Reaction Time for Object Categorization Is Predicted by Representational Distance

Thomas A. Carlson[1,2], J. Brendan Ritchie[1,2], Nikolaus Kriegeskorte[3], Samir Durvasula[2], and Junsheng Ma[2]

Abstract

■ How does the brain translate an internal representation of an object into a decision about the object's category? Recent studies have uncovered the structure of object representations in inferior temporal cortex (IT) using multivariate pattern analysis methods. These studies have shown that representations of individual object exemplars in IT occupy distinct locations in a high-dimensional activation space, with object exemplar representations clustering into distinguishable regions based on category (e.g., animate vs. inanimate objects). In this study, we hypothesized that a representational boundary between category representations in this activation space also constitutes a decision boundary for categorization. We show that behavioral RTs for categorizing objects are well described by our activation space hypothesis. Interpreted in terms of classical and contemporary models of decision-making, our results suggest that the process of settling on an internal representation of a stimulus is itself partially constitutive of decision-making for object categorization. ■

INTRODUCTION

Perception requires the brain to make decisions. Detailed models of the link between stimulus, decision, and behavior have been motivated by this fundamental insight. Perceptual decision-making necessitates two intermediary components: a formulation of an internal representation of the stimulus and readout from this representation for behavioral decisions, such as a motor response indicating a choice. This raises the question of how models of perceptual decision-making are to be mapped to the brain activity that performs this transformation and implements the decision. In recent years, important progress has been made on this question (Gold & Shadlen, 2007; Smith & Ratcliff, 2004). Primate studies investigating vibrotactile working memory (Deco, Rolls, & Romo, 2010; Romo & Salinas, 2003) and visual motion perception (Gold & Shadlen, 2007) have successfully mapped decision models to neural activity. In the domain of motion perception, for example, a causal link between activity in MT and perceptual decisions has been demonstrated by administering a current to bias behavioral responses for motion direction (Hanks, Ditterich, & Shadlen, 2006; Shadlen & Newsome, 2001); activity in MT and lateral intraparietal cortex (LIP) neurons have been linked to decision-making to predict choice, RT, and decision confidence (Kiani & Shadlen, 2009; Roitman & Shadlen, 2002); and RT in a random-dot motion detection task has been modeled using sequential analysis applied to activity recorded in LIP (Gold & Shadlen, 2002; Roitman & Shadlen, 2002).

Detecting visual motion direction and matching the vibrotactile frequency of cutaneous stimulations are sensory discrimination tasks with well-described stimulus parameters. Like these two perceptual tasks, the categorization of visual objects is a core form of perceptual decision-making, in which the brain is able to make rapid decisions (Kirchner & Thorpe, 2006; Thorpe, Fize, & Marlot, 1996; Potter, 1976). Unlike the study of visual motion and tactile stimulation, however, the parameterization of visual object categories is opaque. For motion, the "opposite" of a rightward-moving stimulus (90°) is leftward motion (270°), and for tactile stimulation, the difference between stimulations can be easily quantified (e.g., the difference between 10 and 15 Hz stimulation is 5 Hz). In the domain of object recognition, this is less straightforward. What is the opposite of being human? What is close to being human? Addressing this issue represents a formidable challenge for developing models of decision-making for object recognition.

It is clear from the organization of inferior temporal cortex (IT) that it is a critical region for encoding object category information (Kiani, Esteky, Mirpour, & Tanaka, 2007; Downing, Chan, Peelen, Dodds, & Kanwisher, 2006; Hung, Kreiman, Poggio, & DiCarlo, 2005; Freedman, Riesenhuber, Poggio, & Miller, 2003; Grill-Spector, Kourtzi, & Kanwisher, 2001; Haxby et al., 2001). And akin to the previously described studies on motion perception,

[1]Macquarie University, Sydney, Australia, [2]University of Maryland, [3]MRC Cognition and Brain Sciences Unit, Cambridge, UK

stimulation of IT neurons biases behavioral responses for object recognition (Afraz, Kiani, & Esteky, 2006). This research strongly suggests that IT plays an important role in categorical decision-making. The precise nature of this role, however, remains elusive. To date, we know of no successful attempt at predicting RTs based on activity in IT. One promising approach was to examine the relationship between RTs and neuronal latencies in IT (DiCarlo & Maunsell, 2005). This research, however, found little covariance between the response latencies of IT neurons and RT. Furthermore, at present, successful prediction of RT from neural activity has been restricted to studies using invasive methods; no equivalent results have been reported based on analyses of human data using non-invasive methods such as fMRI or MEG/EEG. Studies employing these methods with human participants have instead identified regions other than IT as the locus of perceptual decision-making for object recognition (Philiastides & Sadja, 2006; Heekeren, Marrett, Bandettini, & Ungerleider, 2004).

Recent studies have sought to characterize the geometry of visual object representations in IT using single unit recordings in primates and fMRI in humans (Kriegeskorte et al., 2008; Kiani et al., 2007) with a brain area's representational geometry being derived from the dissimilarity of brain activity produced by a large number of object exemplars. In their analysis, Kriegeskorte et al. (2008) found a remarkable correspondence in the representational geometry between primate and human IT, providing strong evidence that this geometry is a central organizing principle for object representations. Two key insights from these studies might serve to elucidate the role of IT in the perceptual decision-making for object categorization. First, representational dissimilarity could be used as a proxy for physical stimulus parameters to address the issue of quantifying the difference between object stimuli. Second, one of the key features of the representational structure revealed in these studies was the clustering of object representations according to categories (e.g., faces and bodies) with a prominent distinction between animate and inanimate objects. The delineation of these categories in the representational space suggests the existence of a representational boundary between the animate and inanimate object categories.

