

Humans and monkeys share visual representations

Denis Fize^{a,b,1}, Maxime Cauchoix^{a,b}, and Michèle Fabre-Thorpe^{a,b}

^aUniversité de Toulouse, Université Paul Sabatier, Centre de Recherche Cerveau et Cognition, 31062 Toulouse, France; and ^bCentre National de la Recherche Scientifique, Centre de Recherche Cerveau et Cognition, Faculté de Médecine de Rangueil, 31062 Toulouse, France

Edited by Robert Treichler, Kent State University, Kent, OH, and accepted by the Editorial Board March 23, 2011 (received for review November 3, 2010)

Conceptual abilities in animals have been shown at several levels of abstraction, but it is unclear whether the analogy with humans results from convergent evolution or from shared brain mechanisms inherited from a common origin. Macaque monkeys can access “non-similarity-based concepts,” such as when sorting pictures containing a superordinate target category (animal, tree, etc.) among other scenes. However, such performances could result from low-level visual processing based on learned regularities of the photographs, such as for scene categorization by artificial systems. By using pictures of man-made objects or animals embedded in man-made or natural contexts, the present study clearly establishes that macaque monkeys based their categorical decision on the presence of the animal targets regardless of the scene backgrounds. However, as is found with humans, monkeys performed better with categorically congruent object/context associations, especially when small object sizes favored background information. The accuracy improvements and the response-speed gains attributable to superordinate category congruency in monkeys were strikingly similar to those of human subjects tested with the same task and stimuli. These results suggest analogous processing of visual information during the activation of abstract representations in both humans and monkeys; they imply a large overlap between superordinate visual representations in humans and macaques as well as the implicit use of experienced associations between object and context.

homology | natural scenes | non-human primate | visual categorization

In a demonstration of monkeys' abstraction abilities, Bovet and Vauclair (1) reported that monkeys were able to classify new pairs of real objects as “same” when belonging to the same superordinate category (an apple and a banana, or a cup and a padlock), while classifying other combinations as “different.” The ability to perform a judgment of conceptual identity among categories such as Food or Tools corresponds to an abstract level of conceptualization (2–4), but the nature of the cerebral processes, the mental representations involved, and their similarity with those of humans remain unclear.

In monkeys, the ability to form and access perceptual classes is the most extensively studied abstraction level. Categorizing by perceptual similarity enables the formation of open-ended categories, which generalize to novel elements of the same kind (2, 3). This categorical behavior emerges spontaneously even in macaque monkeys with lesions of the lateral prefrontal cortex (5), a structure in which neurons show category-specific selectivity (6–8). In fact, building “perceptual concepts” (4) such as natural visual categories—trees, cats, dogs, etc. (6, 9)—or artificial ones such as letter categories (10, 11) could rely on relatively simple mechanisms: a set of representative visual features involving particular shapes or typical shadings could be used as diagnostic information to categorize objects at this perceptual level (12). Such processing in humans and monkeys is likely to take place in the inferior temporal cortex, which has long been known to be critical for object visual recognition (13, 14) and generalization of object views (15). However, single-neuron activity within this region does not seem to reflect categorical information (9, 16–18).

At a more abstract level, conceptual behavior implies categorizing beyond the physical similarity between exemplars of a class (3, 4). The few behavioral studies that have investigated this abstraction level in monkeys dramatically increased stimulus variety by using pictures of natural scenes and superordinate

categories such as Food or Animal to avoid diagnosticity from a restrained set of low-level visual cues (1, 19–22). In the Animal category, for example, the large variety of perceptually different instances of mammals, birds, insects, or fishes and the high performance reached on new exemplars suggest an abstract level of categorization and an ability for macaque monkeys to access superordinate representations (23).

However, the claim that abstract representations are used to perform the superordinate categorizations was challenged by the finding that global scene statistics can predict the presence of animals or objects in scene pictures (24). Indeed, to succeed in such tasks, monkeys had been extensively trained by using large image sets extracted from commercial databases (19–21). Because the animal images are much more likely to be pictured on natural scene backgrounds than in urban contexts, the scene statistics that support the distinction between natural and man-made environments (25, 26) and the presence of foreground objects could potentially allow performance above chance. Monkeys could have used such contextual regularities to reach high scores without any conceptual representation of the object category.

Here, we adapted for monkeys the approach used in earlier work investigating contextual effects on object processing during visual scene categorization by humans (27): animal target objects and man-made distractor objects were displayed randomly in either man-made or natural scene backgrounds. Using such stimuli should confuse the monkeys if they rely on global scene statistics to perform the task. Possible similarities between the mental representations used by humans and monkeys to solve such visual categorization tasks were also investigated by using the categorical object/context interference phenomenon. Indeed, in humans, categorization performance is affected in terms of both delayed reaction times and lower accuracies when the object and the background context belong to different superordinate categories (27–31), suggesting that the mental representations of object and background context overlapped during task performance and were “category-sensitive” at the superordinate level (27, 31).

