

# At 120 msec You Can Spot the Animal but You Don't Yet Know It's a Dog

Chien-Te Wu<sup>1,2,3\*</sup>, Sébastien M. Crouzet<sup>1,2,4\*</sup>, Simon J. Thorpe<sup>1,2</sup>,  
and Michele Fabre-Thorpe<sup>1,2</sup>

## Abstract

■ Earlier studies suggested that the visual system processes information at the basic level (e.g., dog) faster than at the subordinate (e.g., Dalmatian) or superordinate (e.g., animals) levels. However, the advantage of the basic category over the superordinate category in object recognition has been challenged recently, and the hierarchical nature of visual categorization is now a matter of debate. To address this issue, we used a forced-choice saccadic task in which a target and a distractor image were displayed simultaneously on each trial and participants had to saccade as fast as possible toward the image containing animal targets based on different categorization levels. This protocol enables us to investigate the first 100–120 msec, a previously unexplored temporal window, of visual object categorization. The first result is a surprising stability of the saccade latency (median

RT ~155 msec) regardless of the animal target category and the dissimilarity of target and distractor image sets. Accuracy was high (around 80% correct) for categorization tasks that can be solved at the superordinate level but dropped to almost chance levels for basic level categorization. At the basic level, the highest accuracy (62%) was obtained when distractors were restricted to another dissimilar basic category. Computational simulations based on the saliency map model showed that the results could not be predicted by pure bottom-up saliency differences between images. Our results support a model of visual recognition in which the visual system can rapidly access relatively coarse visual representations that provide information at the superordinate level of an object, but where additional visual analysis is required to allow more detailed categorization at the basic level. ■

## INTRODUCTION

What can we perceive with just a glance at a scene? How long does it take to recognize an object? These questions have been important topics in studies of human vision since the 1970s. Early influential works (Potter, 1976; Biederman, Rabinowitz, Glass, & Stacy, 1974) suggested that the human visual system can extract a lot of information from scenes with presentation durations as short as ~100 msec. Further studies have shown that humans are fast and accurate at categorizing or detecting animals, vehicles, food objects, human or animal faces in natural scenes even when just flashed for 20–25 msec (Rousselet, Mace, & Fabre-Thorpe, 2003; VanRullen & Thorpe, 2001; Delorme, Richard, & Fabre-Thorpe, 2000; Thorpe, Fize, & Marlot, 1996). In these tasks, the earliest selective behavioral responses were observed at 250–270 msec after stimulus onset. These RTs include the time needed not only for visual processing of images but also for selecting and producing the motor response. In addition to this behavioral evidence, neurophysiological measures also

revealed an early scalp potential differentiation correlated with correct visual categorization responses at ~150 msec after image onset (Thorpe et al., 1996), a latency that cannot be reduced even with extensive training with the stimuli (Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2001).

Other studies demonstrated that this sort of rapid visual categorization can be performed in parallel for two images with virtually no significant behavioral cost (Rousselet, Thorpe, & Fabre-Thorpe, 2004; Rousselet, Fabre-Thorpe, & Thorpe, 2002). This ability to process in parallel led to the development of a new forced-choice saccade task (Kirchner & Thorpe, 2006) that allows behavioral responses to be recorded much earlier than in conventional paradigms using manual responses (Rousselet et al., 2003; VanRullen & Thorpe, 2001; Delorme et al., 2000; Thorpe et al., 1996). In the forced-choice saccadic task, participants are presented simultaneously with two images in the left and right hemifields and required to make a rapid saccade toward the image that contains a prespecified target object. In general, humans perform the task well (>80% correct responses) with the earliest reliable saccades occurring between 100 and 150 msec poststimulus depending on the target category and with faster responses to faces than with animals or vehicles (Crouzet, Kirchner, & Thorpe, 2010). This paradigm therefore can be seen as a valuable tool to investigate the early temporal processing dynamics

<sup>1</sup>Université de Toulouse, <sup>2</sup>CNRS, CerCo, Toulouse, France, <sup>3</sup>National Taiwan University, Taipei, Taiwan, <sup>4</sup>Charité University Medicine, Berlin, Germany

\*These authors contributed equally to this work.

underlying the access of visual object representations at different categorization levels.

In this study, we used this forced-choice saccade task to address a controversial question on the hierarchy of object categorization (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976): whether the visual system can access information at the basic level (e.g., dogs) faster than at the corresponding superordinate level (e.g., animals). A number of earlier studies suggest that basic level representations can be accessed faster than either the superordinate or the subordinate level (Murphy & Wisniewski, 1989; Murphy & Brownell, 1985; Jolicoeur, Gluck, & Kosslyn, 1984; Rosch et al., 1976). It has been suggested that such a basic level advantage could reflect both the similarities between its exemplars and their distinctiveness from other items within the same superordinate category (Murphy & Brownell, 1985). A more recent study (Grill-Spector & Kanwisher, 2005) even suggested that object categorization at the basic level does not require any more processing time than simple object detection; a highly controversial result (but see Mace, Joubert, Nespoulous, & Fabre-Thorpe, 2009; Bowers & Jones, 2008; Mack, Gauthier, Sadr, & Palmeri, 2008).