In this study, we were motivated by the hypothesis that the representational boundary between animate and inanimate objects in IT reflects a decision boundary for perceptual decision-making (see Figure 1). From a decision-making perspective, this hypothesis predicts that object exemplar representations closer to the boundary are less easily distinguishable relative to those farther from the boundary. This is analogous to the idea from signal detection theory (SDT; MacMillan & Creelman, 2005; Green & Swets, 1966) that a signal is easiest to detect when it produces inputs/evidence that is far from the decision criterion, reflecting a substantial difference in the location of the signal and noise distributions, resulting in faster,

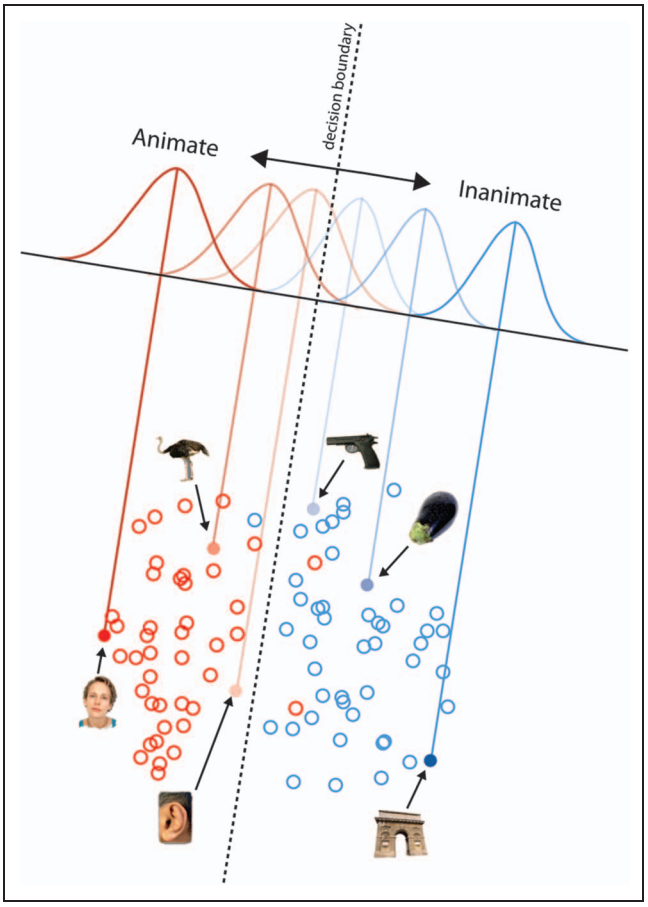

Figure 1. Decision model applied representational space. A 2-D reconstruction of IT's representational geometry with 92 visual object exemplars. A linear boundary (dashed line) divides the space into animate and inanimate objects forming a "decision boundary." Using the Euclidean distance of individual object representations from the decision boundary, the exemplar representations are projected onto a 1-D decision axis with Gaussian noise ("decision noise") added to form probability distribution. Example probability distributions are shown for six of the object exemplars (filled). Classical SDT predicts that as distance from the decision boundary increases discrimination performance will improve.

more accurate, and more confident judgments. This hypothesis generates a natural prediction regarding categorization times: RTs for categorizing individual objects as animate or inanimate would negatively correlate with the distance of these individual objects exemplar representations from the animacy decision boundary in IT; that is, objects closer to the boundary would be more difficult to categorize, yielding longer RTs, whereas exemplars farther from the boundary would yield easier, faster decisions. Our findings confirm this hypothesis in showing that exemplar representations that are distant from the representational boundary are more quickly categorized. We further sought to treat this boundary as a decision boundary for a sequential analysis model, a biologically plausible model of perceptual decision-making (Gold & Shadlen, 2002), which has previously been used in animal studies to predict RT based on neural data (Roitman &

Shadlen, 2002). Using this model, we transformed representational distances to RTs, suggesting that such models might also be applied to the relationship between representational distance and RT.

METHODS

fMRI Experiment and Analysis: Data from Kriegeskorte et al.

Kriegeskorte et al. (2008) characterized the geometry of object exemplar representations in human IT and early visual cortex (EVC) using fMRI multivoxel pattern analysis. The experimental design and analysis procedures for the experiment are described in greater detail in Kriegeskorte et al. (2008) and its supplemental material. Below we provide a brief summary of the methodology and analysis relevant to this study.

Participants ($n = 4$) were presented with 92 images of natural objects against a continuously present gray background for 300 msec, with a 3700 msec ISI (four other images were presented but excluded from analysis; see Kriegeskorte et al., 2008). Stimuli were displayed foveally and spanned 2.9° of visual angle. During each run, participants performed a fixation task, which required them to discriminate changes in the color of the fixation cross overlying the stimuli. Participants were not explicitly instructed to attend to the stimuli or make any form of categorization judgment.

The geometry of exemplar representations in IT and EVC were analyzed using multivoxel pattern analysis techniques. Representational dissimilarity matrices (RDM) were created for IT and EVC. Briefly, an RDM represents the difference in activation patterns between exemplar stimuli. Dissimilarity in the brain activation pattern was measured using the correlation distance $1 - r$ (Pearson correlation). Each entry in the RDM is the dissimilarity in brain activation patterns between two object exemplars. The full RDM includes all possible pairwise comparisons between the object exemplars. Our analysis was conducted on the average IT and EVC RDMs across participants.

Computing Decision Boundaries and Representation Distances

An RDM fully represents the representational distances between the stimuli. Multidimensional scaling (MDS; metric stress criterion) was applied to the RDMs to project the data into a lower-dimensional space suitable for linear discriminate analysis (LDA; Duda, Hart, & Stork, 2001). MDS solutions can vary based on the initial exemplar "seed" position configuration. The results reported are based on the average MDS solution across 100 random initial configurations. In using MDS, we had the option to choose the number of dimensions in the reconstruction. We examined how this affected the main findings of our study by running the analysis on reconstructions of 3–11 dimensions, at which point the MDS minimization of criterion failed to converge for a large proportion of the seed configurations (>.40). The results reported in the main body of the article are for an 8-D reconstruction. The number of dimensions did not impact the findings of the study for IT but did have an impact on the findings for EVC (see Results and Table 1).

LDA was used to find the discriminate axis and decision boundary for animate/inanimate classification (see Figure 1). Table 1 shows the cross-validated classification accuracy for IT and EVC. Cross validation was performed using a leave-one-exemplar-out procedure. In this procedure, one exemplar's data were removed from the data set, the remaining exemplars' data were used to train the classifier, and the trained classifier was then tested with the excluded exemplars' data. The discriminate axis and decision boundary for animate/inanimate classification (see Figure 1) were computed using all the data. After computing the decision boundary using LDA, we measured Euclidean distance of each exemplar's position from the decision boundary surface for IT and EVC.

