Early interference of context congruence on object processing in rapid visual categorization of natural scenes

Olivier R. Joubert

Denis Fize

Guillaume A. Rousselet

Michèle Fabre-Thorpe

Whereas most scientists agree that scene context can influence object recognition, the time course of such object/context interactions is still unknown. To determine the earliest interactions between object and context processing, we used a rapid go/no-go categorization task in which natural scenes were briefly flashed and subjects required to respond as fast as possible to animal targets. Targets were pasted on congruent (natural) or incongruent (urban) contexts. Experiment 1 showed that pasting a target on another congruent background induced performance impairments, whereas segregation of targets on a blank background had very little effect on behavior. Experiment 2 used animals pasted on congruent or incongruent contexts. Context incongruence induced a 10% drop of correct hits and a 16-ms increase in median reaction times, affecting even the earliest behavioral responses. Experiment 3 replicated the congruency effect with other subjects and other stimuli, thus demonstrating its robustness. Object and context must be processed in parallel with continuous interactions possibly through feed-forward co-activation of populations of visual neurons selective to diagnostic features. Facilitation would be induced by the customary co-activation of “congruent” populations of neurons whereas interference would take place when conflictual populations of neurons fire simultaneously.

Keywords: natural scenes, early scene processing, object categorization, object–context interactions, congruence

orientation and location of the object in a scene to make hypotheses about its identity when not given enough information (Oliva & Torralba, 2007).

The idea according to which a consistent context can facilitate object detection, recognition and naming is generally accepted (Biederman, Mezzanotte, & Rabinowitz, 1982; Boyce & Pollatsek, 1992; Boyce, Pollatsek, & Rayner, 1989; Palmer, 1975). These studies employed line drawing of scenes and objects. In Palmer’s study, subjects could analyze a scene for 3 seconds before the presentation of the object they had to identify and all processing stages of the visual pathway had ample time to be influenced by top-down knowledge and expectations. In Biederman’s study the name of the object to look for (in a subsequently cued location) was provided before stimulus presentation, again leaving time for expectation to influence behavior and performance was impaired with incongruent objects. Eye-tracking studies confirm this congruence effect by showing that objects inconsistent with a scene tended to be fixated longer than consistent ones (De Graef, Christiaens, & D’Ydewalle, 1990; De Graef, De Troy, & D’Ydewalle, 1992). However, the hypothesis of a contextual influence on object processing has been challenged by Hollingworth and Henderson (1998, 1999) who reported that after eliminating guesses and response biases, no advantage was found for the detection of consistent objects (over inconsistent ones). They proposed that object identification processes are isolated from knowledge about the world.

If the extent to which object processing is influenced by its context is still an ongoing debate, the time course of context/object interaction is even more controversial. If context and object processing interfere, such interactions could happen late, after activation of semantic information (Ganis & Kutas, 2003). Alternatively, context could also affect the perceptual processing of object and influence early processing, or set constraints on its possible interpretations. A model proposed by Bar and collaborators suggests fast interference between context and object. According to this model, rapid coarse processing of a scene, possibly through the dorsal magnocellular pathway, would be used to activate the most likely possible object(s) in a contextual frame (Bar, 2004; Bar et al., 2006). Indeed, a coarse “blurred” representation of the contextual frame might be sufficient to guide object processing. The structure of a scene image can be estimated by the basis of global image features that provide a “statistical summary” of its spatial layout properties. Thus, natural image statistics could also be used in scene categorization (Fiser & Aslin, 2001; Torralba & Oliva, 2003), allowing feed-forward processing of scene content and providing early contextual information that can influence object processing (Oliva & Torralba, 2006, 2007).

Getting at the gist of a scene can be achieved at a glance (Potter & Faulconer, 1975; Potter & Levy, 1969). Using a rapid categorization task frequently used to study the time course of object processing (Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2001; Macé, Thorpe, & Fabre-Thorpe, 2005; Rousselet, Macé, & Fabre-Thorpe, 2003; Thorpe, Fize, & Marlot, 1996), we have recently shown that categorizing the global gist of a scene is remarkably fast (Joubert, Rousselet, Fize, & Fabre-Thorpe, 2007), with a time course similar to that of object categorization. In comparison, contextual categorization at more detailed levels such as sea, mountain, urban indoor or outdoor contexts is more time consuming (Rousselet, Joubert, & Fabre-Thorpe, 2005). Contrary to everyday life in which the visual system has ample time to be preset by contextual expectations despite eye and head movements, our rapid categorization task uses natural photographs flashed for only 26 ms or less and thus does not allow time for top-down expectancy influences based on context. In daily life this situation is more likely to happen when one zaps between TV channels or turns over the pages of a magazine. In such situations, contextual information has to be processed in parallel with object features.