Results show that monkeys' categorization performance primarily relied on the processing of animal/object information, and that the scene background did not play a major diagnostic role. Furthermore, the categorical interference between foreground objects and background contexts had very similar effects on monkey and human performance over a wide range of parameters. These data suggest a high analogy between monkeys and humans for the use of visual cues to access categorical representation and their cerebral processing.

Results and Discussion

Two macaque monkeys (Dy and Rx) and 11 human subjects performed the Animal/Non-Animal rapid scene-categorization task developed by Thorpe et al. (32) and adapted for monkeys by Fabre-Thorpe et al. (20). The response consisted of a button re-

Author contributions: D.F., M.C., and M.F.-T. designed research; D.F. and M.C. performed research; M.C. analyzed data; and D.F. and M.F.-T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. R.T. is a guest editor invited by the Editorial Board.

¹To whom correspondence should be addressed. E-mail: denis.fize@cerco.ups-tlse.fr.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1016213108/-DCSupplemental.

lease and a screen touch performed in under 1 s when a briefly flashed stimulus (50 ms) contained an animal (go/no-go task; Fig. 1A). The test stimuli were composed of 192 associations of four achromatic scenes counterbalancing superordinate context categories (man-made and natural) and superordinate object categories (animal or man-made object): very varied animal and man-made object exemplars were presented equally often in the natural and man-made background contexts. Test stimuli were only seen once by each subject, and backgrounds and objects were controlled for various low-level visual characteristics (*Materials and Methods* and Fig. 1B).

For monkeys, the sequence of stimuli randomly interleaved test stimuli with familiar scenes to ensure the stability of monkey motivation and performance (33); first-trial responses to the test stimuli were thus of crucial importance. Familiar scenes were taken from the commercial photographs on which monkeys initially learned the Animal/Non-Animal task and had been intermittently trained and tested (for 5–6 y) (*Materials and Methods*); both monkeys performed at 89% correct on these familiar stimuli during the days preceding the present experiment.

Non-Similarity-Based Concepts in Macaque. For both monkeys, first-trial performance on the test stimuli showed accuracy scores significantly above chance, regardless of the nature of the stimuli background (Fig. 1). The monkeys were able to use both man-made objects and animals as the pertinent key target for task performance: the proportion of hits outnumbered very significantly the proportion of false alarms in both monkeys (Fig. 1C). Monkeys were also able to ignore the scene background category because global-task accuracy was well above chance level, regardless of the environment category (Fig. 1D): natural (Rx, 71%; Dy, 67%) or man-made (Rx, 69%; Dy, 63%).

This task was particularly hard to perform given the high ratio of test stimuli introduced among familiar pictures (1/3 test, 2/3

familiar stimuli). In addition, in the test stimulus set, objects averaged only 6% of scene surface (range 0.2–22%), a much smaller object/scene surface range than in familiar stimuli. Object locations in the test stimuli were also extremely varied at eccentricities ranging from 0.2° to 24° from fixation (12° on average), which probably accounts for the conservative strategy exhibited by both monkeys (there were more misses than false alarms: both monkeys, χ^2 tests, $P < 10 \times 10^{-5}$) and the accuracy drop compared with previous studies of our group [usually 90% correct (20, 21)]. In fact, a similar drop in performance accuracy was also observed in the 11 human subjects performing the same task with the same test stimuli: they averaged only 80% correct, a performance that has to be compared with the 94% correct usually reached by humans on novel scenes (19, 32, 34).

Thus, despite their long training with commercial photographs that mostly associate animals with natural backgrounds, the two monkeys were clearly able to use object information and to ignore background information when categorizing new scenes. From the very first presentation of the manipulated stimuli, the bias attributable to the scene background category only accounted for ~2–5% of the global accuracy (Fig. 1D), despite the relatively small size of most objects within the scenes. Such immediate generalization for new man-made objects and animal exemplars presented in unusual scene contexts rules out the possibility that scene background regularities alone could explain performance.