Behavioral RT Experiment

Participants

As the fMRI participants passively viewed the stimuli, behavioral data were acquired on-line using Amazon's Mechanical Turk services. All participants consented to participate in the experiment. Each participant reported being above the age of 18 years (verified by Amazon) and as having normal or corrected-to-normal vision. To ensure compliance in completing the task as instructed, participants were required to complete eight pretest trials of the task. To qualify for the study, participants needed to answer correctly on seven of the eight trials. Participants that qualified and completed the study ($n = 50$) were compensated $2.50 for approximately 10 min of their time.

RT Experiment

Stimuli were images of natural objects (48 animate, 44 inanimate) used by Kriegeskorte et al. (2008), plus the eight pretest stimuli. Each participant performed one trial for each image. The stimuli were presented for 500 msec in random order on a gray background. Participants were instructed to categorize "as quickly and accurately as possible" each image as animate or inanimate, where an animate object was described as one that could "move on its own volition." Participants were given 2000 msec to respond per trial. After each response, the fixation point changed to either green (correct) or red (incorrect) to provide feedback to participants regarding their performance. Between trials, there was an ISI of 1500 msec in which a fixation point at the center of the screen was presented on a gray background. Error trials

Table 1. Effect of Number of Reconstructed Dimensions

Number of Dimensions	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*
LDA Classification Performance									
IT	95.65%	96.74%	96.74%	96.74%	96.74%	96.74%	95.65%	95.65%	94.57%
EVC	61.96%	59.78%	60.87%	55.43%	54.35%	53.26%	64.13%	64.13%	64.13%
Distance Correlation between IT and EVC									
IT and EVC distance (Spearman's ρ)									
ALL	.1767	.1586	.2258*	.2803**	.3285**	.3564***	.3148**	.252*	.3234**
Results for All Exemplars									
IT (Spearman's ρ)									
ALL	−.2801**	−.3373**	.3373***	−.3765***	−.386***	−.3991***	−.4057***	−.4007***	−.4293***
Animate	−.5973***	−.6432***	−.7057***	−.6804***	−.6898***	−.6993***	−.6982***	−.6988***	−.7058***
Inanimate	.0926	.0286	.0396	.0619	.0388	.0245	.0172	.0231	.0169
EVC (Spearman's ρ)									
ALL	−.2208*	−.2028*	−.2464*	−.2296*	−.2324*	−.2665**	−.2252*	−.2246**	−.227*
Animate	−.4112**	−.3015*	−.3548**	−.4328**	−.4089**	−.4935**	−.5055***	−.3655**	−.3678**
Inanimate	−.0309	−.1333	−.1949	−.0884	−.1183	−.0603	.0806	−.0733	−.0764
Results for Correctly Classified Exemplars									
IT (Spearman's ρ)									
ALL	−.2854*	−.3428**	−.3709**	−.3837**	−.3949***	−.4085***	−.4143***	−.4092***	−.4394***
Animate	−.5973***	−.6432***	−.7057***	−.6804***	−.6898***	−.6993***	−.6982***	−.6988***	−.7058***
Inanimate	.1027	.049	.0765	.0845	.0599	.043	.0416	.0548	.0364
EVC (Spearman's ρ)									
ALL	−.179	−.2326*	−.1995	−.125	−.1881	−.1854	−.1841	−.1402	−.1779
Animate	−.3593*	−.4212*	−.3226*	−.2831	−.403*	−.3548*	−.4905**	−.2255	−.3315*
Inanimate	−.0577	−.1493	−.2289	−.0773	−.0407	−.0044	.0601	−.0216	−.0996
IT Results Controlling for EVC									
IT (partial Spearman's ρ)									
ALL	−.2951**	−.3078**	−.3291**	−.3198**	−.3451***	−.344***	−.357***	−.3745***	−.4005***
Animate	−.5266***	−.5624***	−.6159***	−.5533***	−.5826***	−.5834***	−.5592***	−.635***	−.6304***
Inanimate	−.0049	.0197	.051	.0585	.0069	.0072	−.0193	−.0036	−.0289

The top table reports the LDA classification accuracy in percent correct for IT and EVC for different numbers of reconstructed dimensions. The second table shows correlation (Spearman's ρ) between the distance of exemplars from the category boundary for IT and EVC for different numbers of reconstructed dimensions. The third table shows correlation (Spearman's ρ) between the distance of exemplars from the category boundary for IT and EVC and human RT for both correctly and incorrectly classified exemplars for different numbers of reconstructed dimensions. The fourth table shows correlation (Spearman's ρ) between the distance of exemplars from the category boundary for IT and EVC and human RT for only correctly classified exemplars for different numbers of reconstructed dimensions.

*$p < .05$.

**$p < .01$.

***$p < .001$.

and nonresponse trials were excluded from analysis (5.3% of trials). Each participant's RT data was normalized to eliminate variation between participants. From this data, we computed the average normalized RT for each image in the data set for the analysis.

Statistically Evaluating Correspondences between Representational Distance from Decision Boundary and RT

To evaluate statistical significance, we computed the Spearman's rank order correlation between individual exemplar representations' distance from decision boundary and the exemplars' normalized RT from the behavioral study. This correlation value was then compared with a null distribution of correlation values generated using the following procedure: (i) the labels in the RDM were randomly shuffled, (ii) MDS was used to project the RDM data into a low-dimensional space (see above), (iii) LDA was used to determine the discriminant axis and decision boundary, (iv) the distances of individual exemplar representations from the decision boundary were computed, and (v) these distances were correlated with the behavioral RTs. Note the procedure is identical to the procedure for computing the actual correlation with the exception of shuffling the labels in the RDM. The procedure was repeated 1000 times to generate the null distribution of correlation values. The reported p values are based on comparing the actual correlation to the null distribution.

