Activation in the neural network responsible for categorization and recognition reflects parameter changes

Robert M. Nosofsky[a,1], Daniel R. Little[b], and Thomas W. James[a]

[a]Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN 47405; and [b]Psychological Sciences, The University of Melbourne, Melbourne, Parkville VIC 3010, Australia

According to various influential formal models of cognition, perceptual categorization and old−new recognition recruit the same memory system. By contrast, the prevailing view in the cognitive neuroscience literature is that separate neural systems mediate perceptual categorization and recognition. A direct form of evidence is that separate brain regions are activated when observers engage in categorization and recognition tasks involving the same types of stimuli. However, even if the same memory-based comparison processes underlie categorization and recognition, one would not expect to see identical patterns of brain activity across the tasks; the reason is that observers would adjust parameter settings (e.g., vary criterion settings) across the tasks to satisfy the different task goals. In this fMRI study, we conducted categorization and recognition tasks in which stimulus conditions were held constant, and in which observers were induced to vary hypothesized parameter settings across conditions. A formal exemplar model was fitted to the data to track the changes in parameters to help interpret the fMRI results. We observed systematic effects of changes in parameters on patterns of brain activity, which were interpretable in terms of differing forms of evidence accumulation that resulted from the changed parameter settings. After controlling for stimulus and parameter-related differences, we found little evidence that categorization and recognition recruit separate memory systems.

Among the most fundamental cognitive processes are categorization and old−new recognition (1). In categorization, observers make decisions about whether distinct objects belong to the same class (2). By contrast, in recognition, observers judge whether each test object is an exact match to some study object. Despite the difference in task goals, various theories posit that categorization and recognition rely on similar memory representations and cognitive processes (3–7). For example, according to an influential exemplar model of categorization (5–7), which will serve as an analytic tool in the present research, people store individual exemplars in memory, and make categorization and recognition decisions on the basis of how similar test objects are to the exemplars. The model has provided excellent quantitative accounts of categorization and recognition choice probabilities and response times in numerous experiments (7–10).

However, this single-memory system exemplar model has also been challenged on grounds of various dissociations observed between categorization and recognition (11). For example, various influential studies demonstrated that amnesic subjects with poor recognition memory show normal categorization performance, suggesting that separate neural memory systems may mediate the tasks (12–16). Nevertheless, formal modeling has indicated that even these dissociations are consistent with the predictions from the exemplar model (4, 17–19). The general approach in the modeling was to assume that amnesics have reduced ability to discriminate among distinct exemplar traces. This reduced memorial discriminability is particularly detrimental to recognition, which may require the observer to make fine-grained distinctions between old vs. new items. However, as illustrated in the modeling, the reduced discriminability is not very detrimental to categorization, which may require only gross-level assessments of similarity to be made.

A more direct challenge to the exemplar hypothesis comes from brain imaging studies that show that distinct brain regions are activated across recognition vs. categorization tasks (11, 20–23). In a seminal study, Reber et al. (22) conducted categorization tasks in which subjects were presented with statistical distortions of a dot-pattern prototype (24). Following study, subjects were tested with new distortions of the category prototype and with patterns from an unstudied category. In a comparison recognition task, subjects studied five random dot patterns, and then were tested on these five old patterns and new random patterns. Dramatically different patterns of brain activity were observed across the categorization and recognition test phases. In the categorization task, category members were associated with decreased activity in posterior occipital cortex (POC). By contrast, in the recognition task, the old patterns led to increased activity in POC and a variety of other brain regions associated with explicit memory.

Although these differences suggest that categorization and recognition may indeed recruit separate memory systems, alternative possibilities need to be considered. First, note that different stimuli were presented across the tasks. In categorization, subjects endorsed novel patterns that were similar to a single prototype, whereas in recognition they endorsed highly distinct patterns that were exact matches to unrelated, old study items. Thus, rather than reflecting differences in categorization vs. recognition per se, the brain imaging results may have been reflecting these stimulus differences. Second, even if stimulus conditions are held constant, and even if the same basic exemplar similarity processes guide performance, one would still expect to see differences in brain activity across typical versions of the tasks; from a formal perspective, the reason is that subjects would be expected to change parameter settings across the tasks to satisfy the different task goals (6, 7, 9). For example, in recognition, an observer endorses a test item only if it is an exact match to a study item. By contrast, in categorization, a test item is endorsed if it has sufficient similarity to the study items; exact matches are not required. Thus, observers would be expected to vary criterion settings for similarity matching across the tasks, with a much stricter criterion being adopted in recognition than in categorization. Therefore, differences in brain activation across categorization vs. recognition need not be explained in terms of qualitatively different memory systems. An alternative is

that there are quantitative changes in the same cognitive systems. Those quantitative differences reflect changes in parameter settings, such as a shift in criterion for the level of similarity required to endorse a pattern.

In the present study, we address these issues in two ways. First, we conduct brain imaging of categorization and recognition in which stimulus conditions are held constant, with only the task goal being manipulated. Second, we use an instructional manipulation to vary hypothesized parameter settings across conditions, to study how such parameter changes may be implemented in the brain. A formal exemplar model is fitted to the behavioral data to track these hypothesized changes in parameter settings and to help interpret the brain-imaging results.

Tasks

We acquired blood oxygenation-level dependent (BOLD) measurements of brain activation using fMRI while subjects performed a series of categorization and recognition tasks. In all tasks, subjects studied five high-statistical distortions of a novel dot-pattern prototype. At test, subjects were presented with: (*i*) the old high distortions, (*ii*) new distortions of the prototype at both low and high distortion levels, and (*iii*) new random patterns that were unrelated to the prototype. Examples are provided in Fig. 1*A*. In a recognition task (REC), subjects were instructed to endorse a test pattern only if it was old, i.e., an exact match to one of the study patterns. In a categorization task (CAT), subjects were told that they had just experienced members of a single category, and should endorse a test pattern if it was also a member of the category. The instructions made clear that both old and new category members would be presented and that both should be endorsed. Note that the types of study and test stimuli were identical across the CAT and REC conditions, with only the task goal varying. In addition, to hold stimulus-encoding strategies during the study phase constant across conditions (25), subjects were not informed of the detailed task instructions until time of test. Our focused question in this investigation was whether the task goal of categorizing vs. recognizing at time of test, in and of itself, leads to reliance on separate neural memory systems.