Earlier studies have reported contextual influence on object processing with such briefly presented scenes. Davenport and Potter (2004) used manipulated photographs containing a salient object that was consistent (or not) with its context. The scenes were presented for only 80 ms then masked, and subjects were asked to name the object. They reported that objects were named more accurately in semantically related contexts than in non-consistent contexts. Recent studies (Davenport, 2007; Joubert et al., 2007) have also shown that context and object processing could interact in both directions: context can influence object processing but salient objects can also disturb context processing. In both studies the accuracy of background reports was influenced by the consistency of foreground objects. When subjects were just required to categorize scenes as natural or urban in a rapid categorization task (Joubert et al., 2007), the presence of an incongruent salient object in the scene induced an accuracy drop of about 10% correct and delayed correct responses by as much as 80 ms.

The evidence of such interactions in rapid processing suggests that context and object processing must interact early, but the existence of such early interactions has not yet found support from associated brain activity. Analyzing EEG associated with the processing of objects embedded in a congruent or incongruent context, Ganis and Kutas (2003) observed that the earliest signs of activity related to congruity vs. incongruity were recorded in a late 300–500 ms window. This result supports theories postulating a late effect of context on object processing that would depend upon activation of semantic information. Thus, it still remains to be determined when the earliest influences of context on object processing take place. In Davenport’s experiments, subjects were asked to provide a verbal response and no reaction times were provided. The manual go/no-go categorization task often used in our group requires a motor response that has to be
provided “as quickly and accurately as possible.” In such task, median RT are generally around 400 ms or less (Joubert et al., 2007); responses might not even require conscious representations (Thorpe, Gegenfurtner, Fabre-Thorpe, & Bülthoff, 2001). Such rapid responses and the precise quantification of reaction times can provide information about the time course of contextual influence on fast object processing.

This was the aim of the present study. Subjects were asked to perform an animal/non-animal fast categorization task on briefly flashed natural scenes using a very large number of stimuli to avoid possible biases. The brief presentation of the stimuli (26 ms) prevented eye movements and scene exploration. Natural scenes were manipulated so that context and object could be either congruent or not. The robustness of the effect was studied by pasting objects in various congruent or non-congruent contexts. In a preliminary experiment, we controlled for the effect of simply manipulating scenes and pasting objects without interfering with context meaning. In the absence of contextual effects, results from our current study could argue for the “functional isolation” model (Henderson & Hollingworth, 1999). Alternatively, they could support their “priming” model if a delay is needed for scene context to influence object processing. Finally, immediate interactions between context and object processing flows would be compatible with their “perceptual schema” model. Such early effects of context could also support interactions between parallel streams of visual information in a feed-forward wave of processing (Macé et al., 2005; Rousselet, Fabre-Thorpe, & Thorpe, 2002; VanRullen & Thorpe, 2002).

**Experiment 1: Effect of object segregation and stimulus manipulation**

The first experiment was designed to evaluate the impact of pasting an object on a neutral or consistent background on fast object categorization performance.

**Methods**

**Subjects and task**

Twelve volunteers (8 men, mean age 27, range 23–30, 3 of them left-handed) gave their informed written consent. All of them had normal or corrected to normal vision. Subjects performed a rapid visual go/no-go categorization task. They were asked to lift their finger as quickly and as accurately as possible (go responses) each time the picture included an animal, and to withhold their responses (no-go responses) when there was no animal (Thorpe et al., 1996). Each subject performed 12 test series of 96 trials preceded by a training series of 48 trials. In each series, target and distractor trials were equally likely.

**Stimuli**

All horizontal and vertical scenes (768 × 512 pixels, 8° × 5° of visual angle) included a foreground object that was a man-made object for distractor scenes and an animal for target ones. In order to disentangle context influence and object pasting (whether performance could change simply because of stimulus manipulation), each object was seen by every subject in 3 conditions: (1) object in the original non-altered scene, (2) isolated object presented on a gray background, and (3) isolated object pasted in another congruent context. Because contextual processing takes longer at more detailed categorization levels (Joubert et al., 2007; Rousselet et al., 2005), the congruence of an object with its context was defined in terms of “Man-made” vs. “Natural” terms, whereby a sofa is usually found in a man-made environment and a leopard more likely to be found in a natural environment. The order in which these 3 conditions were presented was counterbalanced across subjects. All conditions were randomly interleaved and in equal proportions in each series.

