the 20 individual participants in the feedback-only group. Those on the left were classified as distributed profile on the basis of the relatively flat slope across feature ranks. Those on the right were classified as single feature because one stimulus feature was extremely effective in eliciting a response of "Mog," another stimulus feature was extremely effective in eliciting a response of "Nib," and the other eight features were relatively ineffective in eliciting the category judgments probabilistically associated with those features.

and boundary as factors produced significant main effects of all three factors (all $ps < .01$), as well as interactions: Group $\times$ Condition, $F(4, 74) = 8.72$, $p < .0001$; Condition $\times$ Boundary, $F(2, 74) = 6.59$, $p < .005$; and Group $\times$ Condition $\times$ Boundary, $F(4, 74) = 2.48$, $p = .05$. These were unpacked by follow-up analyses. Near-boundary stimuli were categorized less accurately than far-boundary stimuli for all groups and conditions (all $ts > 7.0$, $ps < .01$). Across conditions, the rule group was more accurate than either of the feedback subgroups, $Fs(1, 28) > 20.3$, $ps < .01$, which did not differ from each other in overall accuracy ($F < 1$). The distributed-profile and single-feature subgroups showed differential accuracies across conditions, however, $F(2, 36) = 4.22$, $p < .05$.

A characteristic feature of the rule participants was higher accuracy for Others than for Categorical and Conflicting conditions, $ts(19) > 4.0$, $ps < .01$, which did not differ from each other. Figure 3 shows that the advantage for Others shown by the rule participants was largely due to higher accuracy for near-boundary Others (those with two Mog or two Nib features) as compared with the other near-boundary stimuli. This was reflected in a smaller boundary effect for Others as compared with either Categoricals or Conflictings in the rule group, near minus far difference, both $ts(19) > 4.5$, $ps < .01$.

A characteristic feature of both feedback groups was poor performance on Conflictings relative to the rule group (see Figure 3). However, Figure 7 shows that the single-feature subgroup was more accurate on Conflictings than was the distributed-profile subgroup, $t(18) = 3.26$, $p < .01$, whereas the two subgroups had equivalent accuracy for Categoricals and Others. The single-feature subgroup was still less successful than the rule group in classifying Conflictings (72.6% vs. 63.5%), $t(28) = 3.60$, $p < .005$, and the relative success in handling these stimuli likely resulted from rather different strategies in the single-feature and rule participants. Basing decisions on only a single feature will tend to make one impervious to mixtures of features from two

categories in the unattended features. In contrast, we assume that the rule group was largely successful in classifying Conflicting stimuli because they did attend to all features but were able to assess the numbers of Mog versus Nib features in light of the rule.

Figure 7 also shows boundary effects on accuracy in both subgroups of feedback participants, $Fs(1, 9) > 147.0$, $ps < .01$, with no significant difference in the magnitude of the boundary effect ($F < 1$). Note that because far-boundary aliens contain more category-congruent features than their near-boundary counterparts, boundary effects are expected even when categorization is based on a single dimension. Because a far-boundary Mog has four Mog features, it has an 80% probability of containing a specific Mog feature, whereas a near-boundary Mog has a 60% probability. Similarly, far-boundary Others have an 80% probability of not containing a specific Mog or Nib feature that would mislead a participant who based his or her decisions solely on that feature, and near-boundary Others have a 60% probability of not containing a given categorical feature. The observed accuracy rates of the single-feature subgroup were quite close to the 80% and 60% rates predicted by a unidimensional categorization rule for far- and near-boundary stimuli, respectively: Accuracies in all six conditions fell within the 99% confidence interval around the predicted rates.

Reaction Time

The single-feature subgroup had faster RTs than the distributed-profile subgroup, $F(1, 18) = 8.73$, $p < .01$, whose members were in turn faster than the rule group, $F(1, 28) = 4.08$, $p = .05$, as shown in Figures 4 and 8. The omnibus ANOVA showed faster RTs for far-boundary stimuli than they did for near-boundary stimuli, $F(1, 37) = 65.0$, $p < .0001$, as well as interactions: Group $\times$ Condition, $F(4, 74) = 14.0$, $p < .01$; Group $\times$ Boundary, $F(2, 37) = 13.2$, $p < .0001$; and Group $\times$ Condition $\times$ Boundary, $F(4, 74) = 4.60$, $p < .01$.

Within the rule group, Conflicting aliens were slower than either Categorical or Others, $ts(19) > 4.0$, $ps < .01$, which did not differ from each other (see Figure 4). In contrast, Figure 8 shows that neither of the feedback subgroups were slowed by Conflicting stimuli, but were instead slowest on Others: Other versus Categorical, $F(1, 18) = 12.8$, $p < .005$; Other versus Conflicting, $F(1, 18) = 11.0$, $p < .005$; $Fs < 1$ for the interactions with subgroup.

Figure 8 also shows that the single-feature and distributed-profile subgroups had equivalent boundary effects—faster responses to typical than atypical aliens, $F(1, 18) = 24.8$, $p < .01$; $F < 1$ for Boundary $\times$ Subgroup interaction. This finding reflects a dissociation between response choice and response speed in the single-feature subgroup. Although the endorsement profiles suggest that these individuals based their final decisions on single stimulus dimensions, those decisions were speeded or slowed by the presence of supporting or conflicting features in the other dimensions.

Discussion of Behavioral Results

The instructional manipulation divided the participants into two groups: those explicitly asked to use a multidimensional conjunctive rule to divide the stimuli into three sets and those trained by feedback only. However, Bayesian analyses of the relationship
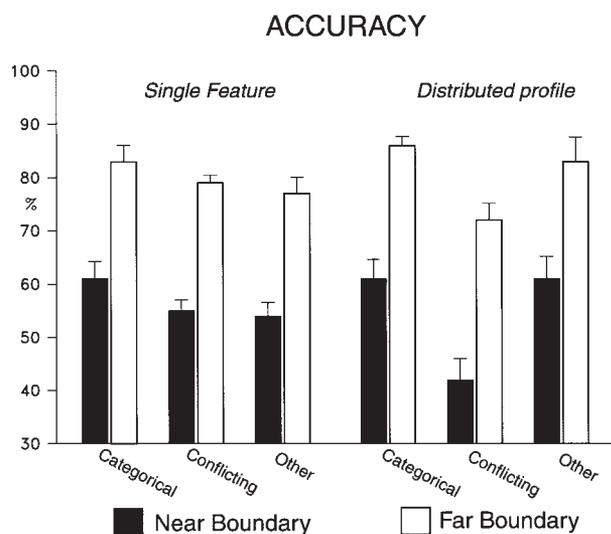
ACCURACY

Figure 7. Accuracies by stimulus type and boundary distance for the two subgroups of feedback-only participants ($n = 10$ for each group). Error bars represent standard errors.
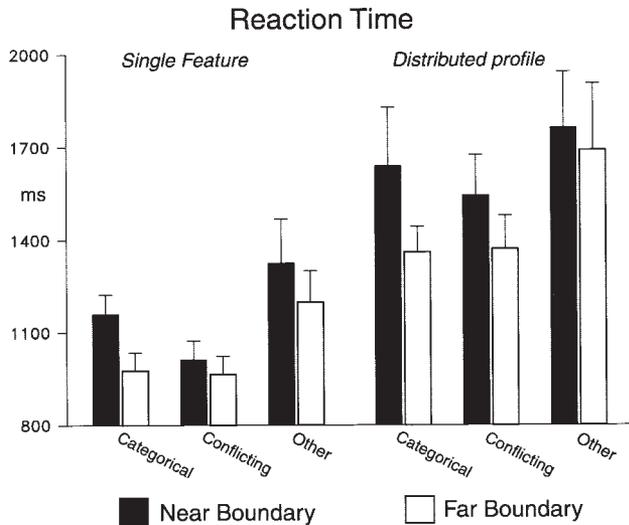
Reaction Time

Figure 8. Reaction times by stimulus type and boundary distance for the two subgroups of feedback-only participants (*n* = 10 for each group). Negative voltage is plotted upward. Error bars represent standard errors.

between stimulus features and categorization choices indicated that individual participants followed three different strategies. The rule-instructed participants largely followed the assigned rule, and as in previous studies, rule instruction resulted in slower but more accurate responses than feedback alone. For half of the feedback-only participants, categorization choices were also driven by multiple dimensions of the stimuli, although these distributed-profile participants allotted greater or lesser weight to particular dimensions in their category assignments. For the other 10 feedback participants, categorizations were strongly driven by a single stimulus dimension, with low weights allotted to the other four dimensions. The response profiles in this last group of participants suggest that they spontaneously adopted a unidimensional rule and responded merely on the basis of detecting a Mog or Nib head (or arms, etc.). The unidimensional strategy resulted in much faster responses than those observed in either the rule or distributed-profile participants. The unidimensional strategy also conferred immunity to the detrimental effects of mixing features from the Mog and Nib categories suffered by the other participants; rule participants were slowest when confronted with Conflicting aliens, and distributed-profile feedback participants showed their lowest accuracy in this condition, but single-feature feedback participants showed neither an accuracy nor a speed deficit in this condition.