Finally, to assess the role of parameter change in influencing the brain imaging results, we also conducted a lax recognition task (LAX). Subjects were given recognition instructions, but the instructions emphasized that it was important to not miss any old items, whereas false alarms to new items were relatively unimportant. The intent was to induce subjects to adopt a lax criterion for making their recognition judgments, and pilot testing revealed that the behavioral data for this LAX condition were similar to those in the CAT condition. Thus, the LAX condition is similar to the REC condition in the sense that the task goal is to recognize, but is similar to the CAT condition in terms of the hypothesized parameter settings from the exemplar model (see below).

Behavioral and Modeling Results

The probability with which subjects endorsed the different pattern types across the three tasks is displayed in Fig. 1*B*. Across all three tasks, subjects endorsed the old items with the highest probability, followed in order by the new low distortions, the new high distortions, and the random patterns. Also, the behavioral data were highly similar across the CAT and LAX conditions, but endorsement probabilities were lower in the REC condition.

We fitted these data using a baseline version of the exemplar model (17). According to the model, the probability that each pattern type i is endorsed in each condition j is given by

$$P(i,j) = S_i/(S_i + K_j),$$ [1]

where S_i is the summed similarity of pattern type i to the old study exemplars, and K_j is a criterion parameter for condition j. To compute the summed similarities, we estimate free parameters s_L, s_H, and s_R, representing the average similarity of the low-distortion, high-distortion, and random patterns to the old study exemplars. (For scaling convenience, the similarity of an old item to itself is set equal to $s_O = 1.0$.) The general version of the exemplar model presumes that selective attention processes may modify similarity relations across different tasks (5–7). For the present types of dot patterns, however, we expect that the role of this selective attention process is relatively minor. Therefore, for simplicity, we assume that the similarity parameters s_L, s_H, and s_R are constant across the CAT, REC, and LAX conditions. Finally, we estimate the criterion setting (K_j) for each task j, which we assume is the only parameter that varies across tasks.

The predictions from the model (solid dots) are shown with the observed data in Fig. 1*B*. The model achieves good fits to the data of all three tasks, in accord with the hypothesis that performance may be mediated by the same exemplar-based retrieval system. The best-fitting parameters, reported in Table 1, are easily interpretable; they show that, on average, the low, high, and random distortions had decreasing levels of similarity to the study exemplars, as would be expected given the statistical distortion techniques used to construct the patterns (and as we have verified in previous scaling work). In addition, the estimated criterion settings were very similar across the CAT and LAX
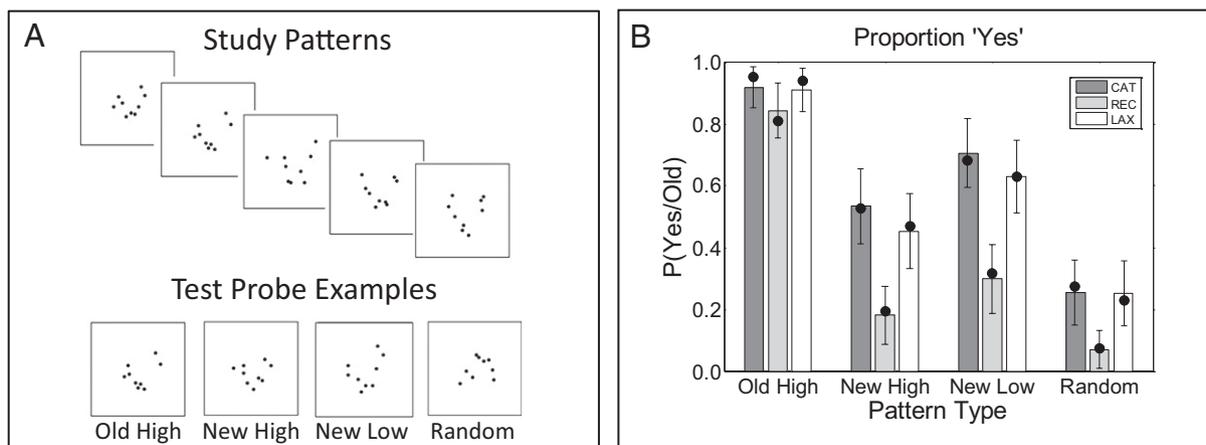

Fig. 1. (*A*) Examples of the study and test patterns presented in each task. (*B*) Proportion of endorsements for each of the pattern types in each of the tasks.

Table 1. Best-fitting parameters from the exemplar model

Parameter	Value
s_L	0.023
s_H	0.012
s_R	0.004
K_{Cat}	0.053
K_{Lax}	0.067
K_{Rec}	0.246

Table 2. Description of clusters found with targeted contrasts

Contrast	Effect	Region label	Coordinates	Brodmann area	Cluster size	Peak Z
CAT/LAX vs. REC	Task	—	—	—	—	—
	Pattern type	—	—	—	—	—
	Interaction	AIC	+32 +23 −8	13/47	3	5.0
		FEF	−26 −5 +52	6	8	5.4
REC/LAX vs. CAT	Task	ACC	−11 +28 +10	24	10	5.2
	Pattern type	—	—	—	—	—
	Interaction	—	—	—	—	—

All clusters were significant at a FWE-corrected level ($\alpha < 0.05$). Coordinates are relative to the MNI template. Cluster size is k_E.

tasks, but a much stricter criterion setting operated in the REC task.

Brain Imaging Analyses

Analytic Approach. For the fMRI analyses, we focus on the planned comparisons of most direct theoretical relevance. Note that, across all three tasks, the "old" and "random" test items have the same correct-response status. That is, for all three of the CAT, REC, and LAX tasks, the correct answer is to endorse "old" items but reject the "random" ones. Thus, the old−random item comparison holds fixed across conditions both the test stimuli and their associated correct responses, while varying only the task and hypothesized parameter settings. If similar patterns of brain activation are observed for both old and random items across the REC and LAX conditions, and these patterns contrast with the CAT condition, it would provide strong evidence of distinct neural processes underlying recognition vs. categorization, regardless of specific parameter settings. Furthermore, if these differences occur in brain regions implicated in memory, it would be convincing evidence for the multiple memory systems view. By contrast, if similar patterns of brain activation are observed across the CAT and LAX conditions, and they contrast with the REC condition, it would suggest a critical role of criterion parameter settings in influencing the patterns of brain activity.