The subset of 384 original pictures (see examples in Figure 1, O) were all selected from a large commercial CD-ROM library (Corel Stock Photo Libraries) and included 192 “Man-made” scenes in which there was one or more man-made objects and 192 “Natural” scenes in which there was one or more animals (natural object). Images were in jpeg 24-bit format (16 million colors). Within each category, images were as diverse as possible.

For half of the stimuli (192), man-made and natural foreground objects were cropped manually from images. To avoid excessively sharp edges we applied Paintshop (version 7.0.0.2, Jasc Software Inc.) progressive transparency on the contour (2 pixels wide). This procedure being very time consuming, the other half of the stimuli were built with objects from the Hemera Photo Objects library. The objects were chosen to be as similar as possible to the object present in the original images and found under the same label in the Hemera library (see Figure 1). Again with these objects, progressive transparency was also used on the contours in order to avoid sharp edges and to allow a good integration of objects in their new background.

The subset of stimuli, objects on gray background, were built by pasting the isolated object in the same location as in the original image on a gray background for which luminance was adjusted to the global luminance of the original scene (see examples in Figure 1, G). To build the set of objects pasted on a new congruent context, we tried to control as many features as possible (see examples in Figure 1, NC1). For each original scene, we chose one picture of the same context category (man-made or natural) with a roughly similar background in terms of
Figure 1. Examples of animal targets and man-made object distractors used in the animal categorization task. Different subsets were used in the 3 experiments: original scenes (O), isolated objects on gray background (G), objects pasted on two different congruent contexts (C1 and C2), and objects pasted on two different non-congruent contexts (NC1 and NC2). Context congruence was considered in terms of natural vs. man-made and scale, position and support relations were respected as much as possible. Examples using objects cropped from original scenes are shown in the two top rows for target and distractors. Examples using similar objects taken from the Hemera library are shown in the two bottom rows for targets and distractors. The first row for distractors illustrates the kind of man-made objects that elicited false alarms when out of its context, especially when pasted on non-congruent (natural) contexts. The number (top left of each image) indicates the number of subjects (out of 12) who correctly withheld their go response when presented with that scene.
orientation, global luminance and spatial layout. Within this new picture, the location of the object was selected to be as close as possible to the original scene, taking into account orientation and coherence (support, interposition, scale, Biederman et al., 1982). The local luminance was evaluated at this position. Using the YCbCr color system (a way of encoding RGB information in which Y is the luminance component and Cb and Cr are the blue and red chrominance components) that allows the color channel values to be preserved, we adjusted the object luminance relative to the background local luminance in order to keep the same local contrast between object and background. Finally, we pasted the adjusted object onto the selected background.

Procedure

Subjects sat in a dimly lit room, at 1 m from a computer screen (resolution 1024 × 768, vertical refresh: 75 Hz) piloted by a PC computer. Image presentation and behavioral response measurements were carried out using Presentation software (NeuroBehavioral Systems, http://nbs.neuro-bs.com/).

Each trial started with a fixation cross (1° of visual angle) that appeared at the center of a black screen for 300–900 ms randomly. As soon as the cross disappeared, the stimulus was displayed for two frames (26 ms), also in the middle of the screen. These brief presentations prevented exploratory eye movements. To start stimulus presentation, subjects had to place their fingers on a response pad equipped with infrared diodes that allowed microsecond precision. After the image presentation, a black screen was displayed for 1000 ms, during which period subjects were required to respond by a finger lift if the image was a target. Longer reaction times were considered as no-go responses. Following this 1-sec period, a black screen was displayed during 300 ms before the next trial started. A trial lasted between 1600 and 2200 ms.

Statistics

We used the non-parametric two-way repeated measure Friedman test to evaluate statistical differences across subjects among the three conditions (original, isolated and pasted). When the Friedman test showed a statistical difference, a paired Wilcoxon test, Bonferroni corrected, was used to perform pairwise comparisons (in the text, p-values are corrected).

In order to provide additional information on the robustness of the effect observed on individual data, we computed confidence intervals using Monte Carlo simulations to test for significant differences between conditions for each subject (Figure 2B) using the following procedure. For each pairwise comparison, responses on each trial (go/no-go or reaction time) from the two conditions, containing n and m number of images, were pooled together and randomly shuffled. Then n image responses were assigned to a ‘fake’ subset 1 and the m others to a ‘fake’ subset 2. Averaged performance was thus computed for the two fake subsets and the difference between them was stored. This procedure was run 2000 times, providing a confidence interval around the null hypothesis that the two conditions were actually sampled from the same population. These confidence intervals are plotted in figures that report individual subjects’ differences.

Results

All individual results for the 3 experimental conditions are summarized in Table 1 and in Figure 2.