Sequential Analysis Model

We implemented a sequential probability ratio test (SPRT) model to show how representational distance can translate to RT, the behavioral measure, using a biologically plausible model that has previously been applied successfully to neural date from animal models (see Gold & Shadlen, 2002). The SPRT model accumulates evidence for a decision by iteratively sampling from a probability distribution. In effect, this test is equivalent to the log-likelihood ratio test of SDT, except that the decision variable v updates at each time step by taking the sum of the ratio for each unit of evidence that has occurred until it reaches a predetermined threshold determined by the stopping rules. The model has four parameters: a starting value (the evidence for a decision at onset), a decision threshold (the amount of evidence necessary to reach a decision), a decision noise parameter for the sampling distribution, and a mean for the sampling distribution. Three of the four parameters were fixed in our implementation. The model was set to have an initial starting value to 0, reflecting an initial state in which there is zero evidence for a decision. The decision thresholds were calculated in the following way: the mean distance from the decision boundary for animate and inanimate exemplars was computed (animate = 0.267; inanimate = −0.240). These distances were used as the means of probability distributions with standard Gaussian noise ($\sigma = 1$) to construct the "average" category distributions for the animate and inanimate portions of the representational space. Following Gold and Shadlen (2007), the values for the stopping rules for the SPRT model were calculated based on the probability of a false alarm occurring for the two categories. We modeled this by determining the proportion of the average category distributions that extended to the other side of the decision boundary ($f_{\text{animate}} = 0.3854$; $f_{\text{inanimate}} = 0.3876$). The stopping rule for animate was then defined using the log-odds ratios of a false alarm: when $v \geq \log(1 - f_{\text{animate}}/f_{\text{animate}})$ terminate; and for inanimate: when $v \leq \log(f_{\text{inanimate}}/1 - f_{\text{inanimate}})$ terminate. When v was between these values, the model accumulated more evidence (i.e., another iteration of the model). This resulted in the following stopping rule values: animate = 0.4666, inanimate = −0.4574. Note negative values are arbitrary outputs from LDA, indicating the side of the representational boundary that an exemplar is located.

The parameter that varied was mean of the sampling probability distribution, which we computed directly from the distance of the exemplar's representation from the representational decision boundary. The exemplars' distances ranged from 0.4928 (animate) to −0.6047 (inanimate). As with the average distances, Gaussian decision noise ($\sigma = 0.05$) was added to construct probability distributions for each model iteration. Crucially, the likelihoods for a unit of evidence were constant and were determined by the averaged category distance distributions. Our rationale was as follows: by holding the stopping threshold values and likelihoods constant across exemplars, variation in model performance would be driven solely by differences in the location of the sampling distributions. Distributions for exemplars with a greater distance from the decision boundary, when sampled, would result in the model converging more quickly, as the ratios produced by the sample at each model iteration would be of a far more positive (or negative) value relative to the ratios produced by sampling from the distributions of exemplars closer to the decision boundary. Because the only parameter that we varied was the mean of the sampling distribution, greater distance from the boundary entails fewer model iterations because the units of evidence increase in size as representational distance increases. Assuming a negative correlation between representational distance and RT, it would follow that a greater number of iterations would positively correlate with RT; that is, given a negative correlation between representational distance and RT, and how we have fixed the parameters of the SPRT model, a positive correlation between model iterations and RT would follow automatically, as number of iterations is a direct reflection of variation in representational distance.

The model was run 100 times to generate a mean and a standard deviation for decision times (computed in units of model iterations) for each exemplar.

RESULTS

Categorization Performance: Human Categorization Accuracy Accords Well with IT, but Not EVC

In the behavioral study, participants were instructed to categorize the object exemplars "as quickly and as accurately as possible." Participant performed very well on the task averaging 95% correct. We first examined classification performance for IT and EVC's representation of the stimuli. LDA was used to determine the category boundary between animate and inanimate objects in the reconstructed representational space. As an illustration, Figure 1 shows the positions of the individual exemplar representations relative to the category boundary (dashed line) for a 2-D space reconstructed from IT data. Using a leave-one-out cross-validation procedure, LDA classification performance was nearly perfect in distinguishing animate and inanimate object exemplars (97% correct) in IT for an 8-D space (see Table 1 for reconstructions with other dimensionality; the following statistics are also based on an 8-D reconstruction). In EVC, however, classification performance was only just above chance (54% correct). That the IT data are in better agreement with participants' choice behavior than the EVC data makes sense in light of the fact that if humans are using representational space in IT to categorize exemplars, then a classifier applied to a reconstruction of this space is in effect using the same information (albeit encoded in the brain activity of distinct fMRI participants). Human performance was very high. Thus, the near ceiling performance of the classifier for IT and near chance performance for EVC suggest that representational distance in IT, but not EVC, is likely determining participants' choice behavior.

The category boundary derived from LDA was imperfect in that some stimuli were located on the wrong side of the boundary (i.e., incorrectly categorized), particularly in EVC. We included these cases, as the focus of the study was the distance from the boundary and to make the comparison between EVC and IT equivalent in terms of number of exemplars used in the analysis. Excluding incorrectly categorized exemplars did not impact the findings for IT but did impact the findings for EVC, which we discuss shortly (see Table 1).

Human RTs Correlate with the Distance of Individual Object Exemplar Representations from Representational Decision Boundary

We next measured the Euclidean distance of individual object exemplar representations from the animate category boundary and compared these distances to human RTs for categorizing the stimuli as animate and inanimate. Our hypothesis predicts a negative correlation between the distance of an exemplar's representation from the animate category decision boundary in human IT and human RTs to categorize the exemplars. The relationship between the distance of an exemplar representation from the boundary and behavioral RT for IT and EVC is shown in the scatter plots in Figure 2. In agreement with our prediction, we found a negative correlation between IT distance and RT (Spearman's $\rho = -.399$; $p < .001$, bootstrap test). This is consistent with the idea that object representations closer to the boundary are near the criterion and so will reflect greater difficulty in discrimination whereas representations far from the boundary will reflect easier to discriminate stimuli (Figure 1). Importantly, when analyzed separately animate (Spearman's $\rho = -.699$; $p < .001$, bootstrap test) but not inanimate (Spearman's $\rho = .025$; $p = .91$, bootstrap test) exemplars correlated with representational distance. We believe this reflects the fact that inanimacy is a negatively defined category, including those things that are not animate, and inanimate exemplars in the study were incredibly heterogeneous (e.g., fruits and vegetables, places, and human artifacts).