However, several recent findings have challenged the general basic level advantage in visual recognition processes by showing that, under certain circumstances, the processing at the superordinate level is faster than at the corresponding basic level (Mace et al., 2009; Joubert, Rousselet, Fize, & Fabre-Thorpe, 2007; Rogers & Patterson, 2007; Large, Kiss, & McMullen, 2004; Tanaka & Taylor, 1991). A potential explanation of these discrepancies could be that most of the earlier studies showing a basic level advantage used experimental paradigms that involve semantic processing (e.g., verbal report or naming association). For example, participants may be simultaneously presented with a drawing of an animal together with a word (e.g., “bird,” or “animal” or “robin”) and required to determine whether the drawing and word agree. However, in the studies that show no basic level advantage, participants typically have to respond whether a particular image belongs to a predefined category based mainly on visual analysis with little involvement of semantic processing (i.e., no object–name associations are required; Mace et al., 2009). One caveat for these studies is that manual RTs have relatively long latencies with mean RTs around 400–500 msec, although the earliest responses can be seen around 250–270 msec. This makes it difficult to tackle the nature of categorization hierarchies in an early temporal window of visual processing.

In the current study, we investigated further the hierarchical nature of rapid visual categorization using the forced-choice saccade task developed by Kirchner and Thorpe (2006). A key characteristic of this forced-choice saccade task is that the saccades to the side containing the target can be extremely fast (starting as early as 120 msec after stimulus onset for animal targets) when compared with conventional manual responses (Bacon-Mace, Kirchner,

Fabre-Thorpe, & Thorpe, 2007). Thus, it allows us to probe the efficiency of the human visual system on object categorization processes within a previously unexplored temporal window, much closer to stimulus onset. To closely simulate the perceptual challenge faced in real life and reduce potential confounding effects due to stimulus repetition, we used a large pool of natural photographs containing objects that can be either categorized at the superordinate level (e.g., animals) or at the basic level (e.g., dogs, birds, cats). In addition to the traditional comparison between the superordinate versus basic level categorization, we also investigated the influence of morphological similarity on the efficiency of visual categorization. Our results again challenge the traditional hierarchical view that object representations are first accessed at the basic level (or “entry level”). Instead, they demonstrate that basic level categorization is initially close to chance level whereas accuracy at the superordinate level can reach up to 80% correct. They also suggest that the morphological similarity both between- and within-object categories is a critical predictor of categorization performance. This view is consistent with a coarse-to-fine model of visual recognition where the visual system first gains access to relatively coarse visual information that can be used to activate superordinate object representations, but where additional processing is required to allow more detailed categorization at the basic level.

## METHODS

### Participants

There were 12 participants (6 men, 6 women, 20–33 years old) in Experiment 1 (Exp1) and 12 participants (4 men, 8 women, 21–33 years old) in Experiment 2 (Exp2; one participant performed both experiments). All participants provided their informed consent before participating in the experiment.

### Apparatus and Saccade Detection

Participants sat in a dimly lit room while stimuli were presented using a 19-in. CRT screen (Iiyama Vision Master PRO 454, resolution of 800 × 600 pixels; refresh rate of 100 Hz). They held their heads on a chin rest to maintain viewing distance at 60 cm. Image centers were 8.5° from the fixation cross, and the image size was 14° × 14°. Stimulus display and eye tracker control were done using Matlab and the Psychophysics Toolbox 3 (Brainard, 1997; Pelli, 1997). Eye movements were monitored with an IView Hi-Speed eye tracker (SensoMotoric Instruments, Berlin, Germany) that uses an infrared tracking system with a sampling rate of 240 Hz. Saccade detection was performed offline using SMI’s BeGaze Event Detection software (saccade based algorithm; Smeets & Hooge, 2003). Only the first saccade (if entering one of the two images) after each stimulus presentation was analyzed, and the

onset latency of this first saccade was defined as the saccadic RT (SRT). Before each block, we used the standard 13-point calibration procedure provided for the IView Hi-Speed Eye tracker. Calibration was confirmed before initiating each trial by requiring the participant to fixate the central cross.

## Procedures

Participants performed a saccadic choice task where two images were simultaneously displayed on each trial. At the beginning of each trial, observers had to keep their eyes fixated on a black fixation cross for a pseudorandom time interval (800–1600 msec), the fixation cross then disappeared for 200 msec before the presentation of the task-related image pair; this gap period allows faster initiation of saccades (Kirchner & Thorpe, 2006; Fischer & Weber, 1993). Two natural scene images, one as target and one as distractor, were displayed on each side of the screen for 400 msec, which allows faster average saccades than with briefly flashed stimuli (Crouzet et al., 2010). This longer presentation time was originally introduced to prevent the unwanted effect of stimulus offset on saccade initiation. Participants were required to make a saccade as fast and as accurately as possible to the side where the image contains the target object category. The background color was set to a midgray level.

## Tasks and Stimuli

The two experiments involved different image categories and thus different categorization tasks for the participants. In Exp1, we manipulated both the level of categorization and the morphological similarity within the sets of targets and distractors. There were four conditions (Figure 1): (1) Animal targets versus non-animal distractors (Ani/nonAni), (2) Dog targets versus non-animal distractors (Dog/nonAni), (3) Dog targets versus other-animal distractors (Dog/oAni), and (4) Dog target versus Bird distractors (Dog/Bird). The study was done as a block design. Before each block of 50 trials, the participant was explicitly told about the target category but also about the set of images that will be used as distractors. Participants were instructed to make saccades toward animal targets in Condition 1 and to dog targets in the three other conditions. The order of all experimental conditions was pseudorandomized across participants. There were 200 target images and 200 distractor images (from the Corel database and the Internet) for each condition (four blocks of 50 trials).