From the behavioral data, we can thus conclude that the rule-instructed participants used a multidimensional rule, and that half of the feedback-only participants used a unidimensional rule. However, both sorts of rule-using participants were also influenced by the similarity between test stimuli and the central tendencies of the categories; they were faster to respond to typical members of a category (far from the categorical boundaries) than to atypical (near-boundary) stimuli. This influence of similarity, even when participants use a rule that should make the similarity irrelevant, is in accord with the results of previous studies.

The correct characterization of the strategy applied by the other half of the feedback-only participants (the distributed-profile, or

DP, group) is less obvious to casual inspection. Aside from the criterion we used to define the DP group (the Bayesian response profiles), other aspects of the behavioral data indicate that these participants used multiple stimulus dimensions in their judgments. Their overall RTs were slower than those of the single-feature group, consistent with prolonged assessment of multiple dimensions. Moreover, they were less accurate on the Conflicting aliens than they were on the stimuli that did not mix features from the two high-frequency categories, indicating that they were sensitive to the presence of multiple features associated with different categories.

From this pattern of results, we might speculate that the DP participants were applying a consistent multidimensional rule, but one that was defective in some regard: a rule that used too few dimensions, or one that misassigned a given feature to the wrong category. However, evaluation of various plausibly defective rules does not provide a close approximation to the behavior of the DP group. Consideration of a "two out of five" rule, a "three out of five" rule with one feature misassigned between Mogs and Nibs, and a rule that overlooks a relevant dimension altogether ("three out of four") all substantially underestimate the observed accuracy of the DP group. Classification of the Other aliens provides the strongest argument that the DP participants did not use a multidimensional rule. Far-boundary Others included only one Mog or Nib feature and four Other features. No multidimensional rule would suggest that these were Mogs or Nibs, and, indeed, the rule group misclassified these on only 3% of the trials. The DP group misclassified these items on 17% of the trials.

How does a similarity-based account fare in accounting for the behavior of the DP participants? The simplest similarity model is that the likelihood of classifying a stimulus as a Mog or Nib is driven by the degree of overlap between a test stimulus and the Mog or Nib prototype. (Recall that there was no prototypical Other alien; instead these stimuli were weakly similar to either Mogs or Nibs.) Far-boundary Others contain one categorical feature of the five comprising the Mog or Nib prototype, near-boundary Others contain two categorical features, near Categoricals contain three categorical features, and far Categoricals contain four categorical features, predicting endorsement rates (as Mogs or Nibs) of 20%, 40%, 60%, and 80%, respectively. In a regression relating number of categorical features to endorsement rate, this would correspond to a slope of 1.0 and a *y*-intercept of 0. The observed endorsement rates of the DP group were 17% (*SE* = 4.2), 39% (*SE* = 4.2), 61% (*SE* = 3.6), and 86% (*SE* = 1.7), respectively. The regression on observed endorsement rates yielded a slope of 1.04 (*SE* = 0.03); the predicted slope of 1.0 falls well within the 95% confidence interval around the observed slope.[1] A similarity-based strategy thus accounts very well for classification of the Categorical and Other aliens in the DP group. Note that a similarity strategy could be based on similarity to the Mog and Nib prototypes, or to the exemplars from the training phase identified as Mogs and Nibs. These variants of similarity strategy are difficult to distinguish, and we do not attempt to do so.

---

[1] This regression model forced the *y*-intercept to be 0. When the *y*-intercept was allowed to vary, it went to 2, and the slope was 1.13 (*SE* = 0.08). The predicted slope of 1.0 continued to fall within the 95% confidence interval around the observed slope.

The simplest similarity model is not quite adequate to account for the behavior of the DP participants in the Conflicting conditions. Classifications based on similarity to the Mog and Nib prototypes would predict 80% accuracy for far-boundary Conflictings and 60% for near-boundary Conflictings; these are substantially better than the observed accuracies of 72% and 42%, respectively. The lower-than-predicted accuracy in these conditions is due to a high proportion of "Other" classifications in the DP group (21.5% for far and 36.2% for near Conflictings), despite the fact that these stimuli contained no Other features. Errors in labeling the Conflictings as "other" were much more likely in this group than in either the rule or single-feature groups, $F(1, 28) = 17.9$, $p < .0005$, and $F(1, 18) = 13.7$, $p < .002$, respectively. When confronted with a stimulus that was somewhat similar to both prototypes (and to exemplars of those categories), the strategy of the DP group frequently led them to give up and label the stimulus as belonging to neither category. A similarity-based strategy, combined with a "choke factor" when that strategy produced two solutions, is thus the most plausible account of the behavior of the DP participants.

Event-Related Potential Results

Overview of the Analysis Plan

We analyzed the ERP results across three groups of participants: the rule-instructed group, the DP subgroup of feedback-only participants, and the single-feature subgroup. Because the current paradigm has no precedent in the human electrophysiological literature, we had no specific predictions about the exact time windows during which categorization effects would occur. Visual inspection of the waveforms revealed two time windows of interest for both between-subjects and within-subject differences. The first was 200–600 ms poststimulus onset, encompassing a negative potential peaking at ~400 ms at most scalp sites. In this early time window, condition, distance from category boundaries, and feature frequency were all influential. The second latency window of interest was 800–1,300 ms, a latency range dominated by a positive potential most evident over frontal scalp that varied little by condition but that was sensitive to categorization strategy. Only ERPs elicited during trials with correct responses were analyzed.

Mean amplitude measurements of both time windows were made with respect to the 200 ms prestimulus baseline in each condition. For both time windows, our general analytic procedure was to first perform an ANOVA on the midline electrodes (Fpz, Fz, Fcz, Cz, Cpz, Pz, Oz) with group (rule, single-feature, DP) as a between-subjects factor and condition (Categorical, Other, Conflicting), boundary, and anterior/posterior (AP; 7 levels) as repeated measures. We then used the same ANOVA factors together with a left–right factor in two analyses of the lateral scalp sites. *Dorsal* refers to sites closer to the midline (Fp1, Fp2, F3, F4, Fc3, Fc4, C3, C4, P3, P4, O1, O2); *ventral* refers to sites farther from the midline, largely over the temporal lobes (F7, F8, Ft7, Ft8, T3, T4, Tp3, Tp4, T5, T6).

To evaluate the onset latency of particular experimental effects, we measured mean amplitudes in briefer time windows at selected scalp sites as described below.

Early Epoch (200–600 ms)

Familiar versus New. Because the ERPs to New and Familiar stimuli appeared almost identical, we compared only these two conditions in a preliminary analysis. No significant main effect of Familiar–New or interactions with group or boundary were observed, so the Familiar and New conditions were collapsed into a single Categorical condition as in the behavioral analyses.

Condition. Figure 9 shows that, for the rule group, Other stimuli elicited the most negative ERPs, followed by Categorical, followed by Conflicting. A similar gradient was evident in both feedback-only subgroups, although condition differences were notably absent at the prefrontal sites in these groups. All three electrode chains (midline, dorsal, and ventral) showed a main effect of condition, $Fs(2, 74) > 17.0$, $ps < .01$, $\varepsilon s > 0.9$.[2] In paired comparisons, ERPs to the Other stimuli were significantly more negative than the ERPs to the Categorical stimuli, $Fs(1, 37) > 13.2$, $ps < .01$, which were in turn more negative than the Conflicting condition, $Fs(1, 37) > 8.65$, $ps < .01$.