Accuracy

Subjects were very efficient at performing the animal categorization task in all three conditions. Global accuracy (correct go and no-go responses) reached 96.2% with original images, 96.1% with isolated objects on a gray background and 94.1% with pasted objects on new natural contexts. Accuracy differences between the 3 conditions were not statistically significant (Friedman test: n.s. \( \chi^2 = 4.696, df = 2, p = 0.096 \), Figure 2A). However, the individual performance analysis revealed a significantly lower accuracy in the pasted condition when compared to the original condition in 9 out of 12 subjects (Figure 2B). Considering separately accuracy on target and distractor trials provided more information (Table 1). The global accuracy decrease was clearly due to target trials (97.7% vs. 94.1%; Friedman test: \( \chi^2 = 17.522, df = 2, p < 0.00001 \); paired Wilcoxon test on go-responses: \( Z = 2.847, p = 0.012 \)), whereas accuracy was not statistically different on distractor trials (94.7% vs. 94%; n.s. Friedman test: \( \chi^2 = 1.644, df = 2, p = 0.439 \)). No differences were observed between original scenes and isolated objects.

An additional observation can be made concerning the subjects’ response bias. When categorizing original scenes and isolated objects, subjects tended to respond “animal.” Incorrect trials were biased toward false alarms (original: 5.5%, isolated: 6.8%), whereas subjects missed very few targets (respectively 2.3% and 1%). This bias toward targets was significant (paired Wilcoxon test, \( Z > 2.552 \) and \( p < 0.033 \) in both conditions). No such bias was observed for the pasted condition (false alarms 6%, missed targets 5.9%; paired Wilcoxon test, \( Z = 0.275, p = 1 \)).

Reaction times

Subjects were also very fast, with median reaction times (RT) around 380 ms. Median RT were computed on correct target trials. Compared to the original condition, the mean median RT was slightly shorter with isolated
targets (376 ms vs. 371 ms; paired Wilcoxon test: $Z = 2.512$, $p = 0.036$), and longer with pasted targets (376 ms vs. 387 ms; paired Wilcoxon test: $Z = 2.867$, $p = 0.012$). Although RT differences were small, they were also very robust at the individual level (see Table 1 and Figure 2B). These differences are also illustrated by the RT distributions (Figure 2C).

To evaluate how accuracy varies with response latency, $d'$ curves were computed for the 3 conditions (Figure 2C). The $d'$ curves showed that the information processing rate was very similar for the original and isolated conditions. By contrast, the $d'$ curve for the “pasted” condition is shifted toward longer latencies from the very beginning and reaches a lower plateau corresponding to an accuracy
drop. Such a clear shift (10–20 ms) from the shortest RT indicates that pasting an object on a new natural context is not a trivial manipulation and is able to slow down object categorization even for the earliest responses.

As described in the Methods section, half of the manipulated “objects” were totally identical to the object in the original scene while the other half were chosen from the Hemera library to be very similar. We thus checked whether these two image subsets lead to similar categorization performances in all 3 conditions. Behavioral performances on the two image subsets were very similar in each condition. In the original condition, the source images of Corel and Hemera subsets were similarly difficult to categorize (accuracy: 98% vs. 97.4%; median RT: 374 ms vs. 379 ms). In the pasted condition, similar drops of performance were observed for both subsets relatively to the original condition (93.1% vs. 95.1%, 10 ms RT increase in both cases). Finally, global accuracy and median RT for the two gray subsets were also very similar (Corel animals: 99.1% and 371 ms; Hemera animals: 98.8% and 370 ms). Friedman tests computed on individual accuracy and median RT using condition (3 levels) and image subset (2 levels) as factors confirmed the absence of subset effect after adjusting for condition effects (accuracy: \( \chi^2 = 1.76, p = 0.1851 \); median RT: \( \chi^2 = 0, p = 1 \)). Thus, we can conclude that the subsets of stimuli using identical and Hemera animals were associated with virtually identical performance and were similarly affected by the experimental manipulations.

To sum up, the first experiment showed that segregating an object from its background has—surprisingly—very little effect if any (at least in an interleaved protocol) on rapid categorization performance: compared with original photographs, accuracy is not improved by isolating target objects and median RT is only shortened by 5 ms. This result is strengthened by the finding that pasting an object on another congruent context has a cost both in terms of response speed (10-ms increase in median RT) and in terms of global accuracy (>2%). This accuracy cost is larger when only considering target trials (>3.5%). Experiment 1 showed that a performance drop can be due to simple stimulus manipulations.