Two crucial factors can influence categorization performance: the morphological similarity among the target and distractor sets and the distance between the target and the distractor categories (Mohan & Arun, 2012; Duncan & Humphreys, 1989). On the basis of the proposal of Rosch et al. (1976), we define a basic category as a level where members of the category are both more similar to each other and more dissimilar to exemplars from other

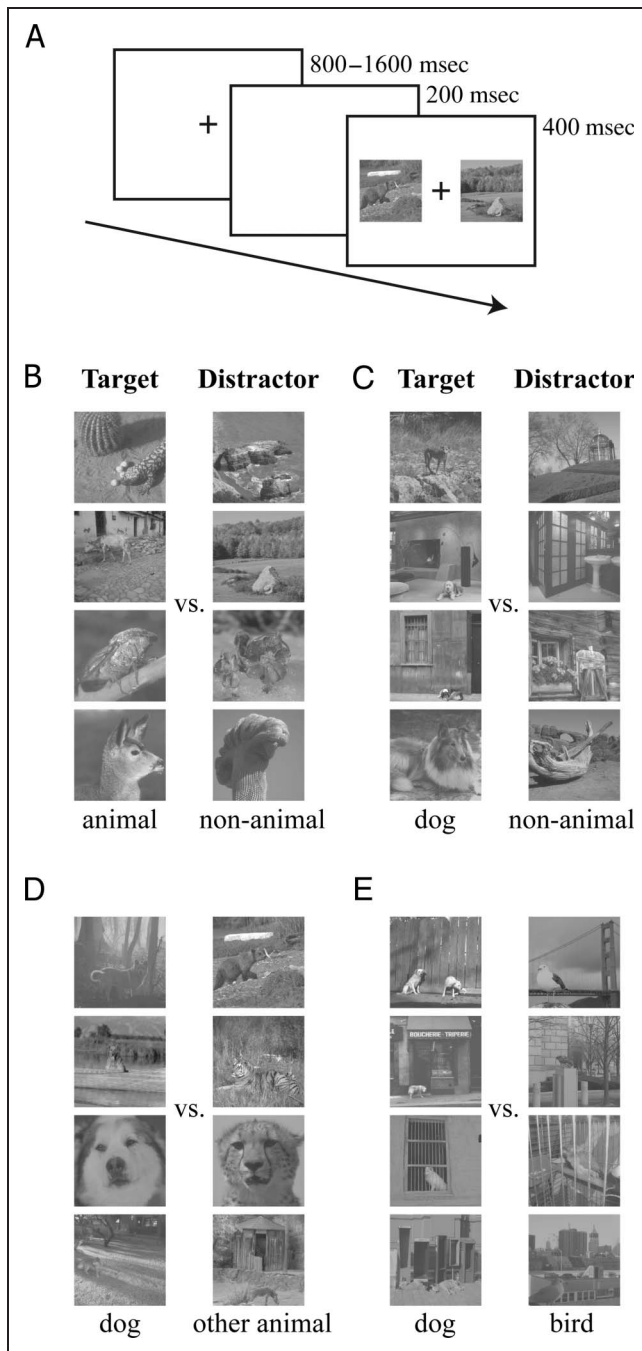
categories. Accordingly, in Exp1, the dog categorization was performed in three conditions. Comparing Conditions 3 and 4 (Dog/oAni vs. Dog/Birds) would reveal the effect of manipulating variability within the distractor set. On the other hand, comparing Conditions 2 and 3 (Dog/nonAni vs. Dog/oAni) would mainly reflect how the physical similarity between the target and the distractor sets affects performance. When targets are presented among distractors from the same superordinate category, the task has to be solved at the basic level. Condition 2 is therefore interesting: Because the dog targets were presented among non-animal distractors, the task can be solved at the superordinate level in which the target set variability was restricted from all animals to dogs (Figure 1).

In Exp2, morphological similarity between the target and the distractor sets was varied by using three different basic categories: Dogs, Cats, and Birds that differ in the number of shared features (e.g., legs, wings, fur, etc.). Dogs and cats (both four-legged mammals) were considered more similar than birds and dogs or birds and cats. There were six conditions: Dog target versus Bird distractor (Dog/Bird), Bird target versus Dog distractor (Bird/Dog), Dog target vs. Cat distractor (Dog/Cat), Cat target versus Dog distractor (Cat/Dog), Cat target versus Bird distractor (Cat/Bird), Bird target versus Cat distractor (Bird/Cat). There were 200 target images and 200 distractor images (from the Corel database and the Internet) for each condition (four blocks of 50 trials). Here again we used a block design, and participants were explicitly and verbally informed before each block about the target category (e.g., cats, dogs, or birds) they had to saccade to and the distractor category they had to avoid; the order of the conditions was randomized and counterbalanced across participants.

For both experiments, the proportions of images on natural and man-made backgrounds were balanced in the target and the distractor sets. We also matched the proportion of close-up, midrange, and long-range views. In addition, in Exp2, targets and distractors were paired on each trial according to the type of view and the type of background. For example, in the Dog/Bird condition, if a target image showed a dog head in a natural context, the corresponding distractor image would be a bird head in a natural context. All images were grayscaled and then equalized for mean luminance and RMS contrast using the average values obtained from the entire original set.

## Data Analysis

To assess the earliest latency at which stimulus information is available for selective behavior in a given task, we computed minimum SRTs (MinSRT) for each condition. This corresponds to the first time bin (of at least five consecutive 10-msec bins) in the SRT distribution in which correct responses significantly outnumber errors. This significance was assessed using a nonparametric bootstrap test, allowing 95% confidence intervals (CI) for the accuracy measurement to be obtained for each bin independently.