Interestingly, we also found a significant correlation between RT and distance from the decision boundary in EVC (Spearman's $\rho = -.267$; $p < .01$, bootstrap test), and similar to IT, there was an animacy-specific effect: IT distances of animate exemplars (Spearman's $\rho = -.494$; $p < .01$, bootstrap test), but not inanimate exemplars (Spearman's $\rho = .060$; $p = .75$, bootstrap test), from the boundary were negatively correlated with RT. This is consistent with Kriegeskorte et al. (2008), who found a strong category effect in IT, but a weaker effect in EVC. We similarly observed a correspondence between IT's and EVC's representations in that the distance from the animate category boundary in IT and EVC were positively correlated (Spearman's $\rho = .356$, $p < .001$). EVC's representation is thus also in good agreement with behavioral RT. However, these findings were not robust across reconstructions when restricted to correctly classified exemplars (see Table 1). Grouping all exemplars, the correlation between representational distance and RT was only significant for a 4-D reconstruction, and when restricted to animate exemplars was not significant for either the 6-D or 10-D reconstructions.

To further examine the impact of low-level feature representations on our findings, we analyzed the IT data, while controlling for the representation of the stimuli in EVC. EVC's influence was controlled for by conducting a partial correlation analysis between RT and distance from the representation boundary in IT, factoring out distance from the boundary in EVC. If the effect in IT is merely a reflection of EVC, then the observed IT effect should be greatly reduced. We found, however, a robust partial correlation between RT and distance from the decision boundary in human IT (Spearman's $\rho = -.344$; $p < .001$, bootstrap test). The animacy effect was also retained after controlling for EVC: animate objects (Spearman's $\rho = -.583$; $p < .001$, bootstrap test), inanimate objects (Spearman's $\rho = -.007$; $p = .854$, bootstrap test). This indicates that the observed effects in EVC are not sufficient to explain the human IT data.

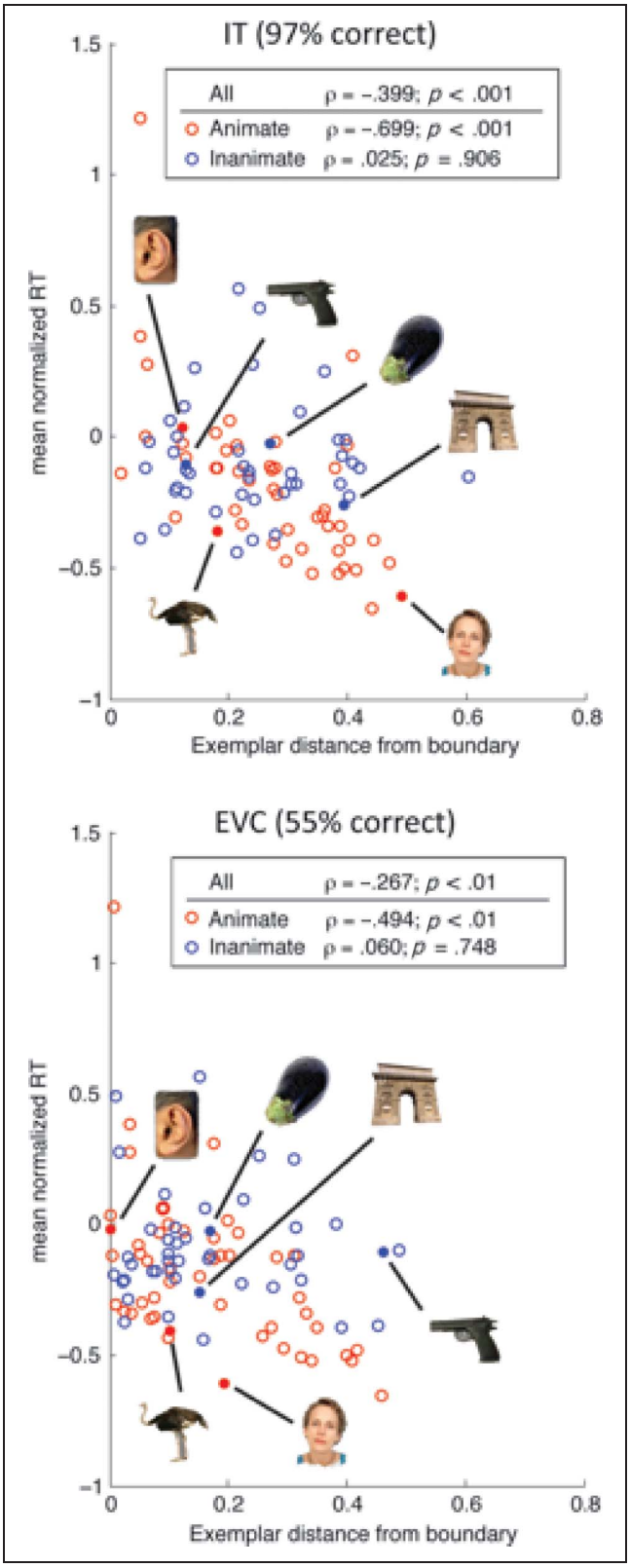

Figure 2. Correlation between distance from representational boundary and RT in IT and EVC. Scatter plots showing the relationship between exemplar distance from the representational decision boundary and RT. Red dots show animate objects; blue dots show inanimate objects. Six exemplars are given as examples (filled). Inset shows correlation values for the full data set and individual correlations for animate and inanimate objects.