Such manipulations may have affected the saliency of foreground objects. To determine whether objects were as physically salient in original vs. pasted context, we used the saliency toolbox (Walther & Koch, 2006), inspired by the computational model of visual attention from Itti and Koch (2001). The most salient zone was found on (or close to) the animal in 74% of the original images but in only 60% of the pasted stimuli (41% vs. 33% for man-made objects). To prevent interferences from such manipulation with the evaluation of context congruence, all stimuli in the following experiments used pasted objects in contextual scenes.

## Experiment 2: Effect of context congruence on object processing

Experiment 2 used only objects pasted on congruent or non-congruent contexts to evaluate the effect of contextual incongruence in rapid visual categorization.

### Table 1

<table>
<thead>
<tr>
<th>Subject</th>
<th>Original</th>
<th>Isolated</th>
<th>Pasted</th>
<th>Original</th>
<th>Isolated</th>
<th>Pasted</th>
<th>Original</th>
<th>Isolated</th>
<th>Pasted</th>
<th>Original</th>
<th>Isolated</th>
<th>Pasted</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRO</td>
<td>98.2</td>
<td>97.7</td>
<td>96.1</td>
<td>98.4</td>
<td>99.0</td>
<td>94.3</td>
<td>97.9</td>
<td>96.4</td>
<td>97.9</td>
<td>399</td>
<td>396</td>
<td>407</td>
</tr>
<tr>
<td>JSN</td>
<td>92.2</td>
<td>92.2</td>
<td>87.8</td>
<td>92.7</td>
<td>97.4</td>
<td>85.9</td>
<td>91.7</td>
<td>87</td>
<td>89.6</td>
<td>324</td>
<td>315</td>
<td>335</td>
</tr>
<tr>
<td>RVR</td>
<td>94.3</td>
<td>96.1</td>
<td>85.9</td>
<td>98.4</td>
<td>99.0</td>
<td>91.1</td>
<td>90.1</td>
<td>93.2</td>
<td>80.7</td>
<td>409</td>
<td>407</td>
<td>410</td>
</tr>
<tr>
<td>JMO</td>
<td>96.6</td>
<td>95.1</td>
<td>93.8</td>
<td>98.4</td>
<td>99.5</td>
<td>94.8</td>
<td>94.8</td>
<td>90.6</td>
<td>92.7</td>
<td>324</td>
<td>313</td>
<td>334</td>
</tr>
<tr>
<td>NGU</td>
<td>94.3</td>
<td>97.4</td>
<td>95.8</td>
<td>99.4</td>
<td>100.0</td>
<td>99.0</td>
<td>90.1</td>
<td>94.8</td>
<td>92.7</td>
<td>371</td>
<td>367</td>
<td>384</td>
</tr>
<tr>
<td>IBA</td>
<td>98.4</td>
<td>97.9</td>
<td>96.4</td>
<td>99.5</td>
<td>97.9</td>
<td>94.8</td>
<td>97.4</td>
<td>97.9</td>
<td>97.9</td>
<td>410</td>
<td>401</td>
<td>413</td>
</tr>
<tr>
<td>JMA</td>
<td>97.4</td>
<td>99.0</td>
<td>95.1</td>
<td>97.4</td>
<td>99.0</td>
<td>93.2</td>
<td>97.4</td>
<td>99</td>
<td>96.9</td>
<td>377</td>
<td>374</td>
<td>400</td>
</tr>
<tr>
<td>LBA</td>
<td>97.4</td>
<td>97.7</td>
<td>94.8</td>
<td>97.4</td>
<td>99.0</td>
<td>92.2</td>
<td>97.4</td>
<td>96.3</td>
<td>97.4</td>
<td>434</td>
<td>430</td>
<td>431</td>
</tr>
<tr>
<td>SBI</td>
<td>95.6</td>
<td>95.6</td>
<td>96.1</td>
<td>98.4</td>
<td>99.5</td>
<td>94.8</td>
<td>92.7</td>
<td>91.7</td>
<td>97.4</td>
<td>351</td>
<td>340</td>
<td>371</td>
</tr>
<tr>
<td>NBA</td>
<td>97.1</td>
<td>93.8</td>
<td>95.6</td>
<td>99.5</td>
<td>99.5</td>
<td>96.9</td>
<td>94.8</td>
<td>88</td>
<td>94.3</td>
<td>348</td>
<td>342</td>
<td>366</td>
</tr>
<tr>
<td>JFO</td>
<td>95.3</td>
<td>98.2</td>
<td>96.4</td>
<td>94.8</td>
<td>99.0</td>
<td>94.8</td>
<td>95.8</td>
<td>97.4</td>
<td>97.9</td>
<td>444</td>
<td>452</td>
<td>457</td>
</tr>
<tr>
<td>MMA</td>
<td>97.7</td>
<td>92.7</td>
<td>95.1</td>
<td>99.5</td>
<td>99.0</td>
<td>97.4</td>
<td>95.8</td>
<td>86.5</td>
<td>92.7</td>
<td>328</td>
<td>319</td>
<td>339</td>
</tr>
<tr>
<td>Mean</td>
<td>96.2</td>
<td>96.1</td>
<td>94.1</td>
<td>97.7</td>
<td>99.0</td>
<td>94.1</td>
<td>94.7</td>
<td>93.2</td>
<td>94.0</td>
<td>376</td>
<td>371</td>
<td>387</td>
</tr>
<tr>
<td>Std.</td>
<td>1.9</td>
<td>2.3</td>
<td>3.5</td>
<td>2.0</td>
<td>0.7</td>
<td>3.4</td>
<td>2.9</td>
<td>4.4</td>
<td>5.0</td>
<td>43</td>
<td>47</td>
<td>39</td>
</tr>
<tr>
<td>Min</td>
<td>92.2</td>
<td>92.2</td>
<td>85.9</td>
<td>92.7</td>
<td>97.4</td>
<td>85.9</td>
<td>90.1</td>
<td>86.5</td>
<td>80.7</td>
<td>324</td>
<td>313</td>
<td>334</td>
</tr>
<tr>
<td>Max</td>
<td>98.4</td>
<td>99.0</td>
<td>96.4</td>
<td>99.5</td>
<td>100.0</td>
<td>99.0</td>
<td>97.9</td>
<td>99.0</td>
<td>97.9</td>
<td>444</td>
<td>452</td>
<td>457</td>
</tr>
</tbody>
</table>