**Figure 1.** Experimental procedures and sample stimuli in Exp1. (A) Each trial started with a fixation cross. After a variable time delay ranging from 800 to 1600 msec, the fixation cross disappeared for 200 msec and a pair of task-relevant images were presented for 400 msec. Participants were required to make a saccade toward the correct target category as fast and accurately as possible. (B–E) Sample images (left column: target, right column: distractor) used in the four different conditions in Experiment 1: (B) Animal versus non-Animals, (C) Dog versus non-Animals, (D) Dog versus other Animals, and (E) Dog versus Bird.

When the lower bound of the CI was above chance level (here, 50% correct), correct responses were considered as significantly outnumbering errors.

To rule out the potential confounding influence of low-level salient features of images on the observed results,

we used the Saliency Toolbox 2.2 (Walther & Koch, 2006) running under Matlab to simulate the saccadic behavior under the pure influence of low-level saliency values. Using Saliency Toolbox 2.2, the input image was processed for low-level feature representations (pixel intensity, orientations, and color) at multiple scales. Because all images used in the current study were in gray-scale, we calculated only the intensity and orientation features. The resulting feature maps were integrated to form a saliency map that was used to predict the locations for the first saccade based on the highest saliency value.

We used the algorithm to simulate each trial that participants experienced during the real experiment. For each trial, the input image was almost the same as the display that participants observed during the real experiment (two images on a gray background), with only one exception that the borders of images were smoothed to prevent any inherent bias toward the borders. The location of the first saccade based on the highest saliency value was recorded. The algorithm has no knowledge of the task requirement and thus would make the saccades simply depending on the most salient location among the two images. But according to the task at hand, if the location was on the target side then it would be taken as a correct response, otherwise as an incorrect response. Such a simulation thus provides insights about the potential contribution of low-level features to the triggering of the saccadic responses observed in the current study.

## RESULTS

### Accuracy

In Exp1, although participants performed the tasks above the chance level of 50% in all four conditions (Figure 2A), the averaged accuracy in the superordinate categorization task (Ani/nonAni) and in the task that could be solved at the superordinate level (Dog/nonAni) was high ( $80 \pm 2\%$ ) and significantly higher ( $t(11) = 12.96, p < 10^{-7}$ ) than for the basic level tasks (Dog/Bird and Dog/oAni:  $60 \pm 1\%$ ). In addition, high degrees of perceptual similarity within either targets or distractors appear to facilitate participants' performance in forced-choice saccade tasks. Specifically, the performance in Dog/nonAni (that can be performed as a superordinate task with high target similarity) was significantly higher than Ani/nonAni (low target similarity;  $t(11) = 6.1, p < 10^{-4}$ ), and at the basic level, performance in Dog/Bird (high distractor similarity) was significantly higher than in Dog/oAni ( $t(11) = 2.24, p < .05$ ).

Furthermore in Exp2, we found a strong influence of target/distractor morphological similarity over the basic level categorization performance across conditions ( $F(5, 55) = 32.02, p < 10^{-6}$ ). Performance was significantly better with dog or cat targets mixed with bird distractors than when they were categorized against each other (Dog/Bird vs. Dog/Cat,  $t(11) = 7.06, p < .0001$ ; Cat/Bird vs. Cat/Dog,  $t(11) = 7.40, p < .0001$ ). Interestingly, however,



when birds were targets with dogs or cats as distractors, the performance dropped to chance levels (Bird/Cat:  $47 \pm 2\%$ , Bird/Dog:  $47 \pm 2\%$ ; Figure 2A). In other words, there seems to be a processing bias toward some object categories. Perhaps dogs or cats are processed more easily than birds, and this effect could counteract the impact of morphological differences between categories.

### Speed of Response

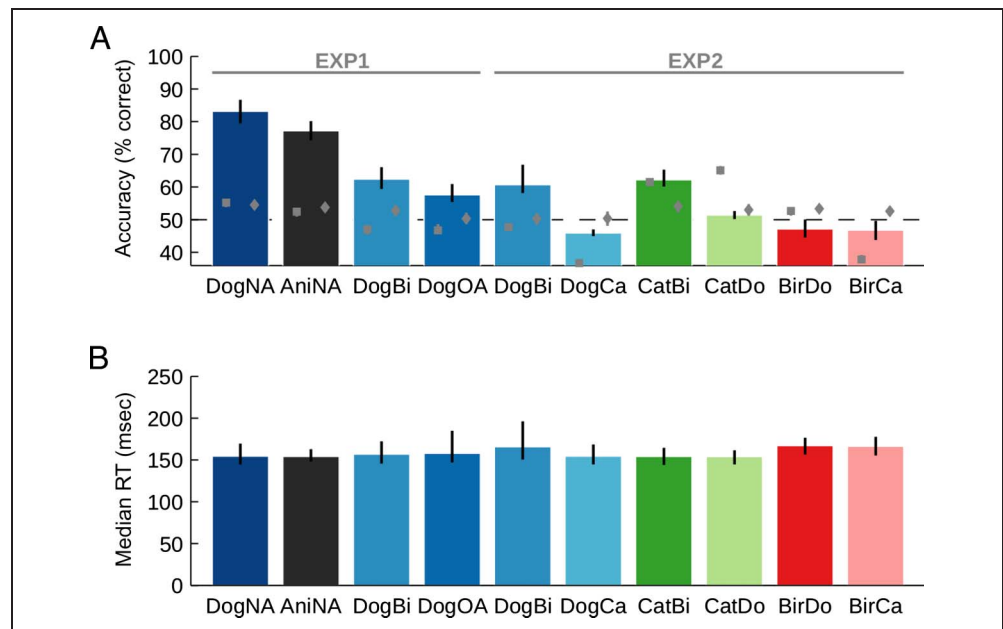
Interestingly, regardless of the categorization task, median RTs showed no difference between conditions: Exp1 (superordinate level: 154 msec vs. basic level: 158 msec,  $t(11) = 0.59$ ,  $ns$ ), Exp2 ( $F(5, 55) = 2.57$ , Greenhouse–Geisser corrected  $p = .11$ ; see Figure 2B).