Thus, we conducted whole-brain analyses of variance using as factors task (CAT, LAX, REC) and item type (old, random). To focus on results that are likely to be highly replicable, we use a stringent voxel-wise threshold with family wise error (FWE) correction for multiple tests ($\alpha < 0.05$), and do not report cases in which only a single voxel from a region met this criterion. (We also occasionally discuss results from a more lenient threshold that uses a voxel-wise P value of 0.001 and a cluster-size threshold of 10 voxels.)

CAT/LAX vs. REC Contrast. The most incisive analyses were focused contrast analyses comparing (CAT + LAX)/2 to REC and (REC + LAX)/2 to CAT. The (CAT + LAX)/2 vs. REC comparison (which should be sensitive to parameter changes) revealed highly interpretable effects (Fig. 2 *A* and *B* and Table 2). Using the stringent threshold, we observed a significant interaction between (CAT + LAX)/2 vs. REC and old vs. random items in the frontal eye field (FEF; −26, −5, 52) and the right anterior insular cortex (AIC; 32, 23, −8). Past research suggests that both areas are associated with forms of evidence accumulation in tasks of perceptual decision-making (26–30).

According to the present theoretical account, the major difference between CAT/LAX and REC is that a more lenient decision criterion is established in the former conditions than the latter. As illustrated schematically in Fig. 3, an implication is that evidence accumulation favoring positive responses to old items should proceed more efficiently in the CAT/LAX conditions than in REC, because the summed similarity for old items exceeds the criterion to a greater extent in the former conditions. Analogously, evidence accumulation favoring negative responses to the random items should proceed more efficiently in REC than in CAT/LAX. The interaction effect (which measures whether the old−random difference changes across the CAT/LAX and REC conditions) arises from the combination of these factors. These interaction effects are summarized in Fig. 2*B*, which shows that in both FEF and AIC, the difference in BOLD percent signal
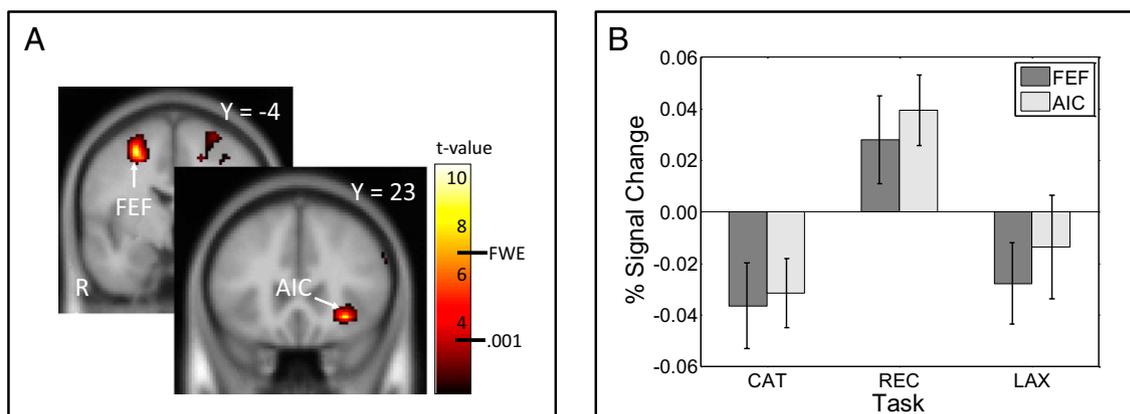
Fig. 2. (*A*) Whole-brain map of the interaction between task and item type. Effects are shown from the CAT/LAX vs. REC contrast. The map is shown at a minimum statistical threshold of $P < 0.001$; however, the two clusters identified by the arrows were also significant using an FWE-corrected threshold. The *t* values corresponding to these two thresholds are indicated on the color legend. Y values indicate coordinates of the coronal slices relative to the Montreal Neurological Institute (MNI) template. (*B*) BOLD percent signal change of the FEF and AIC in each task (difference effect for the old vs. the random patterns).
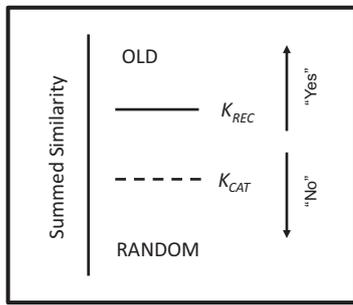
Fig. 3. Schematic illustration of how evidence accumulation relates to criterion placement according to the summed-similarity exemplar model.
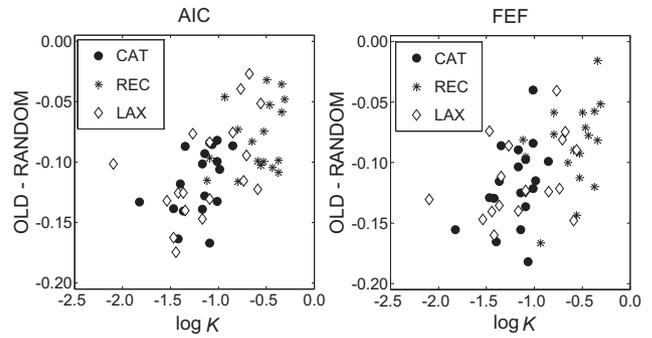

Fig. 4. Scatterplots of old−random BOLD percent signal change against criterion parameter estimates (log K) for all individual subjects in all task conditions.

change for the old−random items was negative in the CAT and LAX conditions, but positive in the REC condition. The forms of evidence accumulation may correspond to the gathering of perceptual information in the case of FEF, and to both modal and amodal forms of evidence accumulation, such as occurs in sequential sampling models of decision making (10, 26, 27, 29, 31–33), in the case of FEF and AIC.

In sum, a significant part of the difference between the brain activation patterns yielded by perceptual categorization and standard old−new recognition is the result of changes in parameter settings. In particular, for the present paradigm, the brain activity differences appear to be related to differences in the efficiency of evidence accumulation resulting from changed criterion settings across the tasks.

To bolster this interpretation, we conducted modeling analyses at the individual-subject level and related the results from these analyses to the brain imaging results. For simplicity, we held the estimated similarity parameters (Table 1) fixed across all conditions and subjects, and computed the values of the criterion parameter K from Eq. **1** that provided maximum-likelihood fits to each individual subject's data in each individual condition. We then computed the correlation between these individual-subject criterion settings (K) and the subjects' old−random brain activation differences within each of the individual task conditions.