Additionally, the midline channels showed a weak Condition × AP interaction, $F(12, 444) = 2.44$, $p < .05$, $\varepsilon = 0.35$. Separate analyses of the three groups showed that this interaction was lacking in the rule and DP participants ($Fs < 1.10$) and marginal in the single-feature subgroup, $F(12, 108) = 1.95$, $p = .10$, $\varepsilon = .44$, and thus reflects the near absence of prefrontal condition differences in the single-feature subgroup and, perhaps, a somewhat weaker overall effect of condition in the distributed-profile subgroup.

Feature frequency. The general pattern of results across conditions shown above—Conflictings most positive, Categoricals next, and Others least positive—is consistent with an interpretation in terms of familiarity with the composite features of the stimuli. Recall that Conflictings are composed of five high-frequency features, whereas Others possess only one or two high-frequency features. To examine the impact of feature frequency, we ordered the stimulus types according to the five levels of average feature frequency shown in Table 2. Across most of the scalp, the ERPs showed a gradient of positivity corresponding to the average frequency of the features making up the aliens, as seen in Figure 10.

We observed main effects of feature frequency in the midline, dorsal, and ventral electrode chains, without significant interactions for group or scalp location factors: midline, $F(4, 148) = 8.98$, $p < .0001$, $\varepsilon = .90$; dorsal, $F(4, 148) = 9.67$, $p < .0001$, $\varepsilon = .88$; ventral, $F(4, 148) = 12.2$, $p < .01$, $\varepsilon = .89$. More importantly, decomposition of the polynomial trends showed that the feature frequency effect was largely linear, meaning that the variation in ERP amplitudes tracked the average feature frequency across conditions, $F_{linear}(1, 37) > 23.7$, $p < .0001$ for all three electrode chains; Fs for quadratic and higher order components < 1.

At posterior scalp sites, the feature frequency effect began quite early after stimulus onset. Onset latency of the feature frequency effect was evaluated via measurements of 50-ms latency windows,

---

[2] For analyses with more than degree of freedom in the numerator, the Huyhn-Feldt correction for nonsphericity of variance is used. Reported are the original degrees of freedom, the epsilon (ε) correction factor, and the corrected probability level.
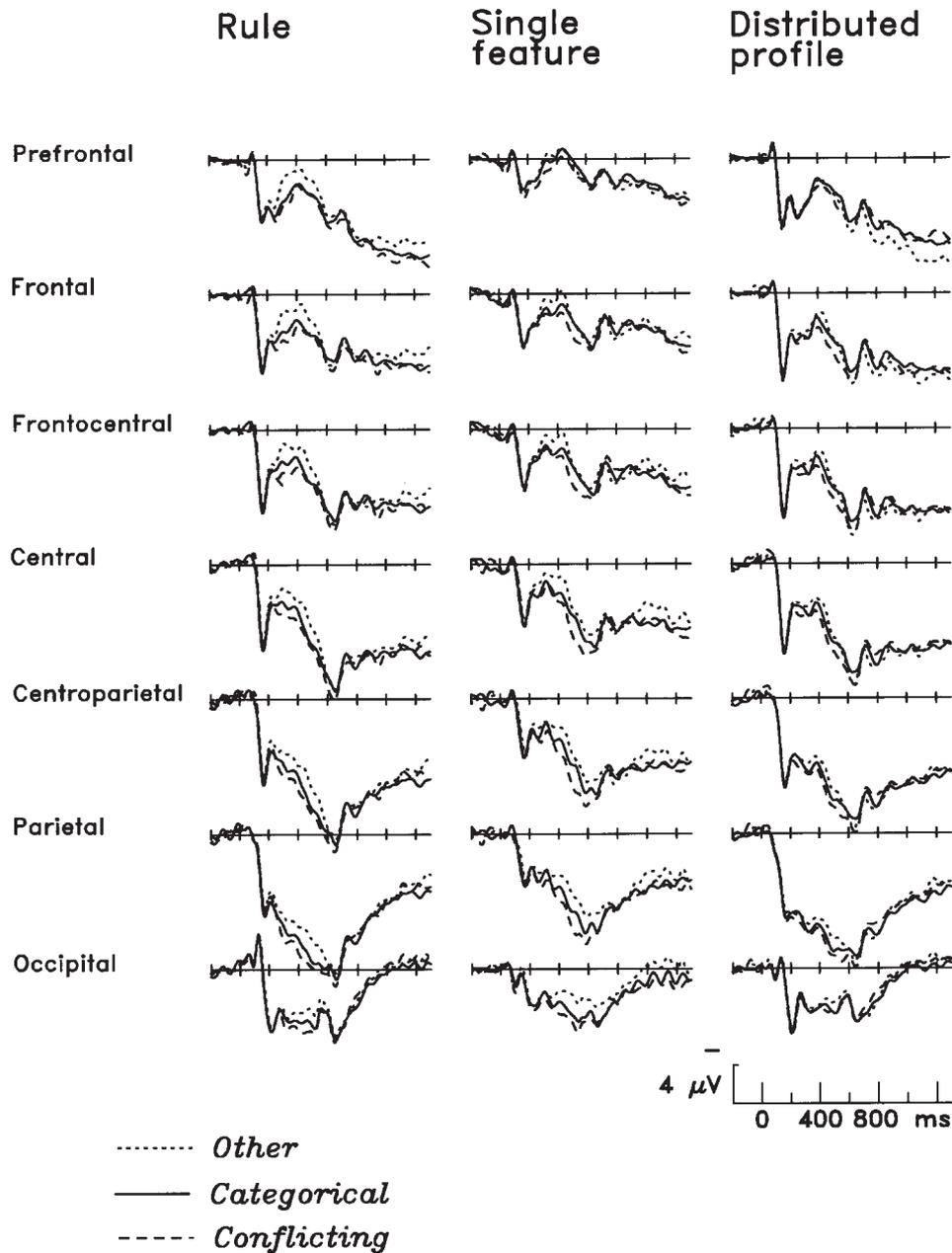
Figure 9. Grand average event-related potentials from midline scalp sites. Electrode sites are arranged from anterior (top) to posterior (bottom). Time 0 marks stimulus onset.

beginning at stimulus onset at parietal and occipital scalp sites (Pz, P3, P4, Oz, O1, O2). The linear gradient first became significant in the 150–200 ms latency window, $F_{linear}(1, 37) = 8.01, p < .01$; no interactions with Group or Condition.

Distance from category boundary. Figure 11 shows that near-boundary stimuli elicited a larger negative peak than far-boundary stimuli, but only in the rule group, and that this boundary effect had a prefrontal scalp distribution. For the rule group, Boundary × AP interactions were thus significant for the midline and ventral electrode chains: midline, $F(6, 114) = 3.85, p < .05, \varepsilon = 0.27$;

ventral, $F(4, 76) = 4.29, p = .05, \varepsilon = 0.29$, and approached significance in the dorsal chain, $F(5, 95) = 3.30, p = .08, \varepsilon = 0.27$. The boundary effect did not interact with condition in the rule group. There was no significant main effect of boundary, or Boundary × AP interactions for either the DP or single-feature subgroups, or for the feedback-only group considered as a whole.

Onset latency of the boundary effect was evaluated via measurements of 50-ms latency windows at the six prefrontal and frontal scalp sites where it was largest (Fpz, Fp1, Fp2, Fz, F3, F4). In the rule group, the boundary effect first became significant in

the 200–250 ms latency range, $F(1, 19) = 4.71$, $p < .05$; earlier window of 150–200 ms, $F < 1$.

Relationship between the boundary effect and other effects in the early latency window. The boundary effect is orthogonal to the feature frequency effect, as far-boundary Others have higher feature frequencies than do near-boundary Others, but the reverse is true for Categorical stimuli (and near and far Conflictings have identical feature frequencies). A second dissociation is that the boundary effect was evident only in the rule group, whereas the feature frequency effect was equivalent in the rule and feedback-only groups. Finally, the boundary effect had a clear prefrontal scalp distribution, whereas the feature frequency effect was widespread across the scalp.