Table 1. Individual accuracy and reaction time in the 3 conditions of Experiment 1: Original scenes, isolated objects on gray background and pasted objects on another congruent context (see examples in Figure 1). The four bottom lines indicate mean, standard deviation, minimal and maximal scores computed from individual scores.
Methods

Subjects

Twelve volunteers (8 men, mean age 28, range 22–33, 3 of them left-handed) gave their informed written consent. All of them had normal or corrected to normal vision. Three of them participated in the first experiment (see Table 2).

Stimuli

In Experiment 2, 768 stimuli were used: (1) 384 foreground objects with congruent context (192 man-made objects and 192 animals), which were those used in the first experiment, and (2) the same 384 foreground objects pasted on a non-congruent context. In this experiment, all stimuli contained a pasted object. Pictures of objects with congruent or non-congruent contexts were built as reported in Experiment 1. Context was defined as congruent if an animal was pasted on a natural context, and as non-congruent if it was pasted on a man-made context, and conversely for a man-made object. Contexts were defined as man-made and natural following Joubert et al. (2007): “natural environment” contexts included sea, mountain, desert, iceberg, forest and field scenes; “man-made environment” contexts included street scenes (with or without pedestrians) and indoor scenes like kitchens, museums, and churches. None of the man-made scenes contained mountains or views of the sea, and none of the natural scenes contained buildings. Scenes came from the same photograph gallery as the one used in Joubert et al. (2007) in which the speed of context categorization per se was evaluated. For each of the 4 subsets of stimuli, man-made objects/animals pasted in congruent/incongruent contexts, the mean eccentricities defined as the distance between the fixation point and the center of the object (man-made or animal) were below 48 pixels corresponding to 0.5° of visual angle. Object eccentricity was thus similar for all subsets.

Task and procedure

Task and procedure were identical to those in Experiment 1. Subjects performed the animal categorization task for 8 series of 96 trials; all series contained, randomly interleaved, 25% animal targets in congruent contexts, 25% animal targets in non-congruent contexts, 25% man-made objects in congruent contexts and 25% man-made objects in non-congruent contexts. Each subject saw all target and non-target objects randomly displayed over the experimental series once in a congruent context, another time in a non-congruent context, but never in the same series.

Results

All global and individual results are summarized in Table 2 and Figure 3.