Note that the results from our between-task contrast analyses revealed a positive relation between the magnitude of K and the old−random brain activations in both FEF and AIC (Table 1 and Fig. 2B). Therefore, the prior (directional) hypothesis from our account is that there should be a positive correlation between the individual-subject estimates of K and the old−random brain activations within each task. The correlations were statistically significant for AIC in the CAT condition [$r(16) = 0.471$, $P < 0.05$], AIC in the LAX condition [$r(16) = 0.593$, $P < 0.05$], and FEF in the REC condition [$r(16) = 0.410$, $P < 0.05$, one-tailed directional hypothesis]. The correlations went in the same positive direction for AIC in the REC condition and for FEF in both the CAT and LAX conditions, but were smaller in magnitude and did not reach statistical significance [mean $r(16) = 0.359$]. In further analyses, we computed average values of old−random activation across FEF and AIC to create an "accumulation network" aggregate score for each subject. The correlations between the individual-subject K estimates and these aggregate brain activity scores were all significant [$r(16) = 0.602$ for CAT ($P < 0.01$); $r(16) = 0.553$ for LAX ($P < 0.02$); and $r(16) = 0.472$ for REC ($P < 0.05$)].

Because each analysis involves a reduced range of the criterion parameter (i.e., most subjects set a low criterion in the CAT and LAX conditions but a strict criterion in the REC condition), finding these positive correlations within each task provides strong support for the hypothesis that the present brain activation differences between categorization and recognition are driven by changes in parameter settings.

The systematic relations described above are illustrated in Fig. 4, which plots the old−random brain activity differences against log K for all subjects in all conditions. (We use the log transformation for this illustration because many of the K estimates were compressed toward zero in the CAT and LAX conditions.) As expected, Fig. 4 reveals that, in both AIC and FEF, K estimates and old−random brain activity differences both tend to be greater in the REC condition than in the CAT and LAX conditions. Impressively, despite the reduced ranges, this positive relation between K and the brain activity differences is seen within each of the tasks as well.

Finally, in a more exploratory approach, we constructed correlation maps of brain activity. These maps were developed by conducting whole-brain analyses based on correlations between the individual-subject K parameters and the old−random BOLD percent signal change at every voxel (*SI Text*). These maps produced results that converged with the results from our (CAT + LAX)/2 vs. REC contrast analyses. In particular, strong correlations were again observed in AIC and FEF. As described in depth in *SI Text* and *Fig. S1*, the maps also pointed to other brain regions that past work suggests are associated with criterion setting and evidence accumulation in decision making, including areas of posterior parietal cortex as well as the caudate (27, 28, 31, 32, 34, 35). Overall, therefore, there is strong converging support for the idea that changes in activity in the neural networks responsible for categorization and recognition reflect changes in parameter settings.

REC/LAX vs. CAT Contrast. The REC/LAX vs. CAT contrast asks whether there are any brain regions sensitive to the process of recognition that differ from those that mediate categorization, once changes in criterion settings across the tasks are removed. This analysis revealed no interactions between task and item type (old vs. random), regardless of whether the strict or the lenient statistical threshold was applied. This null result is consistent with the exemplar model hypothesis that the present forms of categorization and recognition recruit the same memory system.

Interestingly, however, the analysis did reveal a main effect of task in the anterior cingulate cortex (ACC; −11 28 10; Table 2). The effect was seen most strongly near the intersection of the dorsal and ventral divisions of the ACC. Its posterior location, near to corpus collosum, suggests strongly that the activation is generated in Brodmann's area 24, implying the dorsal division. Although unanticipated by the exemplar model hypothesis, a possible explanation is the following: Past research suggests that dorsal ACC is sensitive to the anticipation of error likelihood or more general forms of response conflict (36, 37). Although corrective feedback was not provided in our design, it is clear that in the standard and lax recognition conditions, there are objective right and wrong answers about whether a test item is an exact

match to some study item. By contrast, in the case of categorization, the situation is fuzzier. That is, in the absence of corrective feedback, whether an item should be called a member of a category is a more open-ended type of judgment, and assignment of error is more arbitrary. Thus, rather than reflecting qualitatively different systems, the main effect of recognition vs. categorization in ACC may reflect a quantitative difference in anticipation of error likelihood. One approach to testing this hypothesis would be to conduct designs in which corrective feedback is provided for all tasks. If an observer learns that there are clear, objective right and wrong answers associated with the categorization judgments, then similar patterns of ACC activation should be observed across the CAT, LAX, and REC conditions.

Regardless of the merits of the error-anticipation hypothesis, it is important to keep in perspective our key theoretical proposal. The key idea is that categorization and recognition may rely on similar cognitive processes and memory representations involving largely overlapping neural systems. It would be too strong to claim that there are no differences between the cognitive and neural processes that mediate categorization and recognition. However, these differences may be smaller than suggested by influential separate systems accounts, which posit that categorization and recognition recruit separate memory systems. To our knowledge, none of these accounts have pointed to this region of ACC as a major locus of such separate memory systems.

ROI analysis in POC. As noted previously, in their categorization task, Reber et al. (22) found that category members were associated with decreased activity in POC compared with nonmembers, an effect termed the "category fluency effect" (11, 20). By contrast, in Reber et al.'s (22) recognition task, old items led to increased activity in this region compared with new items. There were important procedural differences between our study and the one conducted by Reber et al. (22). Nevertheless, as reported in *SI Text* and Fig. S2, an ROI analysis comparing old−random activity levels in this same region of POC in our study yielded the same pattern of results as observed by Reber et al. (22). Crucially, however, the same ROI analysis revealed that the brain imaging data from our LAX condition more closely matched the results from our CAT condition, not the REC condition (see Fig. S2 for details); therefore, POC activity levels may be sensitive to changed parameter settings as well. Future research is needed to ascertain the basis for these interesting effects.

Discussion

Conclusions. In contrast to the prevailing view in the field, our study suggests that when stimulus conditions and hypothesized parameter settings are held roughly constant across the tasks, perceptual categorization and old−new recognition are mediated by largely overlapping neural memory systems. Past evidence that categorization and recognition are mediated by separate systems may have been a reflection of changed stimulus conditions, adaptive changes in parameter settings, or both. The results from our behavioral data, formal modeling analyses, and brain imaging analyses are consistent with theoretical proposals that perceptual categorization and recognition are governed by similar representations and processes, with systematic changes in parameter settings occurring across the tasks.