The latency, scalp distribution, and general appearance of the boundary effect resembles the prefrontal difference between the Other aliens and both the Categorical and Conflicting conditions observed only in the rule group (see Figure 9). We thus considered the hypothesis that a single factor accounts for both early prefrontal effects in the rule group. By design, near-boundary stimuli contain more features that suggest the incorrect response than do far-boundary stimuli. If participants form associations between individual stimulus features and categorization responses, responding correctly to near-boundary stimuli will require overriding such stimulus-response associations to a greater degree than will responding correctly to far-boundary stimuli. In the General Discussion section, we thus suggest that the boundary effect observed
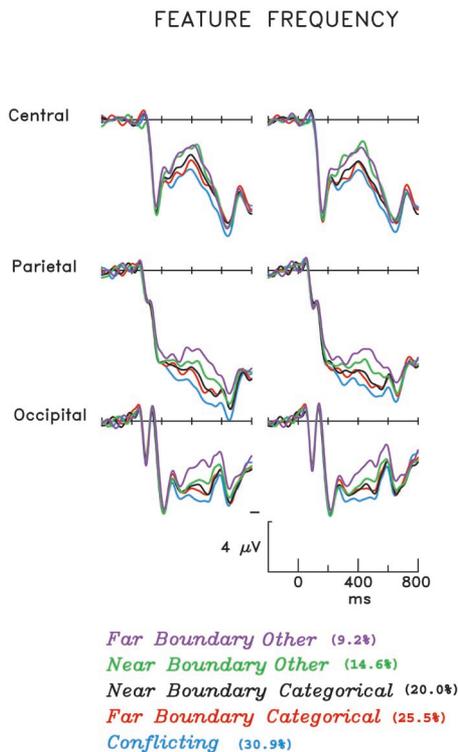
FEATURE FREQUENCY

Far Boundary Other (9.2%)
Near Boundary Other (14.6%)
Near Boundary Categorical (20.0%)
Far Boundary Categorical (25.5%)
Conflicting (30.9%)

Figure 10. Grand average event-related potentials from three left scalp sites in the left column (C3, P3, O1) and three right scalp sites in the right column (C4, P4, O2). The five conditions shown were defined by the frequency with which their component features occurred across the entire experiment (averaged across the five diagnostic features of each stimulus).
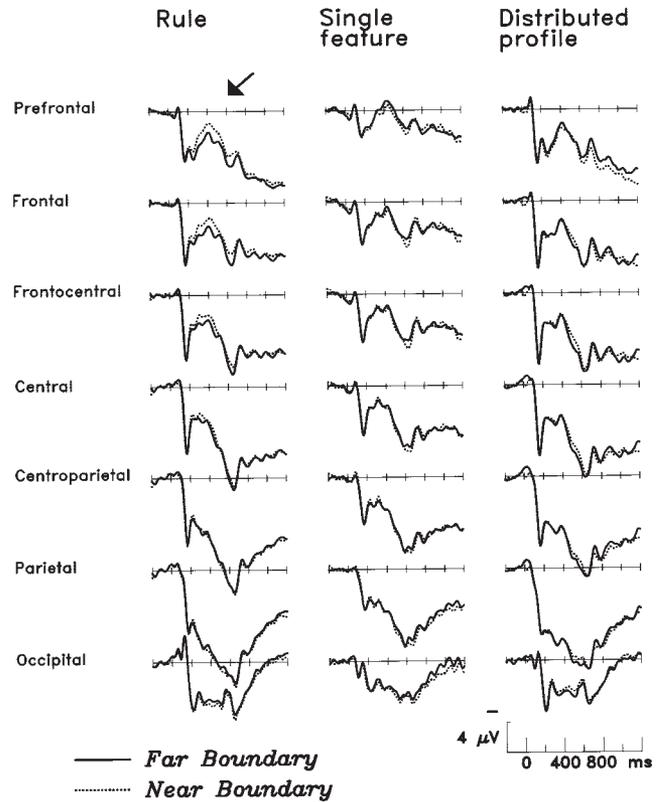

Far Boundary
Near Boundary

Figure 11. Grand average event-related potentials from midline scalp sites.

here is a no-go N2, an ERP component previously linked to response inhibition. However, the relative frequency of the features associated with correct and incorrect responses varies across condition as well as boundary. Other aliens, for example, have frequent "incorrect" features (Mog and Nib features) and infrequent "correct" features, Categorical aliens have the opposite pattern, and Conflicting aliens have frequent correct and incorrect features. It is possible that the additional prefrontal negativity elicited by Others as compared with Categoricals and Conflictings and by near as compared with far-boundary stimuli may both reflect the relative strength of features associated with correct versus incorrect responses. To evaluate this hypothesis, we rank ordered all six conditions by a new metric: (frequency of features associated with correct response) minus (frequency of features associated with incorrect response).[3] On this metric, both near- and

<hr>

[3] For example, near Conflictings contain three features presented on 31% of all trials, and are associated with the correct response of "Mog" or "Nib" for a sum of 93% correct features. The other two diagnostic features are also present on 31% of the total trials, but are associated with an incorrect response; the correct-minus-incorrect difference is thus 31% (greater strength for correct than for incorrect features). For near Others, the difference score is (3.8% $\times$ 3) minus (31% $\times$ 2) = -50.6% (greater strength for incorrect than for correct features). We also examined a ratio measure of correct feature strength divided by incorrect feature strength but found that this accounted for less of the variance in the amplitude of the prefrontal ERP component.

far-boundary Others have negative scores (greater strength of incorrect than correct features), whereas the other four conditions have a range of positive scores, but the scores are always lower for near- than they are for far-boundary. We used an ANOVA with trend analysis to determine whether this metric was linearly related to the amplitude of the early negative component at prefrontal and frontal sites (200–600 ms, Fpz, Fp1, Fp2, Fz, F3, F4). The linear trend across conditions was significant in the rule group, $F(1, 19) = 20.6$, $p < .0005$, and accounted for 63% of the variance associated with the six conditions.

Late Epoch (800–1,300 ms)

Figures 9–11 indicate that ERP differences among stimulus types are mostly over and done by 800 ms poststimulus onset. Figure 12, however, indicates that differences among the three strategy groups persist into this time window. The rule and DP participants had more positive ERPs relative to the single-feature group; the group difference was broadly distributed over the front half of the head but was largest at prefrontal sites.

In comparisons between the rule group and single-feature subgroup, both the dorsal and ventral chains yielded main effects of group: dorsal, $F(1, 28) = 4.11$, $p = .05$; ventral, $F(1, 28) = 5.98$, $p < .05$. The frontal distribution of the strategy effect led to weak Group × AP interactions for the midline and ventral chains: midline, $F(6, 168) = 2.42$, $p < .10$, $\varepsilon = .38$; dorsal, $F(5, 140) = 2.53$, $p < .10$, $\varepsilon = .33$. No Group × Condition interactions were observed, indicating that the impact of strategy was global. Follow-up analyses showed that the group difference was significant at the three prefrontal sites where it was largest, $F(1, 28) = 6.20$, $p < .05$. The top half of Figure 13 shows the frontal distribution of this strategy effect.

Comparisons between the DP and single-feature subgroups yielded similar results. The frontal distribution of the strategy effect yielded a Group × AP interaction for the midline sites, $F(6, 108) = 2.71$, $p = .05$, $\varepsilon = .48$, and a trend toward such an interaction for the dorsal sites, $F(5, 90) = 2.43$, $p = .10$, $\varepsilon = .46$. Follow-up analyses showed that the group difference was significant at the three prefrontal sites where it was largest, $F(1, 18) = 4.80$, $p < .05$.

Comparisons between the rule group and DP subgroup yielded null effects of group for the midline, dorsal, and ventral sites. Group × Condition interactions were evident: dorsal, $F(3, 84) = 3.98$, $p < .05$, $\varepsilon = .97$; ventral, $F(3, 84) = 3.44$, $p < .05$, $\varepsilon = .94$. Visual inspection of the data suggested that these were due to small group differences in the Categorical and Conflicting conditions that were not evident in the Other condition. However, pursuit of the interactions via contrasts of the two groups in each condition yielded no reliable effects. There may indeed be subtle differences between the late ERPs of the rule and DP strategies, but these could not be clearly characterized in the present between-subjects design.

