

# Supporting Information

Fize et al. 10.1073/pnas.1016213108

## SI Materials and Methods

**Test Stimuli.** Test stimuli were 768 gray-level scenes ( $576 \times 384$  pixels) that had the cut out of either an animal or a man-made object pasted onto either a natural or a man-made scene background. Part of the 384 man-made object vignettes had been collected and used by Joubert et al. (1) and were all extracted from either the Corel Stock Photo or Hemera Photo Objects libraries. Animal targets included fish, birds, mammals, reptiles, insects, and amphibians, whereas the man-made objects distractors included means of transport, city and house furniture, as well as various objects such as tools or monuments. The 384 picture backgrounds were mainly selected from the pool used by Joubert et al., with additional scenes selected because of their angle of view. Natural scene backgrounds included pictures from sea, sky, mountains, and forest, and landscape views from various geographical regions; man-made scene backgrounds included inside and outside urban views, such as building and street pictures, airports, or harbors.

The object vignettes and scene backgrounds were organized in 192 object and scene associations of four test stimuli (Fig. 1B). Each association consisted of an animal and a man-made object vignette selected for their similar real size or distance from the photographer, and a natural and a man-made scene background selected to have a similar depth of field. Each object vignette was pasted on both scene backgrounds, providing a set of four test stimuli. Thus, each of the 192 associations provided two object/context congruent stimuli (an animal pasted on a natural background, a man-made object on a man-made background) and two object/context noncongruent stimuli (the animal on the man-made background, the man-made object on the natural background).

Each four test stimuli corresponding to an association were generated simultaneously with a homemade image-processing software (written using Matlab). Particularly, within each four test stimulus associations, the natural and man-made scene backgrounds were adjusted to equalize average luminance and RMS contrast. The associated animal and man-made object vignettes were adjusted to equalize surface areas (alpha layer surfaces in pixels), locations (equal center-of-mass coordinates), average luminance, and RMS contrast. The luminance and contrast values were adjusted to the average values of the scene background surfaces that were occluded by the vignettes, so that the vignettes were fully and equally integrated to their scene backgrounds, with the goal of equalizing saliency. The 768 test stimuli were thus equalized pairwise for foreground objects and background scenes, as well as low-level statistics.

**Procedure.** Monkey subjects were placed in a semidark room, restrained in a primate chair (Crist Instruments) but with their heads free, and sat  $\sim 30$  cm away from a  $1,024 \times 768$  tactile screen (driven by the programmable VSG 2 graphics board; Cambridge Research Systems) controlled by a PC-compatible computer. The subjects placed one hand on a response pad located below the screen at waist level to start stimulus presentation. Pictures were then flashed centrally for 50 ms (three frames at 60 Hz, noninterlaced); the stimuli for monkeys were covering  $\sim 42 \times 28$  degrees of visual angle) on a black background, with a 1.5- to 2-s random intertrial interval between successive images. Because of human arm length and requested movement for behavioral response, the screen on which the same stimuli were displayed was placed at 50 cm away from human subjects (at this distance, stimuli covered  $25 \times 17$  degrees of visual angle). These brief presentations prevented exploratory eye movements and con-

strained the time available for information uptake. Subjects had 1 s to release the response pad and touch the tactile screen when they detected an animal in the flashed image (target stimuli, go response), otherwise they had to keep pressing the button (distractor stimuli, no-go response). Any response after 1 s was considered as a no-go response. A drop of fruit juice for monkeys and a beeping noise rewarded correct go- or no-go decisions. Incorrect go and no-go decisions were followed by a 3-s display of the incorrectly classified stimulus, delaying the next trial and the next possibility of reward and allowing time for ocular exploration. Reaction times (delay between stimulus onset and response pad release) and accuracy were recorded online; stimulus presentation and behavioral monitoring used custom software. Monkeys performed for as long as they wanted (typically 1 h for 1,000 trials) during a daily session.

For monkey subjects, test stimuli were introduced progressively, intermixed with a set of familiar stimuli used for training. The first 12 daily sessions were designed to this end: 100 familiar stimuli (from the pool of pictures categorized many times by the monkeys) were first used to stabilize monkeys' performance; then, 192 test stimuli were introduced randomly together with 384 familiar stimuli (2/3 in proportion); the last part of a session used more than 700 familiar stimuli so that monkeys could perform as long as they wanted until satiation. Stimulus randomization was recomputed for each session. The same 192 test stimuli (one-fourth of the total test stimulus set) were used for three successive sessions, so that, as for human subjects, the robustness of any behavioral effect could be assessed on three successive presentations. Twelve sessions were thus needed to record monkey performance on the complete test stimulus set presented three times. For clarity in the results and analyses, we pooled together the trials performed on the entire test stimulus set (for a given ordinal presentation).