Support for the parameter-change interpretation was obtained from several converging sources of evidence. First, a baseline version of an exemplar model provided a good quantitative account of the choice probabilities from all of the categorization and recognition conditions, simply by allowing for systematic changes in parameter settings. Second, analysis of the averaged data indicated that differences in brain activity patterns across the categorization and recognition tasks were interpretable in terms of changes in evidence accumulation arising from changes in criterion settings. Third, modeling analyses conducted at the

individual subject level revealed significant correlations between the estimated criterion settings and these brain activity patterns within the individual conditions.

Although the major form of hypothesized parameter change was a change in criterion settings, the full version of the exemplar model presumes that other forms of parameter change may occur as well. Most notably, people may attend selectively to component dimensions of stimuli, causing changes in similarity relations across tasks (6). For example, in categorization, people tend to focus attention on the subset of dimensions that is most characteristic of the category members; whereas in recognition, they may spread attention to other dimensions for purposes of making fine-grained discriminations between old and new items (7). To obtain a fuller understanding of the neural processes that underlie categorization and recognition, future research will need to track the brain regions that implement these selective attention processes.

Limits to Generality. Several factors limit the generality of our conclusion that categorization and recognition are mediated by largely overlapping cognitive and neural processes. First, we are referring to similarity-based forms of "prototype extraction" in which a single category is learned from induction over exemplars and without corrective feedback. We focused on such a task because it is perhaps the major one that has been used in past work to argue that separate systems mediate categorization and recognition (11). More elaborate versions of the exemplar model have been applied to designs that involve multiple categories and learning with feedback (5−7, 10).

Second, the present conditions were tested in a within-subjects design. Although our instructions made clear that each task was new and independent, subjects may nevertheless have adopted similar strategies because the tasks were juxtaposed, perhaps even adopting a categorization strategy in the LAX condition. In a preliminary attempt to address this concern, we conducted analyses involving the order in which the tasks were tested (see *SI Text* for details). These analyses revealed that the patterns of brain activity observed in the LAX condition were extremely similar, regardless of whether LAX was tested before or after CAT (Fig. S3). Furthermore, ROI analyses in AIC, FEF, and ACC revealed reasonably similar patterns of brain activity across the CAT, LAX, and REC conditions in the first and second halves of the testing session (Fig. S4). Nevertheless, for purposes of generality, future research should test similar conditions in between-subjects designs.

Third, even when holding stimulus conditions and hypothesized parameter settings constant, we did observe brain activity differences in ACC across the categorization and recognition conditions. One possible interpretation is that these results reflect differences in anticipation of error likelihood across the tasks. However, future research needs to test this interpretation.

Although no single study can be definitive, the present work reopens the interesting possibility that perceptual categorization and old−new recognition are mediated by similar cognitive and neural systems, with only adaptive parameter changes arising across tasks.

Methods

Behavioral Protocol. Eighteen right-handed college students (11 females) participated for monetary compensation. Subjects provided informed written consent in accord with the policies of the Indiana University Humans Subjects Review Board.

For each subject and task, a novel random dot-pattern prototype was generated (24). Statistical distortion procedures (24) were then used to generate five old high distortions of the prototype, four new low distortions, 12 new high distortions, and 20 new random patterns. During the study phase, there were three blocks in which each of the old distortions was presented in a random order for 5,000 ms each. During the test phase, each old distortion was presented four times, and each of the new

distdistortions and random patterns was presented once. Stimuli were presented in random order with a variable interval between the onset of the stimulus and the start of the next trial based on random exponentially distributed jitter (4,000, 6,000, or 8,000 ms). Each stimulus was presented for 2,000 ms with a blank interval for the remainder of the randomly generated time period. In all tasks during the study phase, subjects were instructed to simply view the patterns, and the instructions made clear that new questions would be asked about the patterns at the time of test. The test instructions were as described in the text. Subjects participated in each task twice in two separate series of the three tasks. A Latin Square was used to balance the ordering of the CAT, LAX, and REC conditions in the first series, and a separate Latin Square was used to balance the ordering in the second series.

Imaging Procedures. Details of imaging and statistical analysis methods are provided in *SI Text*.