General Discussion

What's Special About Rule Use?

The late frontal effect and evaluation of multiple stimulus dimensions.   Two distinct ERP effects differentiated participants using different categorization strategies, both with frontal scalp distributions. All stimulus types elicited a larger frontally distributed positivity in participants using either a multidimensional rule or a multidimensional similarity strategy, as compared with those relying largely on a unidimensional rule. The group difference indexes a late processing stage (800–1,300 ms after stimulus onset) but was evident some 700 to 1,000 ms prior to average reaction time in the multidimensional strategies. The pattern of results indicates that rule use per se is not the relevant factor driving this brain activity. Instead, the late frontal effect reflects the evaluation of multiple stimulus features, whether this assessment is in the service of a rule- or a similarity-based strategy.

Although the ability to shift from one unidimensional rule to another has been well studied and associated with the frontal lobe, we are aware of no studies evaluating frontal brain activity during the continued application of unidimensional versus multidimensional categorization rules. The frontal differentiation observed here is consistent with the argument of Holyoak and colleagues that it is problems requiring consideration of multiple dimensions that require prefrontal involvement, whereas unidimensional problems do not (Robin & Holyoak, 1995). The categorization task applied here is different from the matrix problems studied by this group of investigators but converges with their conclusion (Christoff et al., 2001; Kroger et al., 2002; Waltz et al., 1999). One advantage of the categorization paradigm used here was that it allowed contrasts among three strategies and thus revealed a commonality between multidimensional rule use and assessment of multiple features independent of rule application per se.

Consideration of more features before making a decision can be construed as a working memory manipulation—we might imagine that both the rule and DP participants maintained more information in working memory than the single-feature group. However, numerous ERP studies have manipulated working memory load in a variety of paradigms, and these manipulations are generally associated with slow negative potentials rather than the positive potential observed here (King & Kutas, 1995; Mecklinger & Pfeifer, 1996; Mecklinger, Schriefers, Steinhauer, & Friederici, 1995; Münte, Schiltz, & Kutas, 1998; Roeder, Rösler, & Hennighausen, 1997; Ruchkin, Johnson, Grafman, Canoune, & Ritter, 1992, 1997).

In contrast with working memory experiments, frontal positive potentials similar to those observed here have appeared in source and associative memory tasks, which tap long term memory. In source memory experiments, prefrontal positivities in the 800–1,200-ms range are enhanced when participants must judge whether a test stimulus contains the same conjunction of attributes as a studied stimulus (e.g., a word plus the speaker's voice, or a drawing plus its spatial location), as compared with basing a recognition judgment on only a single attribute (Senkfor & Van Petten, 1998; Van Petten, Senkfor, & Newberg, 2000). In a recent associative memory experiment, participants who were better able to distinguish studied word pairs from recombinations of studied words had larger prefrontal positivities than participants who frequently failed to distinguish between conjunctions and individual elements retrieved from long term memory (Van Petten, Luka, Rubin, & Ryan, 2002). Late prefrontal positivities in these recognition experiments thus dissociate tasks that can be performed with attention to one stimulus element from those that require attention to multiple elements and their combinations.
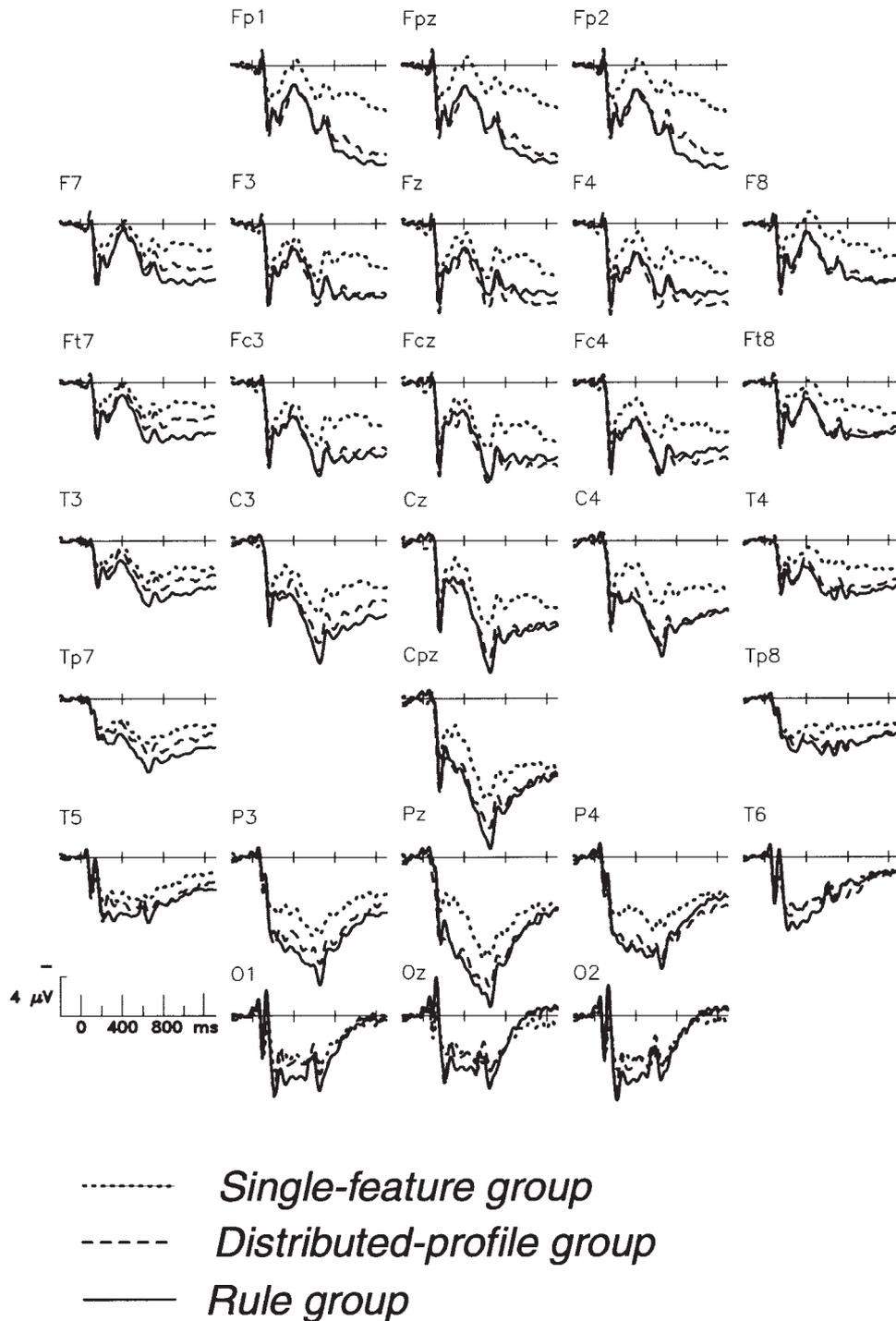
Figure 12. Grand average event-related potentials from all scalp sites. Sites are arranged from anterior (top) to posterior (bottom), and from left to right, roughly corresponding to head location.