Such results obtained in monkeys show that superordinate representations supporting abstract concepts do not necessarily require high-level functions such as linguistic abilities or even elaborate processing far from perceptual modalities (23, 35, 36). Recent results indicate that superordinate categories are the first to be accessed within the visual modality (37, 38). In fact, several processing schemes have been proposed that combine such coarse-to-fine visual processing (26, 39–41) with fast decision mechanisms. For example, coarse visual information could be

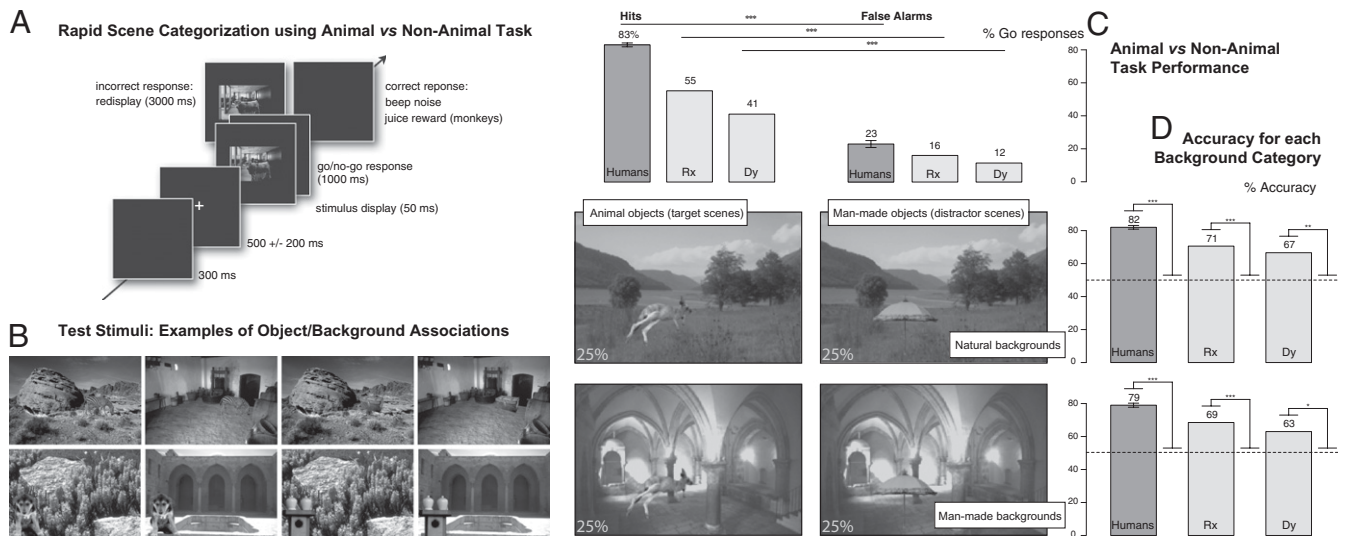


Fig. 1. Animal/Non-Animal categorization performance reached by human and monkey subjects on the first presentation of the test stimuli. (A) Animal/Non-Animal task. Stimuli were presented for three frames unmasked on a black screen. When a scene containing an animal was flashed, subjects had to release a button and touch the tactile screen within 1 s (target stimulus, go response); otherwise, they kept their hands on the button (distractor stimulus, no-go response). Correct (go and no-go) responses were rewarded by a noise (and a juice drop for monkeys); incorrect (go and no-go) responses were punished by the redisplay of the incorrectly categorized stimulus for 3 s. (B) Two examples of the 192 associations of four stimuli. For each association, the man-made and natural backgrounds had equal average luminance and RMS contrast, and the animal and man-made object vignettes had equal surface, center-of-mass location, luminance, and RMS contrast. (C) First-trial hit and false-alarm rates for the group of 11 human subjects and the macaque monkeys Rx and Dy. Bar histograms for hits and false alarms correspond to the stimulus illustrations in the left and right columns, respectively. (Humans: paired *t* test, $t = 26$, $df = 10$, $P < 10 \times 10^{-9}$; Rx: χ^2 test, $\chi^2 = 122$, $P < 10 \times 10^{-9}$; Dy: χ^2 test, $\chi^2 = 86$, $P < 10 \times 10^{-9}$.) (D) First-trial performance accuracy (pooling correct go and no-go responses) computed separately for natural and man-made background stimuli. Bar histograms for the accuracies on natural and man-made backgrounds correspond to the stimulus illustrations in the upper and lower rows, respectively. Dotted line represents chance level; error bars indicate SEM. (χ^2 tests, natural background—humans: contingency table, $\chi^2 = 269$, $n = 11$, $P < 10 \times 10^{-5}$; Rx: $\chi^2 = 13$, $P < 0.0003$; Dy: $\chi^2 = 9$, $P < 0.003$. χ^2 tests, man-made background—humans: contingency table, $\chi^2 = 327$, $n = 11$, $P < 10 \times 10^{-5}$; Rx: $\chi^2 = 11$, $P < 0.001$; Dy: $\chi^2 = 6$, $P < 0.02$.)

rapidly conveyed within the ventral pathway (21) or through the dorsal visual pathway (39) to frontal cortices (42) in order to prime object representations and facilitate subsequent detailed analysis in the ventral visual pathway. Another possibility is that midlevel areas in the ventral visual pathway might use intermediate representations that could be sufficient for categorical judgments (43–45). In both cases, coarse visual information is presumed to be sufficient to trigger prefrontal cortex activity that reflects the precise delineation of categorical boundaries relevant for task performance, as proposed by Freedman, Miller, and collaborators (6, 8).