Translating Representational Distance to RT Using Sequential Analysis

Behavioral RTs have been modeled successfully from LIP activity using sequential analysis (Gold & Shadlen, 2002, 2007). Representational distance is a measure that is static in time, whereas RT is a time varying behavioral measure. We next used sequential analysis to show how representational distances can be translated into RTs. Sequential analysis can be seen as a natural extension of SDT (Pleskac & Busemeyer, 2010). Where it differs from SDT is that, rather than modeling decision-making as a one-off process in which a decision is made for a single unit of evidence or input, sequential analysis models the decision-making process as allowing for the accumulation of evidence over time. Using the IT data, we used sequential analysis to translate the distances of individual exemplar representations from the representational border into a time-varying measure (i.e., model iterations). Figure 3 shows the results of the analysis. As expected given our finding of a negative correlation between representation distance in IT and RT, the data show that after converting the distances into model iterations behavioral RTs are well described by the model (Figure 3B). The overall correlation between behavioral RT and model RT was .357 ($p < .001$, bootstrap test), and as expected from the distinction between animate and inanimate objects describe earlier, the correlation for animate objects was strong ($\rho = .674$; $p < .001$, bootstrap test) and the correlation for inanimate objects was not significant ($\rho = -.002$; $p = .990$, bootstrap test). We further observed a significant correspondence between the variance in model RTs and population behavioral RTs (Figure 3C; $\rho = .280$; $p < .01$, bootstrap test; animate $\rho = .476$; $p < .001$, bootstrap test; inanimate $\rho = .023$; $p = .881$, bootstrap test).

DISCUSSION

In this study, we were motivated by the hypothesis that a representational space in human IT for different stimulus categories reflects not only a representational boundary but also a decision boundary for object categorization. On the basis of this hypothesis, we predicted that distance from the representational decision boundary in IT would predict RT when participants were asked to categorize stimuli as animate or inanimate. We reasoned further that the distance from the decision boundary could be used to construct probability distributions from which samples could be drawn for a sequential analysis model—the same type of model that has been successfully applied to animal data from LIP (Gold & Shadlen, 2002, 2007). Our prediction was born out: we observed a significant negative correlation between RTs and absolute distance from the boundary identified by LDA. Iterations of the sequential analysis model also correlated with the observed RTs and further showed an accord between the variance in human RT data and variance in model iterations (Figure 3).

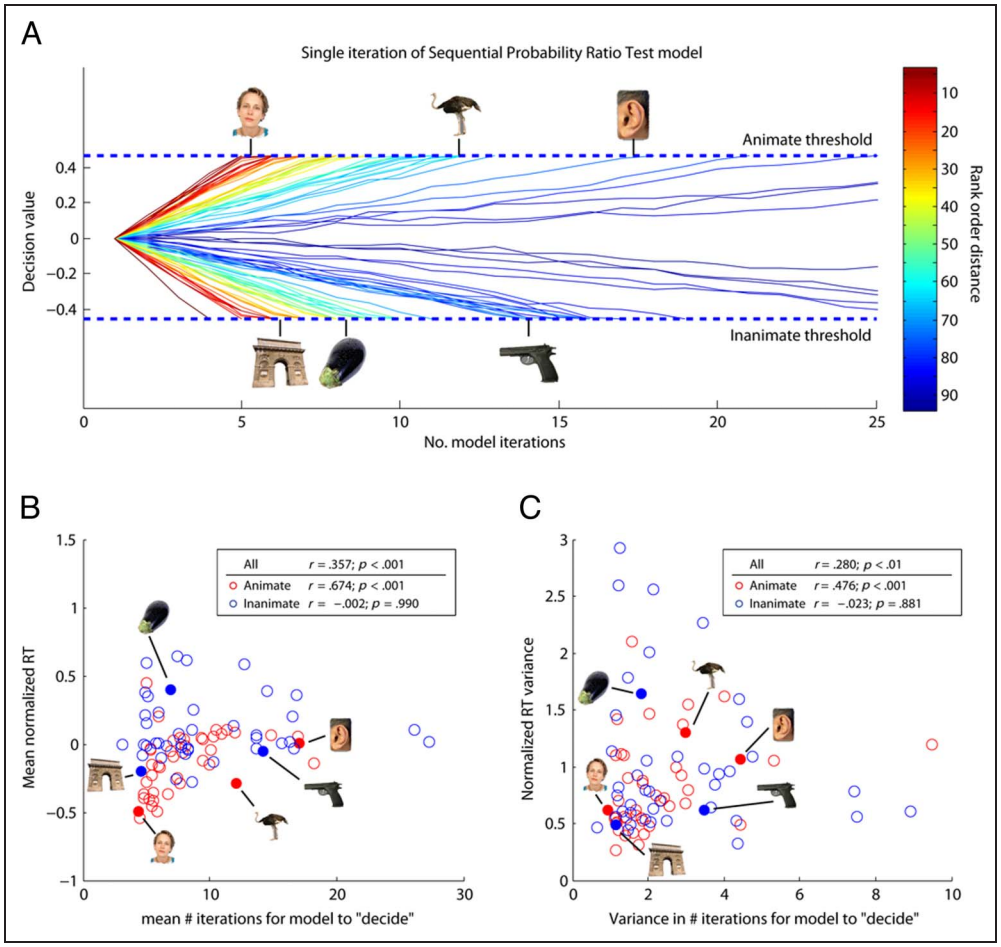

Figure 3. Sequential analysis model applied to representational distance data. (A) Shown are single trials of the SPRT model for each of the object exemplars. The *x* axis is the number of model iterations. The *y* axis is the decision value (accumulated evidence for a decision). The model terminates when the model reaches one of the threshold values for either animacy or inanimacy. Individual exemplars are colored according to their rank order distance from the representational boundary. The number of iterations for six exemplars are displayed. (B, C) Correlation between SPRT model and RT in IT. Red dots denote animate objects; blue dots denote inanimate objects. Six exemplars are given as examples (filled). Inset shows correlation values for the full data set and individual correlations for animate and inanimate objects. (B) Scatter plot shows the relationship between the mean number of iterations needed for the SPRT model to converge on a "decision" exemplar distance and the mean normalized RT. (C) Scatter plot shows the relationship between the variance in the number of iterations needed for the SPRT model to converge on a "decision" and the group variance in the normalized RT data.

The performance of the SPRT follows directly from the observed negative correlation between representational distance and RT. Other sequential analysis models, such as the diffusion model (Ratcliff & McKoon, 2008) or the leaky, competing accumulator model (Usher & McClelland, 2001), would presumably perform in a comparable manner if their parameters were set using simplifying distributional assumptions. Indeed, future research might look to remove such simplifying assumptions and instead fit different kinds of sequential analysis models to the actual distributions of representational distances and RTs across participants (for a review of different sequential analysis models, see Ratcliff & Smith, 2004). However, for present purposes, the importance of SPRT is theoretical and illustrative, as it suggests a different picture of perceptual deciding than the one proposed by those who have applied the model to animal single-cell data.