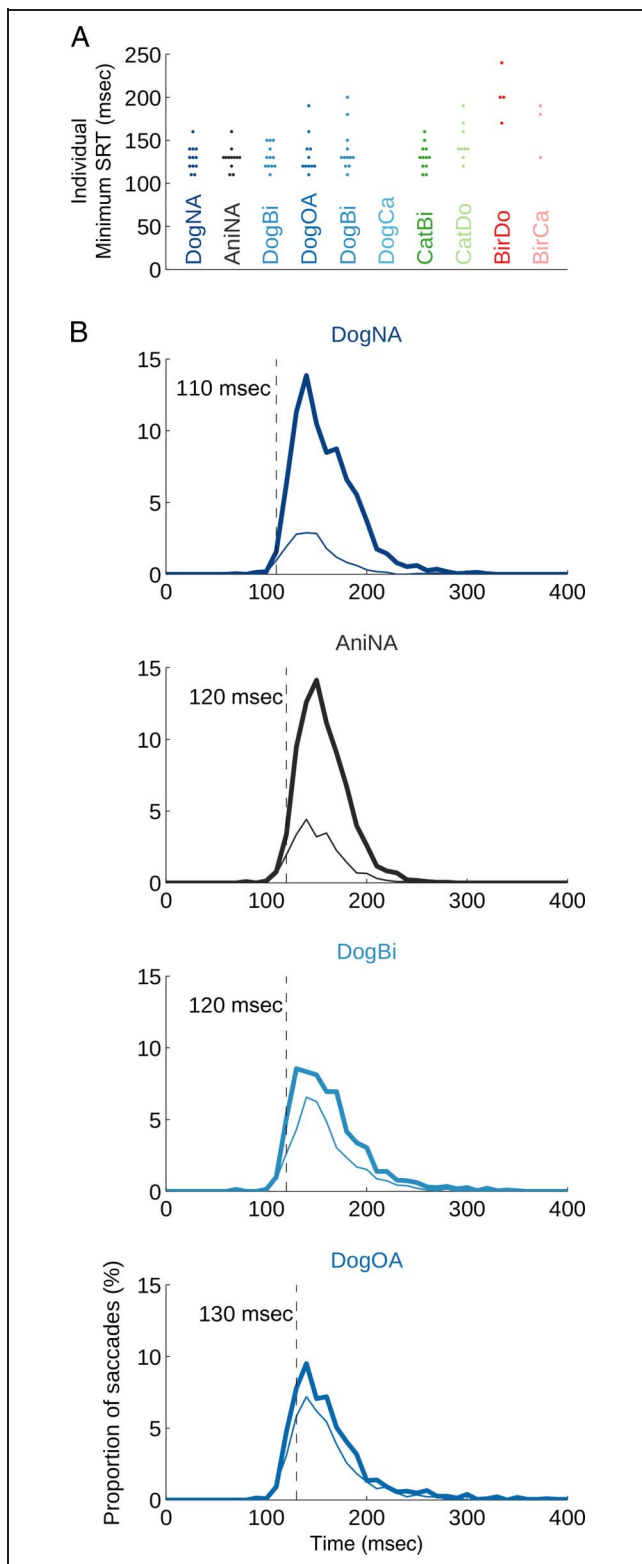
To further evaluate the ability of the visual system to categorize an object within the early processing time window accessible with the saccade task, we calculated a minimum SRT value (MinSRT) for each condition. The measure of MinSRT reflects the shortest processing time for the brain to reliably differentiate between two categories and initiate a saccade (i.e., correct responses significantly exceed incorrect responses, see Methods). As illustrated in Figure 3A, the MinSRTs were comparable (110 msec for the fastest individual participant) in all the conditions where it could be computed (i.e., when the participants' performance was above chance level, thus excluding Dog vs. Cat, Bird vs. Dog and Bird vs. Cat in Exp2).

In contrast to previous findings, which showed processing time advantages for the superordinate over the basic categorization in manual response tasks (e.g., 40–50 msec in Mace et al., 2009) coupled with very similar accuracy scores, the saccadic latencies observed here were remarkably similar for all the categorization tasks. On the other hand, whereas accuracy was good at the superordinate level (80% correct), we observed a large accuracy drop at the basic level of categorization. With saccadic decisions made at such short latencies (~120 msec in the current study), we can thus demonstrate a temporal window where information about superordinate object representations is available but judgments based on basic levels representations are almost at chance. These results thus provide strong evidence for a categorization advantage at the superordinate level and challenge the common view of a processing advantage at the basic level.