On the basis of the striking similarity between the ERP components they elicit, we speculate that there might be some functional similarity between source memory and multidimensional categorization tasks. When categorization is based on multiple features, a similar process of comparing a current stimulus with information retrieved from long term memory may be required. When categorization is based on similarity, the fit between a current stimulus and a categorical prototype or multiple previous exemplars will need to be assessed, rather than the fit with a single episodic trace, as in the source memory tasks. When a rule strategy
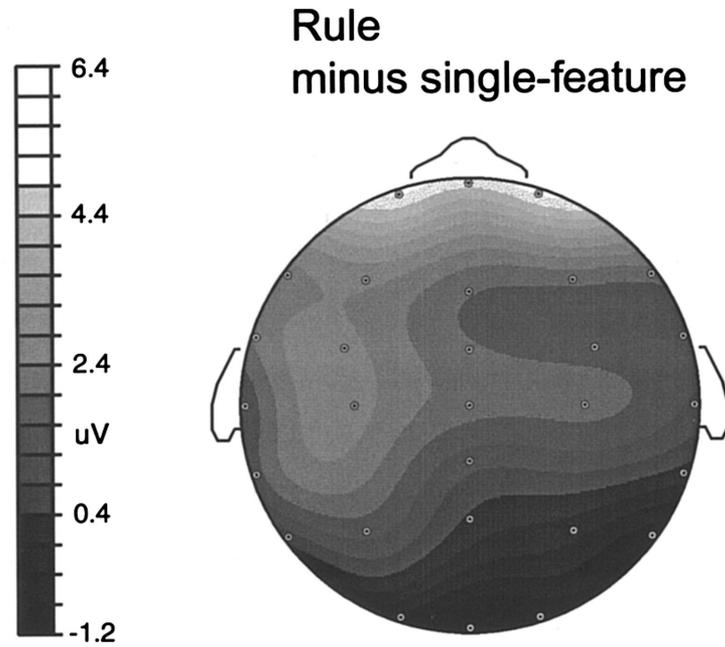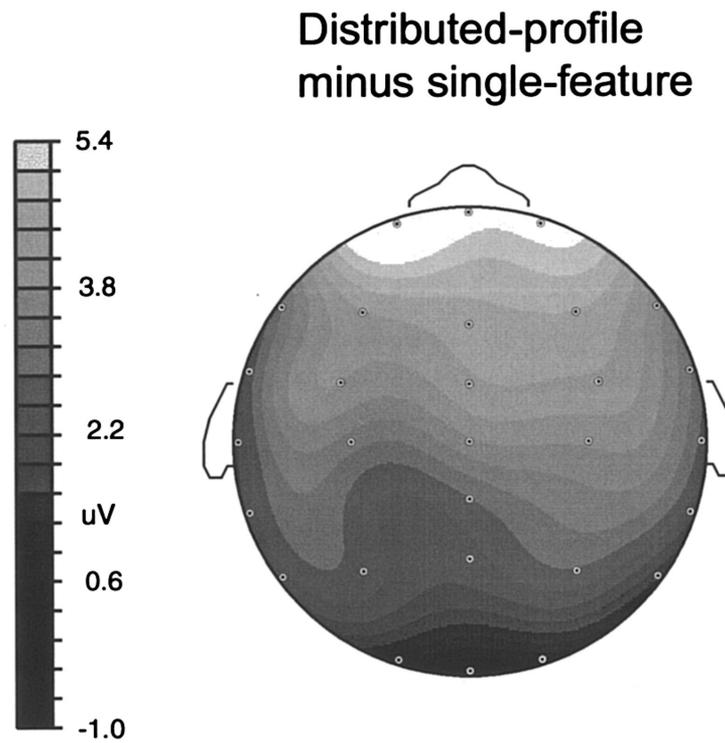
Figure 13. Topographic map of group differences in event-related potential amplitudes in the late time window (800–1,300 ms).

is used, the matching process may operate on somewhat different content, including memory for the verbal portion of the rule ("three out of five") in addition to visual memories of what the features of various categories look like. There were subtle differences between the late ERPs of participants using multidimensional-rule versus similarity strategies, but the between-subjects design was not sufficiently sensitive to specify these in any detail.

The early frontal effect: Inhibition of similarity-based responses? An additional ERP effect with a prefrontal scalp focus was observed only in the rule-instructed participants. A negative component peaking at ~400 ms was larger for stimuli closer to the boundaries separating the three categories than for stimuli farther from the categorical boundaries. As described in the Event-Related Potential Results section, the amplitude of this component was sensitive to the relative strength of stimulus features suggesting an incorrect response and those suggesting a correct response. This prefrontal N2 effect began more than 1 s before the overt categorization response in the rule participants, and is thus likely to reflect only very preliminary analyses of the stimuli. It is important to note that although all three groups of participants showed a boundary effect in their reaction times, only the rule group showed a boundary effect in the N2. Understanding the cognitive process underlying this ERP effect may thus offer some insight into what differentiates multidimensional rule use from the other strategies.

The polarity, scalp topography, and general morphology of the early frontal effect appear identical to the no-go N2 component of the ERP, associated with inhibition of both overt and covert responses and generated in prefrontal cortex (Kopp et al., 1996; Pfefferbaum, Ford, Weller, & Kopell, 1985; Sasaki & Gemba, 1986; Sasaki, Gemba, Nambu, & Matsuzaki, 1993; Schmitt, Münte, & Kutas, 2000; Schmitt, Schiltz, Zaake, Kutas, & Münte, 2001; Simson, Vaughn, & Ritter, 1977; van Boxtel, van der Molen, Jennings, & Brunia, 2001). In the choice-response paradigm used here, each stimulus included features associated with two different responses. We suggest that initial detection of one or two familiar features leads to preparation of the response associated with that category. If features associated with an alternate response are detected as stimulus processing continues, the early response preparation may be inhibited until the assignment of features to categories is complete and the analyzed feature set can be more deliberately compared with information from long term memory. Because near-boundary stimuli included a more equal balance between features associated with different categories (three and two) than far-boundary stimuli (four and one), the probability of encountering equivalent but conflicting evidence for two categories would be higher in the near-boundary case.

Interpreting the N2 boundary effect as response inhibition suggests an intermediate-level mechanism for applying a multidimensional rule, a mechanism absent in similarity strategies. We suggest that one concomitant of following a rule that classifies items into multiple categories is competition between the responses associated with those categories early in processing—evidence for one category is treated as evidence against a different category. Interruption of response preparation based on preliminary stimulus analyses may serve as a brake on incorrect responses but will also result in slower responses when using a rule. In future experimental designs, this interpretation can be tested by crossing this sort of response conflict with feature familiarity. Both factors should

influence the reaction times of rule users, but only the former should result in an N2 effect.

Whereas the similarity (DP) participants also accumulated evidence for multiple categories on each trial (or they would not have been disproportionately inaccurate on the Mog–Nib mixtures in the Conflicting condition), they showed no N2 boundary effect. We thus suggest that their similarity strategy included little in the way of early competition between response alternatives. Instead, evidence indicative of different categories was simply compared with the long term memory representations of Mogs and Nibs after feature analysis was complete (this latter process was reflected in the late frontal positivity). Because there was no prototype for Others as a category distinct from Mogs or Nibs, and no check on the bias to respond on the basis of Mog–Nib features, this late–comparison process would allow even weakly similar Others to be accepted as Mogs or Nibs.

Recognition Memory During the Categorization Task

Across the accuracy, reaction time, and ERP measures, no differences between New and Familiar exemplars emerged in any of the three strategy groups, although Familiar aliens shared all five diagnostic features with those studied in the training phase. The present paradigm discouraged a pure memorization strategy with the use of a large number of stimuli. In previous categorization studies in which participants were faster or more accurate for studied than for new exemplars, researchers used smaller stimulus sets for which exemplar memorization was a viable strategy that conferred an advantage to studied exemplars (Allen & Brooks, 1991; Regehr & Brooks, 1993). The null effects here indicate that successful categorization need not depend on retrieval of whole exemplars, even when a similarity strategy is used.

In all three strategy groups, evidence for recognition of component features was observed: Stimuli with more frequently presented features elicited more positive ERPs in the 200–600 ms latency range than stimuli with less frequent features. The broadly distributed positive potential showed an amplitude gradient that tracked the summed frequency of the features comprising a given alien. In latency, morphology, and scalp distribution, the feature frequency effect greatly resembles the old–new effect observed in prior studies using direct recognition tests and in prior studies examining incidental-repetition effects (see Friedman & Johnson, 2000; Rugg, 1995; Van Petten & Senkfor, 1996, for reviews). In both sorts of paradigms, the old–new effect is reduced in patients with damage to medial temporal lobe structures (Helmstaedter, Grunwald, Lehnertz, Gleibner, & Elger, 1997; Mecklinger, von Cramon, & Matthes-von Cramon, 1998; Olichney et al., 2000; M. E. Smith & Halgren, 1989).