An important virtue of all sequential analysis models is that they distinguishes between the decision variable and the evidence/inputs to the decision process; although individual units of evidence are momentary, the decision variable is dynamic, changing in value as evidence accumulates (Gold & Shadlen, 2002). However, a question arises as to where and how the evidence and decision variable are realized and how in the brain a criterion is applied to the variable. Decision models have traditionally assumed an inner representation that is converted into a motor command and that the decision depends on the content of these representations. Roitman and Shadlen (2002) reasoned that LIP "reads out" inputs from motion processing areas like MT to produce the behavioral response and correspondingly successfully predicted monkey RT from neural activity in LIP. From this perspective, perceptual deciding is at least partially localized to a higher-level association area and perhaps reflects a type of embodiment of perceptual decision-making (Shadlen, Kiani, Hanks, & Churchland, 2008).

Our results suggest that the nature of a sensory representation is key to the decision-making process—representing is in part deciding. The representation of object exemplars in human IT contains a decision boundary for discriminating animate stimuli, and individual exemplar representations vary in their distance from this boundary. The observation that distance from the boundary is a factor in the decision-making process suggests that the form that a representation takes is a constituent part of the decision-making process. It is important to emphasize that participants in the fMRI experiment viewed the stimuli

although engaged in a distractor task and were never given any indication that the categories of animate and inanimate were relevant. These instructions only occurred in our behavioral RT experiment. Thus, our evidence for the representational boundary is independent of any overt decision-making, which provides strong support for the hypothesis that the process of settling on an internal representation of a stimulus is itself partially constitutive of, rather than before, decision-making in perception. At the same time, this independence constitutes a weakness of our findings, as strictly speaking we cannot infer a causal relation between representational distance and RTs because our fMRI and RT data sets came from different groups of participants. Still we would fully expect to see the same negative correlation if fMRI and RT data were collected from the same participants.

The significant negative correlation between RT and distances in EVC might seem to speak against the above interpretation—Are we suggesting that there is also a representational decision boundary for object categorization in EVC? We do not believe such a conclusion is warranted in light of the present results or the above theoretical interpretation. The brain's representation of the stimuli in EVC and IT obviously are not independent, as IT's representation is derived at least in part from the representation in EVC. This is reflected in Kriegeskorte et al.'s (2008) finding of a weak category effect in EVC that contrasted with a strong effect in IT. We similarly found distances between exemplar representations and the representational boundary in EVC and IT are correlated ($\rho = .3564$; $p < .001$, bootstrap test; see Table 1). However, there are a number of reasons why these findings do not support the idea that EVC contains a representational decision boundary for object categorization. First, both the percent correct classification of animacy and the correlation between distance and RT were substantially greater in IT than in EVC, with LDA performance for IT matching expectations given human choice behavior (Results; Figure 2; see also Table 1). Second, when analyzed separately, the results for correctly classified exemplars were far less reliable for EVC than IT across dimensions of reconstruction. Finally, the negative correlation between representational distance in IT and RTs was still robust after the results in EVC were controlled for using a partial correlation.

A difference in treatment of our results for IT and EVC is also warranted on theoretical grounds. DiCarlo and Cox (2007) have suggested that categorical representations can be characterized as "manifolds" in activation space (the collective pattern of neuronal activity in a particular brain representation). Importantly manifolds can be more or less "tangled." Technically, the representation of an object in the retina contains the information necessary to classify it as animate or inanimate. The retinal representation, however, is confounded by environmental variables (lighting, perspective, etc.), thus making it difficult to decode whether or not an object is animate. In an ideal representation, two smooth manifolds (reflecting two categories A and B) might cut through the activation space in parallel. A homunculus using the boundary between these manifolds would then be able to easily categorize stimuli as A or B; that is, the activity produced by the stimulus would be localized to one or the other manifold. Early in visual processing object manifolds are entangled, albeit to lesser extent than at the level of the retina, reflecting the weakness of the formatting of the information for making categorical decisions. Ostensibly some portions of a manifold are less entangled in EVC, possibly because of categorical biases in low-level stimulus features (Honey, Kirchner, & VanRullen, 2008). Exemplars producing activity localized to partially untangled portions of the manifold in EVC would have a subsequent advantage for categorization in IT. From this perspective, it is therefore not surprising that we can observe in EVC the origins of the perceptual decision-oriented formatting of IT's representational space. This interpretation of our results thus accords well with the theoretical notion that the broad goal of processing in the ventral visual pathways is to format information for efficient readout of object properties (DiCarlo, Zoccolan, & Rust, 2012; DiCarlo & Cox, 2007; VanRullen & Thorpe, 2002).

One important feature of the present analysis is that we used representational distance as a proxy for the unknown visual parameters of animacy, but we were nonetheless able to predict RT based on this proxy. This finding is notable in that as stated earlier there currently exists no clear way to parameterize objecthood akin to other perceptual variables like motion and tactile stimulation. Our study therefore suggests that representational distance can stand in as useful a proxy for future research, at least until a better characterization of objecthood is discovered. In our data, we found no significant correlation between distance and RT for inanimate stimuli. One way to interpret this negative finding is that it reflects the absence of an object manifold for inanimacy per se in IT. Future research might consider pursuing other manifolds (for places, artifacts, etc.), which were activated and are themselves distinct from each other. Furthermore, future research might attempt the same analysis in an A or B task, using two positively defined categories, rather than an A or not-A task as in this study.