In addition, although both Dog/Bird and Dog/oAni belong to the same basic level categorization, the higher accuracy observed in the Dog/Bird condition over the Dog/oAni in Exp1 shows that large variations in performance can be observed at the same level of categorization. It seems plausible that they could be caused by different degrees of morphological similarity in the target and distractor category exemplars. Because distractor variability is lower when distractors are restricted to Birds than when they can include any non-dog animals, we hypothesize that bird features might be more easily determined and inhibited. This could result in reduced competition

**Figure 2.** Result summary for both Exp1 and Exp2. (A) Accuracy. Average accuracy across participants in all different tasks (error bars correspond to a bootstrapped 95% CI). Results from Exp1 suggest that, with limited processing time in the choice saccadic tasks, the object categorization processes are rarely completed at the basic level (DogBi: 62% and DogOA: 58%) as compared with the superordinate level (AniNA: 77% and DogNA: 83%). Results from Exp2 further show that, within the basic level categorization, participants can only perform the task above chance levels in DogBi (61%) and CatBi (62%) conditions, which reveals the influence of morphological similarities between targets and distractors in object categorization. The results from a trial-by-trial simulation based on the Itti and Koch Saliency Map model are represented as square and diamond dots. The square dots represent the averaged accuracy of the model to predict target location, and the diamond dots represent the averaged accuracy of the same model to predict participants' choices. The saliency analysis demonstrates that differences of low-level features between images are very unlikely to play a significant role in the behavioral results reported here. (B) Median RT. Average of median RT across participants and corresponding error bars (bootstrapped 95% CI). The median RTs were similar across all tasks and levels of categorization.





**Figure 3.** RT distributions and corresponding minimum SRT. (A) Minimum SRT computed from the distribution of SRT in each task and for each participant. (B) Histograms of correct (bold lines) and incorrect responses (thin lines) as a function of SRT using 10-msec bin size for the different tasks in Exp1, with corresponding minimum SRT (obtained here from the global distribution of SRT with all participants combined) indicated by the vertical dashed line.

between the target and the distractor representations. The morphological differences between dogs (e.g., four legs, fur) and birds (e.g., wings, feathers) were presumably very high; in contrast, when targeting Dogs among Other Animals, the distractors included a variety of different animals that occasionally shared similar morphological features to dogs (e.g., most mammals) and thus probably increased the categorization difficulty. Indeed, Mace et al. (2009) reported that when a distractor shares more common features with the target objects, it tends to increase the probability of producing a false alarm.

### Control of Potential Differences in Image Saliency

The chance level performance in both the Bird/Dog and Bird/Cat conditions was an unexpected result given that the two contrasted categories share the same degree of morphological similarity as the Dog/Bird and Cat/Bird conditions, suggesting that other factors might influence the participants' saccadic responses. Because the median response latency was very short in the current saccadic task (averaged median RT of 154–167 msec in Experiment 2) as compared with that observed in traditional manual tasks (~500 msec), it is reasonable to suspect that there might be some inherent low-level saliency differences between the selected images of different categories (e.g., cats or dogs were perhaps more “salient” than birds). Potentially this could generate perceptual biases toward a specific category, although all images were all transformed into gray-scaled images and equalized for mean luminance and global contrast. However, simulation results based on a saliency map model proposed by Itti and Koch (2000, 2001) indicate that the local saliency values cannot predict participants' saccadic performance, as demonstrated in the chance level prediction performance in Figure 2A, although Cat images seem to show slightly higher saliency values when compared with Bird images (please see SI Appendix for more details). Thus, saliency variations estimated with the model by Itti and Koch cannot account for any of the observed results.

### DISCUSSION

The aim of the current study was to investigate how object representations are activated during the first steps of object processing. The standard view is that basic level representations are the first to be activated and constitute the “entry level” (Rosch et al., 1976). However, using a go/no-go paradigm with manual responses, Mace et al. (2009) found that it is faster to categorize the object as an animal (superordinate) than as a dog. This study provides further evidence in support of the idea that visual information for superordinate categorization is available earlier than that for basic level categorization. Using a forced-choice saccade task involving much faster behavioral responses, our study showed that, in an early temporal window ~120 msec after stimulus onset, access to

basic level representations is poor even when task difficulties are reduced. In contrast, performance on a superordinate level categorization task is well above chance, even for the very fastest responses. Consistent with a previous study using a similar paradigm design (Kirchner & Thorpe, 2006), our results showed that participants were very good at making saccades toward the images containing animals paired with distractor images containing artificial objects but no animals. When restricting the animal target category to dogs and keeping the same distractors, the performance was very similar with even a slight increase of accuracy, suggesting that participants performed the task at the superordinate level and that reducing target variability improved accuracy. Critically, accuracy fell close to chance levels when participants were required to saccade toward dog images among distractors that contained other animals (basic level categorization). Restricting the distractor stimuli to only one category (e.g., birds in the current case) produced some increase in accuracy, but the effect was weak although statistically significant (62% in Dog/Bir vs. 58% in Dog/OA). Our results therefore suggest that there is an initial window of visual processing when you can “spot” the animal, but additional processing time is needed to make a judgment at the basic level.

A surprising finding from Experiment 2 was that the categorization performance seemed to differ for different categories at the basic level. Participants were better at making saccades toward dogs and cats than toward birds. One potential explanation would hypothesize that some low-level features might bias the saccadic responses toward one image set—images with dogs and cats might tend to be more salient than those with birds. However, this hypothesis was tested using Itti and Koch’s saliency map model (Itti & Koch, 2000, 2001), but no evidence was found for such differences. Specifically, our simulations show that their model cannot predict participants’ response pattern in this task, even for the fastest saccadic responses (i.e., the first quartile of the saccade responses in the SRT distribution) on which low-level saliency might be expected to have the strongest effect.