Object and Context Congruency. Are superordinate visual representations similar in humans and macaques, and do they involve analogous mechanisms? To address these questions, we investigated the interference between object and background categories: we defined as categorically congruent the stimuli that associated congruent superordinate object/context associations (animals pasted on natural backgrounds or man-made objects on man-made backgrounds). Conversely, stimuli that embedded animals in man-made backgrounds or man-made objects in natural backgrounds were considered to be noncongruent (31).

From the very first stimulus presentations, both monkeys and humans exhibited 8% accuracy advantage for congruent compared with noncongruent test stimuli (Table S1). Similar object/context congruence biases have previously been described in humans for rapid scene categorization using color stimuli (27–31); such congruence effects were observed even on the earliest behavioral responses, leading to the suggestion that it could result from feed-forward facilitation between neuronal populations that are usually coactivated because of their selectivity to visual features that are highly likely to co-occur within our environment (27). Following this hypothesis, we predicted that this congruence bias would be robust to short-term practice independently from any improvement in task performance.

Monkey and human performances were thus compared for three consecutive sessions in which all test stimuli were seen only once a day. The monkeys further performed ad libitum for several daily sessions on test stimuli only (9 sessions for Dy and 15 for Rx). As illustrated Fig. 2A, we observed a significant accuracy increase with practice in both species, although it was less pronounced in monkey Dy. Such accuracy increase with scene familiarity could be because of a higher success rate on difficult stimuli (34) in particular for incongruent object/background associations. However, no interaction between congruence effect and task practice was observed in either species. In monkeys, a two-way ANOVA (congruence \times session) paired by subjects yielded a main effect for categorical congruence [$F(1, 43) = 90, P < 10 \times 10^{-4}, \eta^2 = 0.17$] and session [$F(10, 43) = 14, P < 10 \times 10^{-4}, \eta^2 = 0.26$], without interaction between congruence and session [$F(10, 43) = 2, P > 0.1, \eta^2 = 0.03$]. Similar ANOVA performed over the group of human subjects showed significant effects of categorical congruence [$F(1, 65) = 299, P < 10 \times 10^{-4}, \eta^2 = 0.39$] and session [$F(2, 65) = 90, P < 10 \times 10^{-4}, \eta^2 = 0.24$] and no interaction [$F(2, 65) = 3, P > 0.05, \eta^2 = 0.009$]. Similar results were obtained on reaction times (SI Materials and Methods). Thus, although global performance could improve, the impairment observed with incongruent object/background associations was not reduced with practice in either species. This result reinforces the hypothesis of hard-wired (dis)facilitation mechanisms between neurons selective for visual features (not) belonging to the same superordinate categories, as mentioned above.

Are scene statistics the visual features that account for background interaction with object category? We tested this hypothesis using Oliva and Torralba's classification model (24) on the stimuli of the current experiment. This model efficiently implements the major principles of scene gist recognition, considering the global scene features (25) used in a number of recent scene-recognition models (e.g., refs. 46 and 47). First, the simulation was successful in selecting the animal scenes within a set of familiar images that monkeys had categorized over the years (768