1. Estes WK (1994) *Classification and Cognition* (Oxford Univ Press, New York).
2. Smith EE, Medin DL (1981) *Categories and Concepts* (Harvard Univ Press, Cambridge, MA).
3. Anderson JR (1991) The adaptive nature of human categorization. *Psychol Rev* 98: 409–429.
4. Love BC, Gureckis TM (2007) Models in search of a brain. *Cogn Affect Behav Neurosci* 7(2):90–108.
5. Medin DL, Schaffer MM (1978) Context theory of classification learning. *Psychol Rev* 85:207–238.
6. Nosofsky RM (1986) Attention, similarity, and the identification-categorization relationship. *J Exp Psychol Gen* 115(1):39–61.
7. Nosofsky RM (1991) Tests of an exemplar model for relating perceptual classification and recognition memory. *J Exp Psychol Hum Percept Perform* 17(1):3–27.
8. Kahana MJ, Sekuler R (2002) Recognizing spatial patterns: A noisy exemplar approach. *Vision Res* 42:2177–2192.
9. Nosofsky RM, Little DR, Donkin C, Fific M (2011) Short-term memory scanning viewed as exemplar-based categorization. *Psychol Rev* 118:280–315.
10. Nosofsky RM, Palmeri TJ (1997) An exemplar-based random walk model of speeded classification. *Psychol Rev* 104:266–300.
11. Smith EE (2008) The case for implicit category learning. *Cogn Affect Behav Neurosci* 8(1):3–16.
12. Bozoki A, Grossman M, Smith EE (2006) Can patients with Alzheimer's disease learn a category implicitly? *Neuropsychologia* 44:816–827.
13. Kéri S, Kálmán J, Kelemen O, Benedek G, Janka Z (2001) Are Alzheimer's disease patients able to learn visual prototypes? *Neuropsychologia* 39:1218–1223.
14. Knowlton BJ, Squire LR (1993) The learning of categories: Parallel brain systems for item memory and category knowledge. *Science* 262:1747–1749.
15. Reed JM, Squire LR, Patalano AL, Smith EE, Jonides JJ (1999) Learning about categories that are defined by object-like stimuli despite impaired declarative memory. *Behav Neurosci* 113:411–419.
16. Squire LR, Knowlton BJ (1995) Learning about categories in the absence of memory. *Proc Natl Acad Sci USA* 92:12470–12474.
17. Nosofsky RM, Zaki SR (1998) Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychol Sci* 9: 247–255.
18. Palmeri TJ, Flanery MA (2002) Memory systems and perceptual categorization. *The Psychology of Learning and Motivation: Advances in Research and Theory*, ed Ross BH (Academic, San Diego), Vol 41, pp 141–189.
19. Zaki SR, Nosofsky RM (2001) A single-system interpretation of dissociations between recognition and categorization in a task involving object-like stimuli. *Cogn Affect Behav Neurosci* 1:344–359.
20. Reber PJ, Gitelman DR, Parrish TB, Mesulam MM (2003) Dissociating explicit and implicit category knowledge with fMRI. *J Cogn Neurosci* 15:574–583.
21. Reber PL, Stark CEL, Squire LR (1998) Cortical areas supporting category learning identified using functional magnetic resonance imaging. *Proc Natl Acad Sci USA* 95: 747–750.
22. Reber PJ, Stark CEL, Squire LR (1998) Contrasting cortical activity associated with category memory and recognition memory. *Learn Mem* 5:420–428.
23. Reber PJ, Wong EC, Buxton RB (2002) Comparing the brain areas supporting nondeclarative categorization and recognition memory. *Brain Res Cogn Brain Res* 14: 245–257.
24. Posner MI, Goldsmith R, Welton KE, Jr. (1967) Perceived distance and the classification of distorted patterns. *J Exp Psychol* 73(1):28–38.
25. Gureckis TM, James TW, Nosofsky RM (2011) Re-evaluating dissociations between implicit and explicit category learning: An event-related fMRI study. *J Cogn Neurosci* 23:1697–1709.
26. Forstmann BU, Brown S, Dutilh G, Neumann J, Wagenmakers EJ (2010) The neural substrate of prior information in perceptual decision making: A model-based analysis. *Front Hum Neurosci* 4:40.
27. Ho TC, Brown S, Serences JT (2009) Domain general mechanisms of perceptual decision making in human cortex. *J Neurosci* 29:8675–8687.
28. Little DM, Thulborn KR (2006) Prototype-distortion category learning: A two-phase learning process across a distributed network. *Brain Cogn* 60:233–243.
29. Purcell BA, et al. (2010) Neurally constrained modeling of perceptual decision making. *Psychol Rev* 117:1113–1143.
30. Sterzer P, Kleinschmidt A (2010) Anterior insula activations in perceptual paradigms: Often observed but barely understood. *Brain Struct Funct* 214:611–622.
31. Gold JI, Shadlen MN (2007) The neural basis of decision making. *Annu Rev Neurosci* 30:535–574.
32. Ploran EJ, et al. (2007) Evidence accumulation and the moment of recognition: Dissociating perceptual recognition processes using fMRI. *J Neurosci* 27:11912–11924.
33. Usher M, McClelland JL (2001) The time course of perceptual choice: The leaky, competing accumulator model. *Psychol Rev* 108:550–592.
34. Forstmann BU, et al. (2008) Striatum and pre-SMA facilitate decision-making under time pressure. *Proc Natl Acad Sci USA* 105:17538–17542.
35. Kuchinke L, et al. (2011) Human striatal activation during adjustment of the response criterion in visual word recognition. *Neuroimage* 54:2412–2417.
36. Brown JW, Braver TS (2005) Learned predictions of error likelihood in the anterior cingulate cortex. *Science* 307:1118–1121.
37. Sohn MH, Albert MV, Jung K, Carter CS, Anderson JR (2007) Anticipation of conflict monitoring in the anterior cingulate cortex and the prefrontal cortex. *Proc Natl Acad Sci USA* 104:10330–10334.

Supporting Information

Nosofsky et al. 10.1073/pnas.1111304109

SI Text

Correlation Map Analysis. To supplement the whole-brain contrast analyses and ROI correlation analyses reported in the article, we also conducted exploratory whole-brain correlation analyses involving the estimated criterion parameters from the exemplar model. In these analyses, we computed correlations between the maximum-likelihood values of K (considered simultaneously across all three tasks and all subjects) and the old−random BOLD percent signal change at every voxel for all of the tasks. Thus, each individual voxel correlation was computed across 54 pairs of scores. One member of each pair was the maximum-likelihood estimate of K for subject i in task j; the other member was the old−random BOLD percent signal change for subject i in task j computed at that voxel. Correction for multiple tests was done using the family wise error (FWE) method (described in main text) to determine brain regions whose activity patterns showed significant correlations with the estimated criterion parameters, making the assumption, for simplicity, that all 54 pairs of observations were independent. These P values need to be interpreted with caution, because each subject contributes three observations to the computation of the correlation, which likely leads to a more liberal criterion than if there were 54 independent subjects. Thus, these analyses should be viewed as an exploratory, descriptive complement to the main contrast analyses reported in the text, which used a strict statistical criterion.

The correlation map derived with this method is displayed in Fig. S1. In convergence with the results from the contrast analyses, the map shows strong activation in the frontal eye field (FEF) and anterior insular cortex (AIC); in the present analyses, this result is revealed as bilateral activation of these regions, whereas in the contrast analyses only unilateral activation of these regions reached the strict statistical threshold. In addition, the correlation map revealed significant activation in the posterior parietal cortex and lateral occipital cortex. Although there were no significant clusters in the parietal or occipital cortex in our contrast analyses using the strict threshold, there were significant clusters with the more lenient threshold, and these regions overlapped with the regions found in the present correlation analysis. As noted in the main text, there is extensive past evidence to suggest that areas of posterior parietal cortex are associated with evidence accumulation in perceptual decision-making (1–4). Likewise, there is also some past evidence that implicates the lateral occipital cortex (2, 5, 6). Finally, the correlation map also showed significant clusters in the head and body of the caudate nucleus, the dorsal ACC, and the dorsolateral prefrontal cortex. None of these regions overlapped with any clusters found in the contrast analysis, even with the lenient threshold. However, all of the clusters described for the correlation analysis have been implicated in accumulation of evidence for categorization or other perceptual decision-making tasks in previous studies (1–4, 7, 8).