Overall, the feature frequency effect suggests that our participants were well able to recognize frequently encountered stimulus features. Indeed, it would be surprising if individuals with normal memory abilities did not recognize features presented some 110 times over a 3-hr period. What cannot be answered on the basis of this study is whether declarative memory for features was necessary for categorization, or epiphenomenal. It is possible that our healthy participants formed and retrieved conscious memories for stimulus features but that these memories did not contribute to their categorization judgments, as argued by Reed et al. (1999) on the basis of their results from amnesic patients. However, as we

note in the introduction, Reed et al.'s test of declarative memory for stimulus features was more stringent than their measure of classification accuracy. Comparison of the poor declarative memory to the good categorization performance in that study was hindered by a change from recognition to recall, a change from visual presentation to oral–verbal cues, different methods of scoring accuracy, and a time delay. In contrast, the feature-memory effect observed here was triggered by the exact same stimuli that elicited categorization judgments. Establishing the causal relationship (or lack thereof) between feature memory and categorization performance may require using a similarly immediate measure of memory in individuals with a memory impairment.

However, the temporal progression of brain activity across the categorization trials is consistent with the view that recognition of the component features of the stimuli was a necessary first step. The earliest impact of experimental manipulation present in the data was the feature frequency effect, beginning 150 ms poststimulus onset at posterior scalp sites. This was followed not much later by the prefrontal boundary effect in the rule group (beginning at 200 ms), which we hypothesize follows from the initial detection of conflicting evidence for more than one category. At the onset of the N2 around 200 ms poststimulus onset, assignment of features into their associated categories is likely not complete—only two features from different categories need be identified as such to generate a response conflict. Moreover, both the prefrontal N2 and the feature frequency effect were some 500 ms in duration and overlapped in time, suggesting that memory retrieval and association of features with categories are cascaded processes. Finally, the ERPs of those participants who used a multidimensional strategy were marked by a larger frontally distributed positivity than were the ERPs of the unidimensional-rule group, beginning around 800 ms poststimulus onset and continuing to at least 1,300 ms. We suggest that this late group difference reflects the comparison of the now-analyzed stimulus with information from long term memory—categorical prototypes, multiple exemplars, and/or the verbal portion of the rule. The participants using a unidimensional rule were likely to skip, or greatly truncate, this last step, so that essentially all of their behavioral responses occurred prior to 1,300 ms. In contrast, this final comparison between multidimensional stimuli and multidimensional memory representations prolonged the RTs of the rule and DP participants, whose earliest behavioral responses very rarely occurred before 1,300 ms.

References

Aizenstein, H. J., MacDonald, A. W., Stenger, V. A., Nebes, R. D., Larson, J. K., Ursu, S., & Carter, C. S. (2000). Complementary category learning systems identified with event-related functional MRI. *Journal of Cognitive Neuroscience, 12,* 977–987.

Allen, S. W., & Brooks, L. R. (1991). Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General, 120,* 3–19.

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

Buckner, R. L., Koutstaal, W., Schacter, D. L., Dale, A. M., Rotte, M., & Rosen, B. R. (1998). Functional-anatomic study of episodic memory using fMRI. II. Selective averaging of fMRI trials to test the retrieval success hypothesis. *NeuroImage, 7,* 163–175.

Cabeza, R., & Nyberg, L. (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience, 12,* 1–47.

Christoff, K., Prabhakaran, V., Dorfman, J., Zhao, Z., Kroger, J. K., Holyoak, K. J., & Gabrieli, J. D. E. (2001). Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *NeuroImage, 14,* 1136–1149.

D'Esposito, M., Postle, B. R., & Rypma, B. (2000). Prefrontal cortical contributions to working memory: Evidence from event-related fMRI studies. *Experimental Brain Research, 133,* 3–11.

Filoteo, J. V., Maddox, W. T., & Davis, J. D. (2001). Quantitative modeling of category learning in amnesic patients. *Journal of the International Neuropsychological Society, 7,* 1–19.

Friedman, D., & Johnson, R., Jr. (2000). Event-related potential (ERP) studies of memory encoding and retrieval: A selective review. *Microscopy Research Techniques, 51,* 6–28.

Goel, V., & Dolan, R. J. (2000). Anatomical segregation of component processes in an inductive inference task. *Journal of Cognitive Neuroscience, 12,* 110–119.

Hahn, U., & Chater, N. (1998). Similarity and rules: Distinct? Exhaustive? Empirically distinguishable? *Cognition, 65,* 197–230.

Heinze, H.-J., Münte, T. F., & Kutas, M. (1998). Context effects in a category verification task as assessed by event-related brain potential (ERP) measures. *Biological Psychology, 47,* 121–135.

Helmstaedter, C., Grunwald, T., Lehnertz, K., Gleibner, U., & Elger, C. E. (1997). Differential involvement of left temporolateral and temporomesial structures in verbal declarative learning and memory: Evidence from temporal lobe epilepsy. *Brain and Cognition, 35,* 110–131.

Henson, R. N., Goshen-Gottstein, Y., Ganel, T., Otten, L. J., Quayle, A., & Rugg, M. D. (2003). Electrophysiological and haemodynamic correlates of face perception, recognition, and priming. *Cerebral Cortex, 13,* 793–805.

Henson, R. N. A., Rugg, M. D., Shallice, T., Josephs, O., & Dolan, R. J. (1999). Recollection and familiarity in recognition memory: An event-related functional magnetic resonance imaging study. *Journal of Neuroscience, 19,* 3962–3972.

King, J. W., & Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience, 7,* 376–395.

Kitchener, E. G., & Squire, L. R. (2000). Impaired verbal category learning in amnesia. *Behavioral Neuroscience, 114,* 907–911.

Knight, R. T., Staines, R. W., Swick, D., & Chao, L. L. (1999). Prefrontal cortex regulates inhibition and excitation in distributed neural networks. *Acta Psychologica, 101,* 159–178.

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

Kolodny, J. A. (1994). Memory processes in classification learning: An investigation of amnesic performance in categorization of dot patterns and artistic styles. *Psychological Science, 5,* 164–169.

Kopp, B., Rist, F., & Mattler, U. (1996). N200 in the flanker task as a neurobehavioral tool for investigating executive control. *Psychophysiology, 33,* 282–294.

Kroger, J. K., Sabb, F. W., Fales, C. L., Bookheimer, S. Y., Cohen, M. S., & Holyoak, K. J. (2002). Recruitment of anterior dorsolateral prefrontal cortex in human reasoning: A parametric study of relational complexity. *Cerebral Cortex, 12,* 477–485.

Lepage, M., Brodeur, M., & Bourgouin, P. (2003). Prefrontal cortex contribution to associative recognition memory in humans: An event-related functional magnetic resonance imaging study. *Neuroscience Letters, 346,* 73–76.

Lezak, M. D. (1995). *Neuropsychological assessment* (3rd ed.). New York: Oxford University Press.

McCloskey, M. E., & Glucksberg, S. (1979). Decision processes in verifying category membership statements: Implications for models of semantic memory. *Cognitive Psychology, 11,* 1–37.

Mecklinger, A., & Pfeifer, E. (1996). Event-related potentials reveal topo-

graphical and temporal distinct neuronal activation patterns for spatial and object working memory. *Cognitive Brain Research, 4,* 211–224.

Mecklinger, A., Schriefers, H., Steinhauer, K., & Friederici, A. D. (1995). Processing relative clauses varying on syntactic and semantic dimensions: An analysis with event-related potentials. *Memory & Cognition, 23,* 477–494.

Mecklinger, A., von Cramon, D. Y., & Matthes-von Cramon, G. (1998). Event-related potential evidence for a specific recognition memory deficit in adult survivors of cerebral hypoxia. *Brain, 121,* 1919–1935.

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

Milner, B. (1963). Effects of different brain lesions on card sorting. *Archives of Neurology, 9,* 90–100.

Münte, T. F., Schiltz, K., & Kutas, M. (1998, September 3). When temporal terms belie conceptual order. *Nature, 395,* 71–73.

Nagahama, Y., Okada, T., Katsumi, Y., Hayashi, T., Yamauchi, H., Sawamoto, N., et al. (1999). Transient neural activity in the medial superior frontal gyrus and precuneus time locked with attention shift between object features. *NeuroImage, 10,* 193–199.