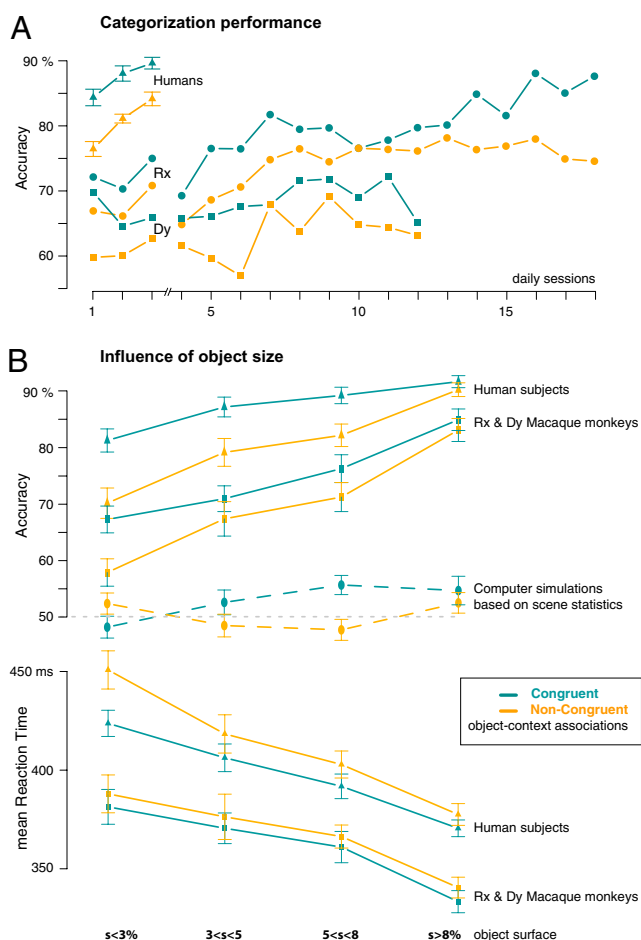


Fig. 2. Global Animal/Non-Animal task performance computed separately for categorical congruent (blue) and noncongruent (orange) object/context associations, reached by human (triangle) and monkey (circle and square) subjects. (A) Performance accuracy for consecutive sessions (test stimuli, all trials). Both species exhibited an accuracy increase with practice (linear regressions—humans: coefficient = 3.2, $R^2 = 0.39$, $df = 2$, $P < 10 \times 10^{-4}$; Rx: coefficient = 0.8, $R^2 = 0.79$, $df = 16$, $P < 10 \times 10^{-6}$; Dy: coefficient = 0.4, $R^2 = 0.24$, $df = 10$, $P = 0.06$) but no interaction between categorical congruence and practice (see text). (B) Accuracy and mean reaction time as a function of object surface assessed as a percentage of the whole image; stimuli were divided in four sets of equal size. The accuracy performance reached with computer simulations (dashed lines) was computed for the Animal vs. Non-Animal categorization task performed on the test stimulus set, after learning had been achieved by using the monkey's familiar training image set. Dotted line represents chance level; error bars indicate SEM between object/background associations.

photographs, model average accuracy 74%; note, however, that Rx and Dy scored 95% and 96%, respectively, on such familiar images). Second, it was successful in distinguishing between man-made and natural scene backgrounds by using the current 768 test stimuli (averaging 83% accuracy). These results stress the fact that processing image statistics could be a straightforward mechanism for using contextual information. However, the model failed completely at categorizing the manipulated test stimuli as containing animals: it only reached 53% accuracy on congruent and 50% on noncongruent object/context associations (average accuracy vs. chance level, χ^2 test, not significant; congruent vs. noncongruent, χ^2 test, $P < 0.005$).

This failure, however, could be accounted for by the small animal size relative to previous experiments. We tested this hypothesis by calculating simulation results by quartiles of the object sizes, corresponding to 2%, 4%, 7%, and 11% of the whole image.

Although model accuracy increased mildly with object surface, computer simulations only reached 54% accuracy on the largest objects: the model based on scene statistics failed in object categorization.

On the contrary, this analysis performed on subjects' behavioral data showed an important size effect that was similar in both species: performance improved with increasing object size for both accuracy and mean reaction time (Fig. 2B; ANOVAs are detailed in *SI Materials and Methods*). Interestingly, the accuracy advantage for congruent compared with noncongruent stimuli was highest for small objects (both species >9%), whereas the smallest accuracy advantage was observed for the biggest objects (both species <2%; Table S2 and ANOVAs in *SI Materials and Methods*). The level of interference between context category and object processing was thus similarly related in both species to the background and object surfaces.

These results suggest that background statistics could play a key role in the context interference observed when humans and monkeys categorize objects in natural scenes. Even if background statistics are not the diagnostic cues used by either species for object categorization, they could play a role analog to the contextual cueing reported by Chun and Jiang in visual search tasks (48): these authors showed that repeated associations between spatial configurations and target locations helped subjects in spatial attention tasks, despite the fact that subjects were not aware of these repeated associations and were not subsequently able to explicitly recognize them. In the present case, similar implicit learning could have occurred through repeated exposures to natural backgrounds associated with the presence of animals. Such implicit learning could thus involve low-level visual cues, including global scene statistics, consistent with their proposed role in triggering the fastest responses during context-categorization tasks (49).

These data suggest a high analogy between monkeys and humans in the way they use visual cues to access categorical representation, the underlying cerebral mechanisms, and their impact on behavior. They suggest the existence of analogous context and object superordinate representations in both species. This proposition and its consequences in terms of processing speed are further tested below.

Fast Access to Abstract Representations and Their Local Interaction. If the processing mechanisms responsible for object/context interactions are analogous in humans and monkeys, any advantage in terms of processing speed should be similar in the two species. More precisely, the underlying hypothesis of feed-forward facilitation implies that this processing speed advantage should be significant from the fastest responses. Indeed, in humans, previous investigations of the influence of object/context congruence on rapid go/no-go scene categorization (27, 31) reported that scenes in which the object conflicted with the surrounding scene backgrounds required an additional processing time of ~10–20 ms for the earliest responses. We thus took advantage of the severe time constraints imposed by the task to investigate whether the early response onset would be delayed when object and context categories are in conflict. We made the assumption that this delay previously documented in humans would also exist for macaques, despite the facts that the fastest reaction times were associated with the biggest objects, and the biggest objects were associated with the smallest congruence effect on average performance.