### **Distractor Impacts Accuracy but Not Response Speed**

An earlier study using the same saccadic choice paradigm found large differences in processing time for different target categories, with a MinSRT of 100 msec for faces, 120 msec for animals, and 140 msec for vehicles (Crouzet et al., 2010). Interestingly, in the current study, all the different conditions led to very similar MinSRT values (~120 msec), consistent with previous saccadic response studies using animals as targets (Crouzet et al., 2010; Bacon-Mace et al., 2007; Guyonneau, Kirchner, & Thorpe, 2006; Kirchner & Thorpe, 2006). The similar MinSRT across multiple conditions from different studies strongly suggests that the onset latency of saccade initiation might not be related to the difficulty of the task per se but might

rather depend on the nature of the superordinate target category. Specifically, task difficulty seems to only be reflected in the performance accuracy. In other words, in this task, it is as if the saccadic eye movement is initiated when the perceptual threshold for one alternative is reached regardless of the other alternative. When the distractors are more similar to simultaneously presented targets, the threshold might be reached faster for the distractor image than for the target image, inducing an accuracy drop. From a perceptual decision-making perspective, it suggests that in the saccadic choice task, the competition between alternatives is driven by a process of independent accumulation of information rather than mutual inhibition between two alternatives, as is often assumed in many protocols (Smith & Ratcliff, 2004).

One particularly intriguing finding was the observation that in the basic level tasks, participants were better at saccading to dogs and cats when paired with birds than the other way round. Indeed they were even slightly below chance when the target was a bird paired with either a dog or a cat. What might explain this bias in favor of dogs and cats? We have effectively ruled out the possibility that the bias could result from a low-level saliency difference between image sets. Could it be that we have an inbuilt bias toward mammals? Or could it be that we are more familiar with dogs and cats because they are frequently encountered as household pets? Clearly, further research will be needed to tease apart these various alternatives, but the difficulty of a given basic categorization task seems to be defined by both the target category and its relation to the distractor category.

### **Neuronal Latencies and the Saccadic Choice Task**

The fact that there were no RT differences between conditions here is indeed consistent with the idea that RTs are mainly driven by the latency of neurons selective for the superordinate target object category (Crouzet et al., 2010). In the current study, because the object category is always an “animal,” RTs do not vary between conditions. If we consider (i) that the decision in the saccadic choice task is based on the read-out of sensory populations selective for each superordinate object category involved (Heekeren, Marrett, & Ungerleider, 2008; Gold & Shadlen, 2007; Glimcher, 2003) and (ii) that this latency differs between object categories (e.g., in IT, Kiani, Esteky, & Tanaka, 2005), then a possible underlying mechanism to explain the present results could be that the neuronal populations selective for different basic categories are less separable than those selective for superordinate categories, a result that is suggested by cluster analysis of both monkey single unit studies and fMRI studies in humans (as suggested by cluster analysis, Kriegeskorte et al., 2008; Kiani, Esteky, Mirpour, & Tanaka, 2007). The performance in the choice saccade task would thus reflect the ability of the oculomotor system to “tune-in” to the relevant populations of sensory neurons. An interesting

remaining question concerns the generalization of the present results to other nonbiologically relevant object categories such as vehicles for example. A very recent study with a go/no-go manual task design provides some insight on this issue (Poncet & Fabre-Thorpe, 2014). In that study, our group found that a similar superordinate advantage can be obtained even with vehicles as testing stimuli, albeit with a smaller amplitude. Therefore, we would predict that the current results may well generalize to other object categories.

### From Superordinate to Basic Category

This study appears to support a “coarse-to-fine” temporal dynamic during the visual processing of objects. In contrast to earlier studies that suggested a processing advantage at the basic level as compared with its corresponding superordinate level (e.g., Large et al., 2004; Gauthier, Anderson, Tarr, Skudlarski, & Gore, 1997; Murphy & Wisniewski, 1989; Murphy & Brownell, 1985; Jolicoeur et al., 1984; Rosch et al., 1976), the present findings indicate that information about the superordinate category is available earlier during visual object processing. Indeed, one recent study revealed not only that explicit reports of image details become more common with increasing image exposure duration (i.e., information gathering), but that the visual system took more time to recognize objects at the basic level than at the superordinate level (Fei-Fei, Iyer, Koch, & Perona, 2007). A possible underlying mechanism might be that an animal can be recognized at its superordinate level on the basis of a single feedforward sweep (which would be the window offered by the saccadic choice task), but that additional recurrent processing would be needed for any more detailed analysis, such as that required for basic and subordinate level categorization (Fabre-Thorpe, 2011). Interestingly, a recent study used TMS to interfere with visual processing at various latencies while participants had to classify pictures of animals at the basic level. In that study, performance for categorizing animals as mammals or birds was significantly impaired when MS pulses were applied at 100 and 220 msec after image onset (Camprodon, Zohary, Brodbeck, & Pascual-Leone, 2010). Assuming that the 100-msec time window potentially corresponds to the first feedforward stage whereas the 220-msec time window corresponds to the recurrent stage of visual processing, their results indicate that visual processing based solely on the first feedforward sweep does not seem to enable a reliable object categorization at the basic level. Given the findings from our study, we would predict that object categorization at the superordinate level might only be impaired when TMS pulses are applied during the first time window (i.e., ~100 msec).

A fundamental remaining question is whether object similarity and morphological differences can account for the small variations of performance within different levels of categorization (such as the performance variations between different basic level categories) or if they could

actually account for the entire level of categorization effect. The data presented here cannot yet address this question because quantitative measurements of the similarity between categories are not currently available. Further studies using quantitative scene representations (e.g., Kriegeskorte & Kievit, 2013) may open a way to understand the mechanisms underlying visual categorization.

### Acknowledgments

We would like to acknowledge the grant support from CNRS to C.-T. Wu and Délégation Générale pour l'Armement (DGA) and Fondation pour la Recherche Médicale (FRM) to S. M. Crouzet.

Reprint requests should be sent to Michele Fabre-Thorpe, CNRS CERCO UMR 5549, Pavillon Baudot CHU Purpan, BP 25202, 31052 Toulouse Cedex, France, or via e-mail: michele.fabre-thorpe@cerco.ups-tlse.fr.