Thereafter, the monkeys continued performing for a number of daily sessions (9 for Dy and 15 for Rx); these daily sessions only used the 768 test stimuli randomly presented ad libitum. On the test stimulus set, monkeys Dy and Rx performed a total of 7,940 trials and 14,560 trials, respectively.

**Data Analysis.** For each subject, trials with reaction times longer than 3 SDs were not considered for reaction-time analysis ( $\sim 0.1\%$ ). Validation of all ANOVAs was performed by using residual analysis (variance homogeneity and normality); reaction times were log-transformed. Global results of Fig. 2 were assessed using paired *t* tests, contingency tables, and ANOVAs. Global accuracy and mean reaction times as a function of object size (Fig. 3B) were first computed per image considering all subjects' trials; mean and SEM were further computed for each size (object surfaces grouped per quartiles). Minimal reaction times were computed by using binomial tests considering the cumulated average theoretical proportion of response for each 10-ms time bin.

**Effect of practice and object/background category congruence on response speed (Fig. 2A).** On reaction times that correspond to human and monkey performances illustrated in Fig. 2A, monkeys exhibited a small categorical congruence effect (two-way ANOVA, congruence  $\times$  session) [ $F(1, 43) = 9, P < 0.02, \eta^2 = 0.01$ ] and session effect [ $F(10, 43) = 26, P < 10 \times 10^{-4}, \eta^2 = 0.28$ ], without interaction between congruence and session [ $F(10, 43) = 1, P > 0.5, \eta^2 = 0.01$ ]. Similar ANOVA on mean reaction times performed over the group of human subjects showed significant but small effects of categorical congruence [ $F(1, 65) = 46, P < 10 \times 10^{-4}, \eta^2 = 0.01$ ] and session [ $F(2, 65) = 25, P < 10 \times 10^{-4}$ ],

$\eta^2 = 0.01$ ] and no interaction [ $F(2, 65) = 3, P > 0.05, \eta^2 = 0.002$ ]. Thus, for reaction time, as for accuracy, the performance impairment observed when object and background belonged to different man-made and natural categories was not reduced with practice in either species.

**Effect of object surface on categorization accuracy (Fig. 2B).** Two-way ANOVAs (object surface  $\times$  categorical congruence) yielded a strong main effect of object surface on performance accuracy in both species [humans:  $F(3, 87) = 231, P < 10 \times 10^{-4}, \eta^2 = 0.54$ ; monkeys:  $F(3, 15) = 1,847, P < 10 \times 10^{-4}, \eta^2 = 0.66$ ] and confirmed the main effect of categorical congruence [humans:  $F(1, 87) = 267, P < 10 \times 10^{-4}, \eta^2 = 0.21$ ; monkeys:  $F(1, 15) = 508, P < 10 \times 10^{-3}, \eta^2 = 0.06$ ]. The interaction between object surface and congruence was significant in both species [humans:  $F(3, 87) = 23, P < 10 \times 10^{-4}, \eta^2 = 0.05$ ; monkeys:  $F(3, 15) = 55, P < 0.005, \eta^2 = 0.02$ ]; the largest accuracy advantage for congruent compared with non-congruent stimuli was recorded for the smallest objects (both species  $>9\%$ ) compared with the biggest objects (both species  $<2\%$ ). Table S2 shows data and significance of the effect using paired *t* tests.

**Effect of object surface on reaction times (Fig. 2B).** As for accuracy, two-way ANOVAs (object surface  $\times$  categorical congruence) on reaction times showed similar trends. In both species, there were main effects of object surface [humans:  $F(3, 87) = 344, P < 10 \times 10^{-4}, \eta^2 = 0.22$ ; monkeys:  $F(3, 15) = 13, P < 0.004, \eta^2 = 0.18$ ]. In humans, categorical congruence affected reaction times [ $F(1, 87) = 79, P < 10 \times 10^{-4}, \eta^2 = 0.02$ ] with a small interaction with object

surface [ $F(3, 87) = 6, P < 0.003, \eta^2 = 0.003$ ]. No effect of categorical congruence nor any interaction with object surface were observed on response speed in monkeys [main:  $F(1, 15) = 0.85, P > 0.5, \eta^2 = 0.003$ ; interaction:  $F(3, 15) = 0.25, P > 0.8, \eta^2 = 0.004$ ].