For this analysis, a large number of trials was needed that could not be restricted to first-trial performance. For both species, all go responses over all sessions were expressed by using 10-ms time bins. Minimal reaction time (corresponding to the minimal input–output processing time) was then determined as the first time bin for which correct responses significantly exceeded false alarms, using binomial tests. This temporal analysis was performed for each subject individually.

Individual results showed that monkeys Rx and Dy exhibited, respectively, 30-ms and 20-ms delays between minimal reaction time on congruent and noncongruent object/context associations

(Table S3 and Fig. S1). In humans, similar results were observed despite a higher variability (0- to 40-ms individual delays), but the reverse effect was never observed; this intrinsic individual variability was likely enhanced by the fact that human subjects performed far fewer trials than the monkeys did. Interestingly, in all subjects, the earliest false alarms were produced when natural backgrounds were presented: scene backgrounds efficiently biased behavior from the fastest responses.

Pooled across subjects (Fig. 3A), the distributions of go responses showed a general 50-ms advantage for monkeys compared with humans. However, an equal delay of 30 ms in minimal reaction time was observed in both species between the responses to congruent and noncongruent stimuli (monkey, 200 vs. 230 ms; human, 250 vs. 280 ms): when the object category conflicts with the surrounding context, the additional processing time is similar in the two species from the earliest responses. This finding was further observed with a d' analysis designed to evaluate how accuracy varies with response latency independently of the subject's strategies. Stimulus detectability was computed over time by using cumulative d' scores, by computing $d' = z(\text{hit}) - z(\text{fa})$ by 10-ms time bin (Fig. 3A *Inset*). Data showed that performance with congruent object/context associations reached higher d' values than with noncongruent ones with a similar temporal shift toward shorter response latencies for humans and monkeys. Considered in the frame of signal detection theory, this result indicates that human and macaque cerebral mechanisms are equally sensitive to the congruence of the visual features that determine object and context categories.

To quantify how object size could affect the congruence effect from the fastest responses, we computed minimal reaction times with stimuli containing either the largest or the smallest objects (Fig. 3B). Between the earliest responses to small and large objects, a 30-ms delay was observed in both species, showing that “larger is faster” (50) not only on average performance but also for the fastest responses. For these large objects, object/context incongruence delayed the earliest responses by 10 ms in both species (Fig. 3C); this value can be compared with the 30-ms average delay for all object sizes in both species reported above.

Such similar delays in monkeys and humans are surprising. Direct behavioral comparisons have, until now, reported macaque response latencies at about two-thirds of the human ones. This principle seems to hold for ocular and manual responses for both absolute latencies as well as for the delays induced by task manipulations (21, 51, 52). A convincing explanation could be the slow speed of intracortical connections that makes brain size a critical factor in determining the time of information transfer between cortical areas (20, 21). If it is the case, the similar delays in humans and monkeys responses reported here are the likely signatures of analogous local computations or, in other words, analogous mechanisms that are embodied in restricted cortical regions all along the processing pathways. For example, the above 30-ms delay related to object size observed during object categorization by human and monkey plausibly reflects similar neural integration mechanisms in the ventral pathways of the two species: indeed Sripathi and Olson (53) recently reported that shape selectivity of macaque inferotemporal neurons could develop 30 ms earlier for large stimuli compared with small ones. Here we propose that the highly similar temporal dynamics of object/context interactions observed behaviorally are the signature of analogous fast visual mechanisms that locally process features for object and scene category.

In summary, these results demonstrate first that macaque monkeys can really perform Animal/Non-Animal categorization tasks based on animals as a category and that they can do the task irrespective of scene background content. Moreover, despite their faster reaction times, monkeys exhibited highly similar behavior to human subjects when facing object/context incongruence: there was a similar accuracy impairment in object categorization, similar reaction-time delays observed from the fastest responses, and similar sensitivity to the object/background surface ratio. In both species, the Animal representations used to perform the task

birds (23% and 15%), insects (2% and 6%), reptiles and amphibians (8%), and fish and crustaceans (10% and 6%). In the test stimuli, man-made objects included means of transport (16%), urban furniture (15%), house furniture (16%), kitchen and water containers (17%), tools and toys (13%), interior decoration (8%), and other various objects (15%). The familiar stimuli were commercial photographs in which animals were usually large (or even very large), centered, and focused by the photographer. In the test stimuli, a wide range of object sizes and locations were used. Views of all of the test stimuli with associated first-trial performances and a subset of familiar stimuli are available at <http://cerco.ups-tlse.fr/~denis/fizePNAS2011>. For further details about test stimulus generation, see *SI Materials and Methods*.