### REFERENCES

- Bacon-Mace, N., Kirchner, H., Fabre-Thorpe, M., & Thorpe, S. J. (2007). Effects of task requirements on rapid natural scene processing: From common sensory encoding to distinct decisional mechanisms. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 1013–1026.
- Biederman, I., Rabinowitz, J. C., Glass, A. L., & Stacy, E. W., Jr. (1974). On the information extracted from a glance at a scene. *Journal of Experimental Psychology*, *103*, 597–600.
- Bowers, J. S., & Jones, K. W. (2008). Detecting objects is easier than categorizing them. *Quarterly Journal of Experimental Psychology*, *61*, 552–557.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.
- Camprodon, J. A., Zohary, E., Brodbeck, V., & Pascual-Leone, A. (2010). Two phases of V1 activity for visual recognition of natural images. *Journal of Cognitive Neuroscience*, *22*, 1262–1269.
- Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: Face detection in just 100 ms. *Journal of Vision*, *10*, 16.1–16.17.
- Delorme, A., Richard, G., & Fabre-Thorpe, M. (2000). Ultra-rapid categorisation of natural scenes does not rely on colour cues: A study in monkeys and humans. *Vision Research*, *40*, 2187–2200.
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*, 433–458.
- Fabre-Thorpe, M. (2011). The characteristics and limits of rapid visual categorization. *Frontiers in Psychology*, *2*, 243.
- Fabre-Thorpe, M., Delorme, A., Marlot, C., & Thorpe, S. (2001). A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *Journal of Cognitive Neuroscience*, *13*, 171–180.
- Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision*, *7*, 10.
- Fischer, B., & Weber, H. (1993). Express saccades and visual attention. *Behavioral and Brain Sciences*, *16*, 553–567.
- Gauthier, I., Anderson, A. W., Tarr, M. J., Skudlarski, P., & Gore, J. C. (1997). Levels of categorization in visual recognition studied using functional magnetic resonance imaging. *Current Biology*, *7*, 645–651.
- Glimcher, P. W. (2003). The neurobiology of visual-saccadic decision making. *Annual Review of Neuroscience*, *26*, 133–179.



- Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*, 535–574.
- Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition: As soon as you know it is there, you know what it is. *Psychological Science*, *16*, 152–160.
- Guyonneau, R., Kirchner, H., & Thorpe, S. J. (2006). Animals roll around the clock: The rotation invariance of ultrarapid visual processing. *Journal of Vision*, *6*, 1008–1017.
- Heekeren, H. R., Marrett, S., & Ungerleider, L. G. (2008). The neural systems that mediate human perceptual decision making. *Nature Reviews Neuroscience*, *9*, 467–479.
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*, 1489–1506.
- Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*, 194–203.
- Jolicoeur, P., Gluck, M. A., & Kosslyn, S. M. (1984). Pictures and names: Making the connection. *Cognitive Psychology*, *16*, 243–275.
- Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision Research*, *47*, 3286–3297.
- 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., Esteky, H., & Tanaka, K. (2005). Differences in onset latency of macaque inferotemporal neural responses to primate and non-primate faces. *Journal of Neurophysiology*, *94*, 1587–1596.
- 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., & Kievit, R. A. (2013). Representational geometry: Integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, *17*, 401–412.
- 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.
- Large, M. E., Kiss, I., & McMullen, P. A. (2004). Electrophysiological correlates of object categorization: Back to basics. *Cognitive Brain Research*, *20*, 415–426.
- Mace, M. J., Joubert, O. R., Nespoulous, J. L., & Fabre-Thorpe, M. (2009). The time-course of visual categorizations: You spot the animal faster than the bird. *PLoS One*, *4*, e5927.
- Mack, M. L., Gauthier, I., Sadr, J., & Palmeri, T. J. (2008). Object detection and basic-level categorization: Sometimes you know it is there before you know what it is. *Psychonomic Bulletin & Review*, *15*, 28–35.
- Mohan, K., & Arun, S. P. (2012). Similarity relations in visual search predict rapid visual categorization. *Journal of Vision*, *12*, 19.
- Murphy, G. L., & Brownell, H. H. (1985). Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 70–84.
- Murphy, G. L., & Wisniewski, E. J. (1989). Categorizing objects in isolation and in scenes: What a superordinate is good for. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 572–586.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.
- Poncet, M., & Fabre-Thorpe, M. (2014). Stimulus duration and diversity do not reverse the advantage for superordinate-level representations: The animal is seen before the bird. *European Journal of Neuroscience*, *39*, 1508–1516.
- Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning & Memory*, *2*, 509–522.
- Rogers, T. T., & Patterson, K. (2007). Object categorization: Reversals and explanations of the basic-level advantage. *Journal of Experimental Psychology: General*, *136*, 451–469.
- Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439.
- Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nature Neuroscience*, *5*, 629–630.
- Rousselet, G. A., Mace, M. J., & Fabre-Thorpe, M. (2003). Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *Journal of Vision*, *3*, 440–455.
- Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004). How parallel is visual processing in the ventral pathway? *Trends in Cognitive Sciences*, *8*, 363–370.
- Smeets, J. B., & Hooge, I. T. (2003). Nature of variability in saccades. *Journal of Neurophysiology*, *90*, 12–20.
- Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, *27*, 161–168.
- Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology*, *23*, 457–482.
- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*, 520–522.
- VanRullen, R., & Thorpe, S. J. (2001). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artificial objects. *Perception*, *30*, 655–668.
- Walther, D., & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks*, *19*, 1395–1407.