**Computer Simulations.** Simulations used the code distributed by A. Oliva and A. Torralba (2), which is available at <http://people.csail.mit.edu/torralba/code/spatialenvelope/>. All simulated task performances used 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. The 768 familiar stimuli used for the monkey experiments (Corel Stock Photo) were used for a first simulation of the Animal vs. Non-Animal categorization task; the 768 test stimuli were used to simulate a natural vs. man-made scene context-categorization task. A simulated Animal vs. Non-Animal categorization task with the test stimuli was then used to simulate monkey's performance on the first presentations of the test stimuli. Thus, for each simulation, 384 stimuli were randomly selected within the pool of 768 familiar stimuli and used in the learning phase; 384 stimuli were randomly selected within the pool of the 768 test stimuli and used in the testing phase. We also checked that the software obviously failed to learn the Animal vs. Non-Animal task using the test stimuli only.

1. 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.

2. 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.



**Table S1. Individual results using first-trial performance on the test stimulus set**

Subjects	Accuracy, %			Median reaction time, ms			Mean reaction time, ms		
	Congruent	Noncongruent	Difference	Congruent	Noncongruent	Difference	Congruent	Noncongruent	Difference
Dy	69.8	59.8	10	386	403	-18	411	421	-10
Rx	72	66.8	5.2	324	367	-43	354	378	-25
H1	87.1	81.6	5.5	325	343	-19	343	362	-19
H2	78.9	70.4	8.5	411	412	-1	432	433	-1
H3	84.2	77.7	6.5	377	391	-14	392	396	-4
H4	80.2	75.1	5.1	405	403	2	422	424	-2
H5	80.8	73.9	6.9	366	376	-11	372	382	-10
H6	81.5	73.2	8.3	317	341	-24	332	351	-19
H7	81.1	74.2	6.9	469	453	16	480	471	9
H8	90.3	81.8	8.5	409	411	-2	425	431	-6
H9	88.5	77.3	11.2	352	349	4	363	361	2
H10	89.3	81.1	8.2	428	444	-16	443	448	-4
H11	86.6	74	12.6	330	330	0	348	354	-6
Mean									
Monkeys	70.9	63.3	7.6	355	385	-30	382	400	-17
Humans	84.4	76.4	8	381	386	-6	396	401	-6
SE									
Monkeys	1.1	3.5	2.4	31	18	13	29	21	8
Humans	1.2	1.2	0.7	15	13	4	14	13	2

Dy and Rx are macaque monkeys, and H1–11 are human subjects.

**Table S2. Global accuracy for each individual (all trials) as a function of object size**

Subjects	Object size congruence			
	<i>s</i> < 3% congruent	<i>s</i> < 3% noncongruent	<i>s</i> > 8% congruent	<i>s</i> > 8% noncongruent
<b>Humans</b>				
H1	85.8	74.5	94.4	94.4
H2	81.3	66.6	88.9	87.0
H3	84.5	70.0	89.2	88.4
H4	75.8	66.9	89.2	90.6
H5	78.6	69.6	85.8	86.4
H6	78.1	71.7	89.0	89.6
H7	74.9	67.9	89.2	87.1
H8	86.6	69.5	96.1	94.4
H9	77.6	70.6	95.1	89.6
H10	86.7	74.9	96.5	93.8
H11	82.5	68.8	95.1	91.9
Mean	81.1%	70.1%	91.7%	90.3%
SE	1.3%	0.8%	1.1%	0.9%
Congruent–noncongruent		<b>11.0%</b>		<b>1.4%</b>
Paired <i>t</i>		<i>P</i> < 10 × 10 <sup>-6</sup>		<i>P</i> < 0.022
<b>Monkeys</b>				
Dy	60.5	51.9	79.7	78.3
Rx	70.9	61.3	87.5	85.7
Mean	65.7%	56.6%	83.6%	82.0%
SE	5.2%	4.7%	3.9%	3.7%
Congruent–noncongruent		<b>9.1%</b>		<b>1.6%</b>
Paired <i>t</i>		<i>P</i> < 0.018		<i>P</i> < 0.031

Object size is indicated as a percentage of stimulus surface. Accuracy scores are significantly higher when object and context belong to congruent superordinate categories, but this congruence effect is much larger for small objects. Bold indicates the size of the congruence effect on accuracy scores. Dy and Rx are macaque monkeys, and H1–11 are human subjects.