Procedure. In a session, human subjects performed the Animal/Non-Animal categorization task once using the 768 test stimuli presented randomly (about half an hour). Each human subject performed three sessions on a daily basis to assess the robustness of the behavioral measures with three repetitions. For monkeys, test stimuli were introduced progressively, intermixed with familiar stimuli. Twelve daily sessions were needed to record monkey performance on the complete test stimulus set presented three times. Monkeys were further tested on daily sessions (9 for Dy and 15 for Rx) using

the 768 test stimuli randomly presented ad libitum. Monkeys Dy and Rx performed, respectively, a total of 7,940 and 14,560 trials on test stimuli.

Computer Simulations. Simulations used the code distributed by Oliva and Torralba (24) using the default software parameters. Each simulated task included 500 simulations: the performance accuracy indicated in the text result from their average value (SEM ranged from 1.07% to 1.2%). Each simulation involved randomly shuffling the stimuli into two equal sets that were considered for the subsequent phases of learning and testing.

More detailed descriptions of the methods can be found in *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank Simon Thorpe, Rufin VanRullen, Leila Reddy, and Thomas Serre for valuable comments on the manuscript. Computer simulations were possible thanks to the code freely provided by Aude Oliva and Antonio Torralba. This work was supported by the Centre National de la Recherche Scientifique; the Université de Toulouse, Université Paul Sabatier; and the Fondation pour la Recherche Médicale. The Délégation Générale de l'Armement provided financial support to M.C.