Accuracy

There was a clear effect of context congruence on global accuracy. Subjects scored 92.9% correct when animals and distractor objects were presented in a congruent context (respectively natural and man-made), but they only scored 85.2% when both target objects and distractor objects were presented in a non-congruent context. A paired Wilcoxon test showed that this 7.7% (95% percentile bootstrap confidence interval: 6.9–8.7%) decrease in accuracy was statistically significant ($Z = 3.059$, $p = 0.002$, Figure 3A) and Monte Carlo simulations revealed that this decrease was significant for all subjects (Figure 3B). It is interesting to notice that when analyzing separately correct and incorrect go responses, subjects tended to produce more go responses with natural scene contexts than with man-made ones. This was true for correct go responses toward animals (91.5% and 81.3%, respectively, paired Wilcoxon test: $Z = 3.062$, $p = 0.002$, Figure 3A) and individually significant for 11 subjects among 12 (Figure 3B). This was also true for the incorrect go responses (false alarms) produced toward man-made

### Table 2

<table>
<thead>
<tr>
<th>Subject</th>
<th>Global (Accuracy (%))</th>
<th>Targets</th>
<th>Distractors</th>
<th>Reaction time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C</td>
<td>NC</td>
<td>C</td>
<td>NC</td>
</tr>
<tr>
<td>NGU</td>
<td>95.6</td>
<td>85.9</td>
<td>95.8</td>
<td>87.5</td>
</tr>
<tr>
<td>SCR</td>
<td>91.9</td>
<td>84.6</td>
<td>92.2</td>
<td>80.7</td>
</tr>
<tr>
<td>MLA</td>
<td>94.8</td>
<td>87.8</td>
<td>94.8</td>
<td>85.9</td>
</tr>
<tr>
<td>TMA</td>
<td>95.1</td>
<td>88.5</td>
<td>92.7</td>
<td>82.3</td>
</tr>
<tr>
<td>RVR</td>
<td>88.5</td>
<td>78.1</td>
<td>86.5</td>
<td>71.4</td>
</tr>
<tr>
<td>APA</td>
<td>95.3</td>
<td>83.3</td>
<td>92.7</td>
<td>80.2</td>
</tr>
<tr>
<td>NBO</td>
<td>95.6</td>
<td>89.1</td>
<td>96.9</td>
<td>87.5</td>
</tr>
<tr>
<td>SBE</td>
<td>88.5</td>
<td>79.2</td>
<td>83.3</td>
<td>70.3</td>
</tr>
<tr>
<td>JMA</td>
<td>94.8</td>
<td>86.7</td>
<td>96.9</td>
<td>88.5</td>
</tr>
<tr>
<td>CHE</td>
<td>93.2</td>
<td>86.5</td>
<td>89.1</td>
<td>78.1</td>
</tr>
<tr>
<td>LLA</td>
<td>93.2</td>
<td>87.8</td>
<td>92.2</td>
<td>86.5</td>
</tr>
<tr>
<td>FRE</td>
<td>90.6</td>
<td>84.6</td>
<td>85.4</td>
<td>76.6</td>
</tr>
<tr>
<td>Mean</td>
<td>92.9</td>
<td>85.2</td>
<td>91.5</td>
<td>81.3</td>
</tr>
<tr>
<td>Std.</td>
<td>2.6</td>
<td>3.5</td>
<td>4.5</td>
<td>6.3</td>
</tr>
<tr>
<td>Min</td>
<td>88.5</td>
<td>77.1</td>
<td>83.3</td>
<td>70.3</td>
</tr>
<tr>
<td>Max</td>
<td>95.6</td>
<td>89.1</td>
<td>96.9</td>
<td>88.5</td>
</tr>
</tbody>
</table>

Table 2. Individual accuracy and reaction time in the 2 conditions of Experiment 2: Pasted on congruent context (C) and pasted on non-congruent context (NC, see examples C1 and NC1 in Figure 1). The four bottom lines indicate mean, standard deviation, minimal and maximal scores computed from individual scores. Subjects NGU, RVR and JMA have also performed Experiment 1.
objects that were embedded in natural (non-congruent) contexts (false alarms: 10.9%). False alarms were considerably reduced when man-made objects were in man-made contexts (5.6%). This difference was statistically significant across subjects (paired Wilcoxon test $p = 0.002$, $Z = 3.065$; Figure 3A) and very consistent considering individual results (Figure 3B). The number of false alarms produced toward man-made objects presented in the context of man-made scenes were very similar in Experiments 1 and 2 (6% vs. 5.6%), but a decrease in correct go responses toward targets presented in a natural context was observed (94.1% in Experiment 1 and 91.5% in Experiment 2). This decrease might be due to the group of subjects tested but more likely (as accuracy is usually a robust measure) to the fact that the congruent pictures used in Experiments 1 and 2 were intermixed with conflictual stimuli.

Reaction times

Reaction times were also affected by congruence. Subjects categorized animals with an averaged median
RT of 406 ms in the congruent context condition and 421 ms in the non-congruent context condition (Figure 3A). This 15 ms (95% confidence interval: 11–20 ms) RT increase was statistically significant across subjects (paired Wilcoxon test: $Z = 3.059$, $p = 0.002$) and individually for 5 subjects (Figure 3B).