Our findings also have significant methodological implications. First, our study provides one means of addressing the problem of how to fit perceptual decision models to brain data from humans collected by noninvasive procedures such as fMRI. Minimally, what is required is a boundary that would map onto a decision criterion. Traditional activation-based methods for analyzing fMRI data at most provide evidence of differential "involvement" of brain areas in a task and do not provide a means of modeling a decision boundary. However, information-based methods such as LDA attempt to uncover the organization of the information that is encoded in different brain regions (Mur, Bandettini, & Kriegeskorte, 2009): For such methods, what matters are the patterns of activation and their similarities and differences based on the category of

the stimulus that produced the response. Such details are obscured by traditional methods, which simply compare the average activity in an ROI to a control condition (cf. Carlson, Schrater, & He, 2003). A simple difference in activation levels in response to two different categories (such as animate or inanimate) does not identify any sort of boundary, which can be modeled as a decision criterion. In form, it is just such a boundary that is required by perceptual decision models. Thus, we believe that the present analysis suggests a promising new use for advanced fMRI analysis techniques, one that could be replicated and improved on by other researchers, working either on the visual system or other sense modalities, and perhaps expanded to other techniques such as EEG or MEG. Indeed, the present analysis might be applied to motion coherence tasks and tactile working memory tasks, which have been the focus of so much research on the neural underpinnings of perceptual decision-making. Finally, the present analysis speaks to an inferential weakness of advanced analysis techniques such as multivoxel pattern analysis, which only support the positive inference that a certain brain region encodes information regarding the stimuli; however, this does not necessarily entail that the brain is using such information in making a decision (Williams, Dang, & Kanwisher, 2007). The present finding of a connection between the representational decision boundary in IT and RT speaks to this issue regarding what inferences can be drawn from decoding analysis. For, if RT can be predicted from a representational boundary, this suggests that the information is not merely being encoded in that region but is being used by the brain to make the categorization response.

In summary, our findings show that the distance of an objects' representation from the representational boundary that distinguishes object categories accurately reflect the time necessary to categorize an object according to this boundary. Our findings support the theoretical notion that the form of a representation may partially constitute a decision and that one of the goals of visual processing in the ventral temporal pathway is to format visual information for making categorical distinctions between objects (DiCarlo & Cox, 2007). Importantly, these findings, taken in conjunction with earlier research showing stimulation of IT neurons biases behavioral responses for object recognition (Afraz et al., 2006), support IT's central role in decision-making for object categorization.

Reprint requests should be sent to Thomas A. Carlson, Centre for Cognition and Its Disorders, Macquarie University, Sydney, NSW 2109, Australia, or via e-mail: thomas.carlson@mq.edu.au.

REFERENCES

Afraz, S.-R., Kiani, R., & Esteky, H. (2006). Microstimulation of inferotemporal cortex influences face recognition. *Nature, 442,* 692–695.

Carlson, T., Schrater, P., & He, S. (2003). Patterns of activity in the categorical representation of objects. *Journal of Cognitive Neuroscience, 15,* 704–717.

Deco, G., Rolls, E., & Romo, R. (2010). Synaptic dynamics and decision-making. *Proceedings of the National Academy of Science, 107,* 7545–7549.

DiCarlo, J. J., & Cox, D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences, 11,* 333–341.

DiCarlo, J. J., & Maunsell, J. (2005). Using neuronal latency to determine sensory-motor processing pathways in reaction time tasks. *Journal of Neurophysiology, 93,* 2974–2986.

DiCarlo, J. J., Zoccolan, D., & Rust, N. (2012). How does the brain solve visual object recognition? *Neuron, 73,* 415–434.

Downing, P. E., Chan, A. W.-Y., Peelen, M. V., Dodds, C. M., & Kanwisher, N. (2006). Domain specificity in visual cortex. *Cerebral Cortex, 16,* 1453–1461.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification*. New York: Wiley.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *Journal of Neuroscience, 23,* 5235–5246.

Gold, J. I., & Shadlen, M. N. (2002). Banburismus and the brain: Decoding the relationship between sensory stimuli, decisions, and reward. *Neuron, 36,* 299–308.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience, 30,* 535–574.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research, 41,* 1409–1422.

Hanks, T. D., Ditterich, J., & Shadlen, M. N. (2006). Microstimulation of macaque area LIP affects decision-making in a motion discrimination task. *Nature Neuroscience, 9,* 682–689.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Piterini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293,* 2425–2430.

Heekeren, H. R., Marrett, S., Bandettini, P. A., & Ungerleider, L. G. (2004). A general mechanism for perceptual decision-making in the human brain. *Nature, 431,* 859–862.

Honey, C., Kirchner, H., & VanRullen, R. (2008). Faces in the cloud: Fourier power spectrum biases ultrarapid face detection. *Journal of Vision, 8,* 1–13.

Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science, 310,* 863–866.

Kiani, R., Esteky, H., Mirpour, K., & Tanaka, K. (2007). Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *Journal of Neurophysiology, 97,* 4296–4309.

Kiani, R., & Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science, 324,* 759–764.

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research, 46,* 1762–1776.

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., et al. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron, 60,* 1126–1141.

MacMillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.
Mur, M., Bandettini, P. A., & Kriegeskorte, N. (2009). Revealing representational content with pattern-information fMRI—An introductory guide. *Social Cognitive and Affective Neuroscience, 4,* 101–109.
Philiastides, M. G., & Sadja, P. (2006). Temporal characterization of the neural correlates of perceptual decision making in the human brain. *Cerebral Cortex, 16,* 509–518.
Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review, 117,* 864–901.
Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning, 2,* 509–522.
Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation, 20,* 873–922.
Ratcliff, R., & Smith, P. L. (2004). Comparison of sequential sampling models for two-choice reaction time. *Psychological Review, 111,* 333–367.
Roitman, J. D., & Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of Neuroscience, 22,* 9475–9489.
Romo, R., & Salinas, E. (2003). Flutter discrimination: Neural codes, perception, memory and decision making. *Nature Reviews Neuroscience, 4,* 203–218.
Shadlen, M. N., Kiani, R., Hanks, T. D., & Churchland, A. K. (2008). Neurobiology of decision making: An intentional framework. In C. Engel & W. Singer (Eds.), *Better than conscious? Decision making, the human mind, and implications for institutions* (pp. 71–101). Cambridge, MA: MIT Press.
Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology, 86,* 1916–1936.
Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences, 27,* 161–168.
Thorpe, S. J., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual cortex. *Nature, 381,* 520–522.
Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review, 108,* 550–592.
VanRullen, R., & Thorpe, S. J. (2002). Surfing a spike wave down the ventral stream. *Vision Research, 42,* 2593–2615.
Williams, M. A., Dang, S., & Kanwisher, N. (2007). Only some spatial patterns of fMRI response are read out in task performance. *Nature Neuroscience, 10,* 685–686.