Neville, H. J., Kutas, M., Chesney, G., & Schmidt, A. (1986). Event-related brain potentials during the initial encoding and subsequent recognition memory of congruous and incongruous words. *Journal of Memory and Language, 25,* 75–92.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General, 115,* 39–57.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review, 101,* 53–79.

Nyberg, L. (2000). Can the brain tell the difference between old and new? In E. Tulving (Ed.), *Memory, consciousness, and the brain* (pp. 293–306). Philadelphia: Psychology Press.

Olichney, J., Van Petten, C., Paller, K., Salmon, D., Iragui, V., & Kutas, M. (2000). Word repetition in amnesia: Electrophysiological evidence of spared and impaired memory. *Brain, 123,* 1948–1963.

Paller, K. A., & Kutas, M. (1992). Brain potentials during memory retrieval provide neurophysiological support for the distinction between conscious recollection and priming. *Journal of Cognitive Neuroscience, 4,* 375–391.

Paller, K. A., Kutas, M., & McIsaac, H. (1995). Monitoring conscious recollection via the electrical activity of the brain. *Psychological Science, 6,* 107–111.

Patalano, A. L., Smith, E. E., Jonides, J., & Koeppe, R. A. (2001). PET evidence for multiple strategies of categorization. *Cognitive, Affective, and Behavioral Neuroscience, 1,* 360–370.

Pfefferbaum, A., Ford, J. M., Weller, B. J., & Kopell, B. S. (1985). ERPs to response production and inhibition. *Electroencephalography and Clinical Neurophysiology, 60,* 423–434.

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

Reber, P. J., Stark, C. E., & Squire, L. R. (1998b). Cortical areas supporting category learning identified using functional MRI. *Proceedings of the National Academy of Sciences of the United States of America, 95,* 747–750.

Reber, P. J., Wong, E. C., & Buxton, R. B. (2002). Comparing the brain areas supporting nondeclarative categorization and recognition memory. *Cognitive Brain Research, 14,* 245–257.

Reed, J. M., Squire, L. R., Patalano, A. L., Smith, E. E., & Jonides, J. (1999). Learning about categories that are defined by object-like stimuli despite impaired declarative memory. *Behavioral Neuroscience, 113,* 411–419.

Regehr, G., & Brooks, L. R. (1993). Perceptual manifestations of an

analytic structure: The priority of holistic individuation. *Journal of Experimental Psychology: General, 122,* 92–114.

Robin, N., & Holyoak, K. J. (1995). Relational complexity and the functions of prefrontal cortex. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 987–997). Cambridge, MA: MIT Press.

Roeder, B., Rösler, F., & Hennighausen, E. (1997). Different cortical activation patterns in blind and sighted humans during encoding and transformation of haptic images. *Psychophysiology, 34,* 292–307.

Rogers, R. D., Andrews, T. C., Grasby, P. M., Brooks, D. J., & Robbins, T. W. (2000). Contrasting cortical and subcortical activations produced by attentional-set shifting and reversal learning in humans. *Journal of Cognitive Neuroscience, 12,* 142–162.

Rubin, S. R., Van Petten, C., Glisky, E. L., & Newberg, W. M. (1999). Memory conjunction errors in younger and older adults: Event-related potential and neuropsychological evidence. *Cognitive Neuropsychology, 16,* 459–488.

Ruchkin, D. S., Johnson, R., Grafman, J., Canoune, H., & Ritter, W. (1992). Distinctions and similarities among working memory processes: An event-related potential study. *Cognitive Brain Research, 1,* 53–66.

Ruchkin, D. S., Johnson, R., Jr., Grafman, J., Canoune, H., & Ritter, W. (1997). Multiple visuospatial working memory buffers: Evidence from spatiotemporal patterns of brain activity. *Neuropsychologia, 35,* 195–209.

Rugg, M. D. (1995). ERP studies of memory. In M. D. Rugg & M. G. H. Coles (Eds.), *Electrophysiology of mind: Event-related brain potentials and cognition* (Vol. 15, pp. 132–170). New York: Oxford University Press.

Rugg, M. D., Fletcher, P. C., Chua, P. M.-L., & Dolan, R. J. (1999). The role of the prefrontal cortex in recognition memory and memory for source: An fMRI study. *NeuroImage, 10,* 520–529.

Sasaki, K., & Gemba, H. (1986). Electrical activity in the prefrontal cortex specific to no-go reaction of conditioned hand movement with colour discrimination in the monkey. *Experimental Brain Research, 64,* 603–606.

Sasaki, K., Gemba, H., Nambu, A., & Matsuzaki, R. (1993). No-go activity in the frontal association cortex of human subjects. *Neuroscience Research, 18,* 249–252.

Schmitt, B. M., Münte, T. F., & Kutas, M. (2000). Electrophysiological estimates of the time course of semantic and phonological encoding during implicit picture naming. *Psychophysiology, 37,* 473–484.

Schmitt, B. M., Schlitz, K., Zaake, W., Kutas, M., & Münte, T. F. (2001). An electrophysiological analysis of the time course of conceptual and syntactic encoding during tacit picture naming. *Journal of Cognitive Neuroscience, 13,* 510–522.

Senkfor, A. J., & Van Petten, C. (1998). Who said what? An event-related potential investigation of source and item memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 1005–1025.

Simson, R., Vaughan, H. G., & Ritter, W. (1977). The scalp topography of potentials in auditory and visual Go/NoGo tasks. *Electroencephalography and Clinical Neurophysiology, 43,* 864–875.

Smith, E. E., Patalano, A. L., & Jonides, J. (1998). Alternative strategies of categorization. *Cognition, 65,* 167–196.

Smith, E. E., & Sloman, S. A. (1994). Similarity- versus rule-based categorization. *Memory & Cognition, 22,* 377–386.

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, M. E., & Halgren, E. (1989). Dissociation of recognition memory components following temporal lobe lesions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 50–60.

Squire, L. R. (1987). *Memory and brain.* New York: Oxford University Press.

Squire, L. R., & Knowlton, B. J. (1995). Learning about categories in the

absence of memory. *Proceedings of the National Academy of Sciences of the United States of America, 92,* 12470–12474.

Strange, B. A., Henson, R. N. A., Friston, K. J., & Dolan, R. J. (2001). Anterior prefontal cortex mediates rule learning in humans. *Cerebral Cortex, 11,* 1040–1046.

Stuss, D. T., Leevien, B., Alexander, M. P., Hong, J., Palumbo, C., Hamer, L., et al. (2000). Wisconsin card sorting test performance in patients with focal frontal and posterior brain damage: Effects of lesion location and test structure on separable cognitive processes. *Neuropsychologia, 38,* 388–402.

van Boxtel, G. J. M., van der Molen, M. W., Jennings, R. J., & Brunia, C. H. M. (2001). A psychophysiological analysis of inhibitory motor control in the stop-signal paradigm. *Biological Psychology, 58,* 229–262.

Van Petten, C., Luka, B. J., Rubin, S. R., & Ryan, J. P. (2002). Frontal brain activity predicts individual performance in an associative memory exclusion task. *Cerebral Cortex, 12,* 1180–1192.

Van Petten, C., & Senkfor, A. J. (1996). Memory for words and novel visual patterns: Repetition, recognition, and encoding effects in the event-related brain potential. *Psychophysiology, 33,* 491–506.

Van Petten, C., Senkfor, A. J., & Newberg, W. M. (2000). Memory for drawings in locations: Spatial source memory and event-related potentials. *Psychophysiology, 37,* 551–564.

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.

Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., de Menezes Santos, M., et al. (1999). A system for relational reasoning in human prefrontal cortex. *Psychological Science, 10,* 119–125.

Ward, T. B., & Scott, J. (1987). Analytic and holistic modes of learning family resemblance concepts. *Memory & Cognition, 15,* 42–54.

Zaki, S. R., & Nosofsky, R. M. (2001). A single-system interpretation of dissociations between recognition and categorization in a task involving object-like stimuli. *Cognitive, Affective, and Behavioral Neuroscience, 1,* 344–359.

Zaki, S. R., Nosofsky, R. M., Jessup, N. M., & Unverzagt, F. W. (2003). Categorization and recognition performance of a memory-impaired group: Evidence for single-system models. *Journal of the International Neuropsychological Society, 9,* 394–406.