ROI Analysis in Posterior Occipital Cortex. We conducted an ROI analysis comparing old−random activity levels in the region of posterior occipital cortex (POC) that had been identified by Reber et al. (9) as giving rise to different activation patterns for categorization vs. recognition (Talairach coordinates 12 −93 17; MNI coordinates 12 −93 13). The results are shown in Fig. S2. As can be seen in the figure, despite differences in procedure across the two studies, we observed the same patterns of activity as did Reber et al. (9), with negative activations for categorization but positive activations for recognition. Crucially, however,

the same ROI analysis conducted on the brain imaging data from our lax recognition condition (LAX) more closely matches the results from our categorization condition, not the recognition condition (Fig. S2).

Analysis of Order Effects. In one set of analyses, we tested whether the patterns of brain activity in the regions of interest [AIC, FEF, and anterior cingulate cortex (ACC)] differed for the LAX condition depending on whether it was tested before or after the categorization task (CAT) condition. Note that there were nine subjects in the pre-CAT condition and nine subjects in the post-CAT condition. [This mode of analysis ignores the positioning of the recognition task (REC) condition.] For AIC and FEF, the unit of analysis was the old - random BOLD percent signal change, whereas for ACC the unit of analysis was the old + random BOLD percent signal change. The results are displayed in Fig. S3. None of the differences approached statistical significance [$t(16) = 0.27$, $P = 0.79$ in AIC; $t(16) = 0.41$, $P = 0.68$ in FEF; $t(16) = 1.21$, $P = 0.25$ in ACC].

In a second set of analyses, we tested whether the patterns of performance across all three conditions (CAT, LAX, and REC) varied during the experimental session. Recall that each of the tasks was tested twice in two repeated triads (*Methods*). The brain activity results for each task in each repetition are displayed in Fig. S4, with a separate panel for each brain ROI (AIC, FEF, and ACC). Again, for AIC and FEF, the unit of analysis is the old - random BOLD percent signal change, whereas for ACC the unit of analysis is the old + random BOLD percent signal change. We analyzed the data in each brain region using a 2 (repetition) × 3 (task) repeated-measures analysis of variance. The main effect of task was highly significant in all three brain regions, as revealed by the whole-brain analyses reported in the main text [$F(2, 17) = 9.1$, $P = 0.001$ for AIC; $F(2, 17) = 8.7$, $P = 0.001$ for FEF; $F(2, 17) = 20.8$, $P = 0.000$ for ACC]. The important new question is whether these patterns of brain activity interacted with repetition. Although there are hints of certain changes from the first to the second repetition, the directions of change are not consistent across the different brain regions. Most important, none of the interaction effects between task and repetition approached statistical significance [$F(2, 17) = 0.87$, $P = 0.43$ in AIC; $F(2, 17) = 1.40$, $P = 0.26$ in FEF; $F(2, 17) = 1.98$, $P = 0.15$ in ACC]. A reasonable summary statement is that, averaged across the repetitions, the patterns of brain activity line up for CAT and LAX and differ from REC in AIC and FEF; whereas the patterns of brain activity line up for REC and LAX and differ from CAT in ACC.

Imaging and Statistical Analysis Methods. *Imaging.* MRI data were acquired using a Siemens MAGNETOM Tim Trio 3 Tesla whole-body MRI equipped with a 32-channel phased-array head coil. The field of view was 24 × 24 × 11.6 cm, with an in-plane matrix of 96 × 96 pixels and 33 axial slices per volume (whole brain), creating a voxel size of 2.5 × 2.5 × 3.5 mm. Images were collected using a gradient echo EPI sequence (TE = 30 ms, TR = 2,000 ms, flip angle = 70°) for BOLD imaging. High-resolution T1-weighted anatomical volumes were acquired using TurboFLASH 3D sequence (TI = 1,100 ms, TE = 3.93 ms, TR = 14.375 ms, flip angle = 12°) with 160 1-mm sagittal slices and an in-plane field of view of 224 × 256 (voxel size = 1 mm^3). Imaging data were preprocessed using the SPM8 toolbox for MATLAB (Wellcome Department of Cognitive Neurology, London; http://www.fil.ion.ucl.ac.uk/spm). Anatomical volumes were normal-

ized to a common stereotactic space [Montreal Neurological Institute (MNI)]. Functional volumes were coregistered to the normalized anatomical volumes and then transformed to the common space using a 12-parameter affine transformation. Functional data underwent 3D spatial Gaussian filtering (FWHM 8 mm), slice scan-time correction, and 3D motion correction. Maximum allowable movement was 1.5 mm for gradual drift throughout a functional run and 1 mm for transient motion spikes.

Statistical analysis. Normalized functional volumes were analyzed using a random effects general linear model procedure to produce statistical parametric maps. For each participant, the design matrix included regressors for correct trials for old high distortions, new distortions, and random distortions, as well as a combined "wrong"

regressor for all incorrect trials. (In the CAT, REC, and LAX conditions, correct responses were defined as "yes/old" responses for old items and "no/new" responses for random items. In addition, for REC, "no/new" responses for new distortions were defined as correct responses, whereas for CAT and LAX, "yes/old" responses for new distortions were defined as correct responses). Trials were represented as boxcar functions convolved with a canonical 2γ hemodynamic response function. In addition to these experimental regressors, the design matrix also included regressors for the six movement parameters derived from the motion-correction preprocessing stage and a constant for each condition replication. Contrast images for all participants were entered into second-level random-effects group contrast analyses.