1. Bovet D, Vauclair J (2001) Judgment of conceptual identity in monkeys. *Psychon Bull Rev* 8:470–475.
2. Herrnstein RJ, Loveland DH (1964) Complex visual concept in the pigeon. *Science* 146: 549–551.
3. Zayan R, Vauclair J (1998) Categories as paradigms for comparative cognition. *Behav Processes* 42:87–99.
4. Lazareva OF, Wasserman EA (2008) *Learning Theory and Behavior*, ed Menzel R (Elsevier, Oxford), pp 197–226.
5. Minamimoto T, Saunders RC, Richmond BJ (2010) Monkeys quickly learn and generalize visual categories without lateral prefrontal cortex. *Neuron* 66:501–507.
6. Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2001) Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* 291:312–316.
7. Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2002) Visual categorization and the primate prefrontal cortex: Neurophysiology and behavior. *J Neurophysiol* 88:929–941.
8. Cromer JA, Roy JE, Miller EK (2010) Representation of multiple, independent categories in the primate prefrontal cortex. *Neuron* 66:796–807.
9. Vogels R (1999) Categorization of complex visual images by rhesus monkeys. Part 1: Behavioural study. *Eur J Neurosci* 11:1223–1238.
10. Schrier AM, Angarella R, Povar ML (1984) Studies of concept formation by stump-tailed monkeys: Concepts humans, monkeys, and letter A. *J Exp Psychol Anim Behav Process* 10:564–584.
11. Vauclair J, Fagot J (1996) Categorization of alphanumeric characters by baboons (*Papio papio*): Within and between class stimulus discrimination. *Curr Psychol Cogn* 15:449–462.
12. Sigala N, Gabbiani F, Logothetis NK (2002) Visual categorization and object representation in monkeys and humans. *J Cogn Neurosci* 14:187–198.
13. Logothetis NK, Pauls J, Poggio T (1995) Shape representation in the inferior temporal cortex of monkeys. *Curr Biol* 5:552–563.
14. Tanaka K (1996) Inferotemporal cortex and object vision. *Annu Rev Neurosci* 19: 109–139.
15. Weiskrantz L, Saunders RC (1984) Impairments of visual object transforms in monkeys. *Brain* 107:1033–1072.
16. Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2003) A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *J Neurosci* 23: 5235–5246.
17. Kiani R, Esteky H, Mirpour K, Tanaka K (2007) Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *J Neurophysiol* 97:4296–4309.
18. Sigala N, Logothetis NK (2002) Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature* 415:318–320.
19. 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 Res* 40: 2187–2200.
20. Fabre-Thorpe M, Richard G, Thorpe SJ (1998) Rapid categorization of natural images by rhesus monkeys. *Neuroreport* 9:303–308.
21. Macé MJ, Richard G, Delorme A, Fabre-Thorpe M (2005) Rapid categorization of natural scenes in monkeys: Target predictability and processing speed. *Neuroreport* 16:349–354.
22. Roberts WA, Mazmanian DS (1988) Concept learning at different levels of abstraction by pigeons, monkeys, and people. *J Exp Psychol Anim Behav Process* 14:247–260.
23. Fabre-Thorpe M (2003) Visual categorization: Accessing abstraction in non-human primates. *Philos Trans R Soc Lond B Biol Sci* 358:1215–1223.
24. Oliva A, Torralba A (2001) Modeling the shape of the scene: A holistic representation of the spatial envelope. *Int J Comput Vis* 42:145–175.
25. Oliva A, Torralba A (2006) Building the gist of a scene: The role of global image features in recognition. *Prog Brain Res* 155:23–36.
26. Hughes HC, Nozawa G, Kitterle F (1996) Global precedence, spatial frequency channels, and the statistics of natural images. *J Cogn Neurosci* 8:197–230.
27. Joubert OR, Fize D, Rousselet GA, Fabre-Thorpe M (2008) Early interference of context congruence on object processing in rapid visual categorization of natural scenes. *J Vis* 8:11.
28. Davenport JL, Potter MC (2004) Scene consistency in object and background perception. *Psychol Sci* 15:559–564.
29. Davenport JL (2007) Consistency effects between objects in scenes. *Mem Cognit* 35: 393–401.
30. Fei-Fei L, Iyer A, Koch C, Perona P (2007) What do we perceive in a glance of a real-world scene? *J Vis* 7:10.
31. Joubert OR, Rousselet GA, Fize D, Fabre-Thorpe M (2007) Processing scene context: Fast categorization and object interference. *Vision Res* 47:3286–3297.
32. Thorpe S, Fize D, Marlot C (1996) Speed of processing in the human visual system. *Nature* 381:520–522.
33. Minamimoto T, La Camera G, Richmond BJ (2009) Measuring and modeling the interaction among reward size, delay to reward, and satiation level on motivation in monkeys. *J Neurophysiol* 101:437–447.
34. 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. *J Cogn Neurosci* 13:171–180.
35. Barsalou LW, Kyle Simmons W, Barbey AK, Wilson CD (2003) Grounding conceptual knowledge in modality-specific systems. *Trends Cogn Sci* 7:84–91.
36. Humphreys GW, Forde EM (2001) Hierarchies, similarity, and interactivity in object recognition: “Category-specific” neuropsychological deficits. *Behav Brain Sci* 24: 453–476, discussion 476–509.
37. Macé MJ, Joubert OR, Nespoulous JL, Fabre-Thorpe M (2009) The time-course of visual categorizations: You spot the animal faster than the bird. *PLoS ONE* 4:e5927.
38. Fabre-Thorpe M, *The Making of Human Concepts*, eds Mareschal D, Quin PC, Lea S (Oxford Univ Press, Oxford).
39. Bullier J (2001) Integrated model of visual processing. *Brain Res Brain Res Rev* 36: 96–107.
40. Schyns PG, Oliva A (1994) From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. *Psychol Sci* 5:195–200.
41. Navon D (1977) Forest before trees: The precedence of global features in visual perception. *Cognit Psychol* 9:353–383.
42. Bar M, et al. (2006) Top-down facilitation of visual recognition. *Proc Natl Acad Sci USA* 103:449–454.
43. Mirabella G, et al. (2007) Neurons in area V4 of the macaque translate attended visual features into behaviorally relevant categories. *Neuron* 54:303–318.
44. Ullman S, Vidal-Naquet M, Sali E (2002) Visual features of intermediate complexity and their use in classification. *Nat Neurosci* 5:682–687.
45. Delorme A, Richard G, Fabre-Thorpe M (2010) Key visual features for rapid categorization of animals in natural scenes. *Front Psychol* 1:21.
46. Grossberg S, Huang TR (2009) ARTSCENE: A neural system for natural scene classification. *J Vis* 9:6.
47. Siagian C, Itti L (2007) Rapid biologically-inspired scene classification using features shared with visual attention. *IEEE Trans Pattern Anal Mach Intell* 29:300–312.
48. Chun MM, Jiang Y (1998) Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognit Psychol* 36:28–71.
49. Joubert OR, Rousselet GA, Fabre-Thorpe M, Fize D (2009) Rapid visual categorization of natural scene contexts with equalized amplitude spectrum and increasing phase noise. *J Vis* 9:2.
50. Vogels R (2009) Visual perception: Larger is faster. *Curr Biol* 19:R691–R693.
51. Busetini C, Fitzgibbon EJ, Miles FA (2001) Short-latency disparity vergence in humans. *J Neurophysiol* 85:1129–1152.
52. Fischer B, Weber H (1993) Express saccades and visual attention. *Behav Brain Sci* 16: 553–567.
53. Sripathi AP, Olson CR (2009) Representing the forest before the trees: A global advantage effect in monkey inferotemporal cortex. *J Neurosci* 29:7788–7796.
54. Perrett DI, Oram MW, Ashbridge E (1998) Evidence accumulation in cell populations responsive to faces: An account of generalisation of recognition without mental transformations. *Cognition* 67:111–145.
55. Macé M, Delorme A, Richard G, Fabre-Thorpe M (2010) Spotting animals in natural scenes: Efficiency of humans and monkeys at very low contrasts. *Anim Cogn* 13:405–418.