1. Little DM, Thulborn KR (2006) Prototype-distortion category learning: A two-phase learning process across a distributed network. *Brain Cogn* 60:233–243.
2. Ploran EJ, et al. (2007) Evidence accumulation and the moment of recognition: Dissociating perceptual recognition processes using fMRI. *J Neurosci* 27:11912–11924.
3. Daniel R, et al. (2011) Assessing the neural basis of uncertainty in perceptual category learning through varying levels of distortion. *J Cogn Neurosci* 23:1781–1793.
4. Grinband J, Hirsch J, Ferrera VP (2006) A neural representation of categorization uncertainty in the human brain. *Neuron* 49:757–763.
5. James TW, Gauthier I (2006) Repetition-induced changes in BOLD response reflect accumulation of neural activity. *Hum Brain Mapp* 27(1):37–46.
6. Philiastides MG, Sajda P (2007) EEG-informed fMRI reveals spatiotemporal characteristics of perceptual decision making. *J Neurosci* 27:13082–13091.
7. Forstmann BU, et al. (2008) Striatum and pre-SMA facilitate decision-making under time pressure. *Proc Natl Acad Sci USA* 105:17538–17542.
8. Kuchinke L, et al. (2011) Human striatal activation during adjustment of the response criterion in visual word recognition. *Neuroimage* 54:2412–2417.
9. Reber PJ, Stark CEL, Squire LR (1998) Contrasting cortical activity associated with category memory and recognition memory. *Learn Mem* 5:420–428.
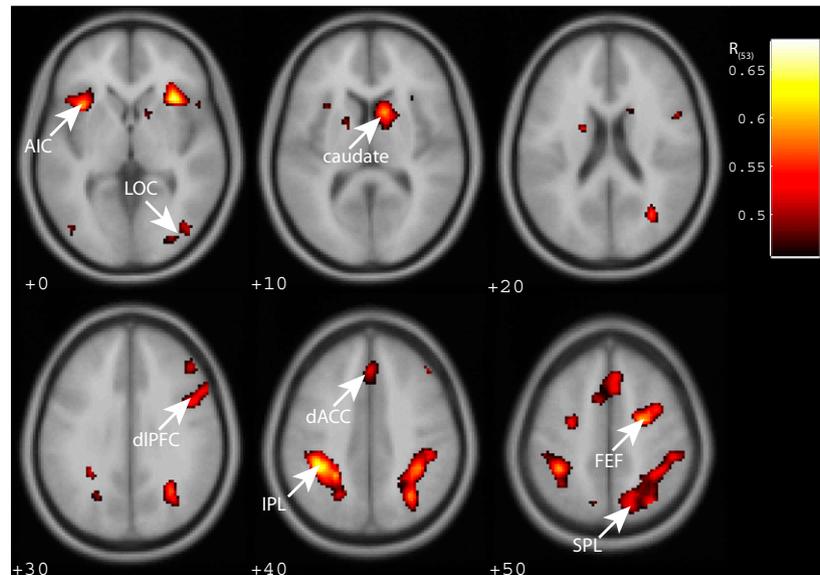
Fig. S1. Whole-brain correlation map relating old–random BOLD percent signal change to the maximum-likelihood *K* parameters derived by fitting the exemplar model to each individual subject's data in each individual task condition. Numbers in the bottom left corner of each panel reflect position along the *z* axis in the MNI reference. caudate, head and body of the caudate; dlPFC, dorsolateral prefrontal cortex; IPL, inferior parietal lobule; LOC, lateral occipital cortex; SPL, superior parietal lobule.
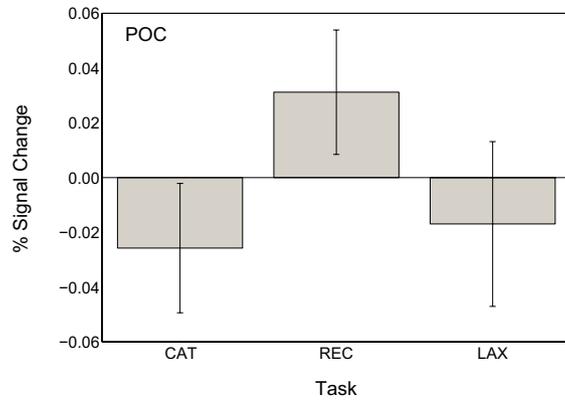
Fig. S2. Mean BOLD percent signal change in the POC for each task (difference effect for the old vs. the random patterns).
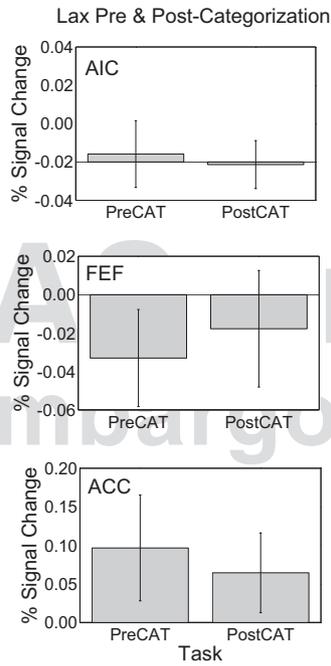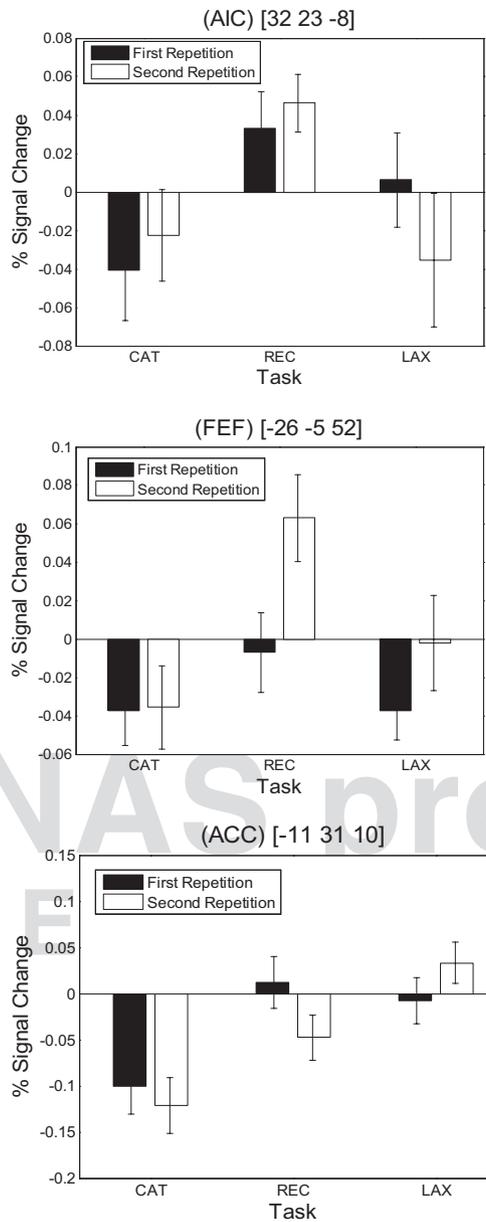

Fig. S3. Mean BOLD percent signal change in the LAX condition as a function of its order of testing with respect to the CAT condition. The results are shown separately for AIC, FEF, and ACC.

Fig. S4. Mean BOLD percent signal change in the CAT, REC, and LAX conditions as a function of task repetition (first vs. second triad of testing). The results are shown separately for AIC, FEF, and ACC.