Although RT distributions for correct go responses (Figure 3C) were nearly superimposed in the initial portion of the distribution, the two global distributions clearly differed from 300 ms onward as attested by $\chi^2$ tests performed between the correct go distributions in congruent and non-congruent conditions ($\chi^2$ test: $p = 0.0246$). Moreover, a higher proportion of correct go vs. incorrect go responses was observed in the congruent context condition. Consistent with these observations, the $d'$ curves computed for each condition clearly differed. The $d'$ curve for the non-congruent context condition was shifted toward longer latencies and reached a lower

![Figure 4](http://jov.arvojournals.org)
plateau. This shift of about 20–30 ms at minimal RT value (Figure 3C) shows that incongruence between object and context affects the information accumulation rate even for the earliest responses.

To summarize, Experiment 2 revealed very early interactions between object and context processing. In a rapid visual go/no-go task in which stimuli are just flashed for 26 ms, context congruence can influence performance. Such performance is biased toward go responses with a natural context, and toward no-go responses with a man-made context. Non-congruent contextual information delays object processing even for the earliest responses.

**Post-hoc analysis on object saliency**

Because natural contexts are usually simpler and more uniform than richer man-made contexts, the congruence effect observed in Experiment 2 could be partially explained by a higher saliency of animal and man-made objects in natural contexts. Such higher saliency would not necessary lead to better processing as shown by the worse rejection of man-made objects as distractors in non-congruent natural context (Figure 3A). We used the same saliency toolbox as in Experiment 1 to measure object saliency in congruent and non-congruent contexts, in order to evaluate the congruency effect as a function of object saliency. For each stimulus used in Experiment 2, we computed a saliency map which allowed us to define the most physically salient area of the scene based on different properties like luminance, orientations and colors. According to this attentional model, this most salient area should be the most probable land location for the first eye movement if photographs were flashed longer. However, the brief image presentation used in the present study prevented any eye movement exploration. Indeed a bias was found such that animal and man-made objects were both less salient in artificial contexts. The most salient zone was found outside the animal (or man-made) object in 67% (69%) in artificial contexts, but in only 42% (58%) of the case in natural context. We thus analyzed separately the effect of contextual congruence on 3 different subsets of stimuli: most salient area (1) on animals or man-made objects, (2) sat astride animal/object boundary or (3) outside the foreground object. Results are illustrated in Figure 4. This analysis demonstrated that the effect of contextual congruence observed in Experiment 2 was independent of object saliency. Regardless of object saliency (animal or man-made), an accuracy drop and an RT increase were observed for objects pasted in non-congruent contexts (vs. congruent). The robustness of the congruence effect after adjusting for saliency effects was confirmed by non-parametric repeated measure Friedman tests with 2 factors (congruence and saliency) on global accuracy (χ² = 45.19, p < 0.0001), correct go (χ² = 33.78, p < 0.0001), correct no-go (χ² = 13.51, p = 0.0002) and median RT (χ² = 4.55, p = 0.0329). Such results confirm that the performance impairments observed in this experiment must be attributed to a congruence effect between object and context.

**Experiment 3: Robustness of the congruence effect**

A third experiment was designed to assess the robustness of the results obtained in Experiment 2 by associating the same target and non-target objects used previously with different congruent and non-congruent backgrounds.

**Methods**

**Subjects**

Twelve volunteers (6 men, mean age 24, range 21–27, 2 of them left-handed) gave their informed written consent. All of them had normal or corrected to normal vision. None of them had participated in the first two experiments.

**Stimuli**

Experiment 3 employed 384 stimuli identical to those in Experiment 2, and 384 stimuli in which paired objects switched context (congruent and non-congruent, Figure 1). This second set was built as follows.

Half of the stimuli from Experiment 2 were first randomly selected. These 384 stimuli were divided in 4 sets of 96 stimuli (targets in congruent contexts, targets in non-congruent contexts, non-targets in congruent contexts, non-targets in non-congruent contexts). Within each set, stimuli were organized in 48 pairs; care was taken to pair stimuli so that scales and relative positions of the two paired objects were as coherent as possible in both contexts. Then, within each pair, objects were context switched.

Thus, 8 image subsets were considered for analysis: 96 animal objects pasted on two different but congruent natural contexts (C1 and C2), the same 96 animals pasted on two different but non-congruent man-made contexts (NC1 and NC2); 96 man-made objects pasted on two different but congruent man-made contexts (C1 and C2), and the same 96 man-made objects pasted in two different but non-congruent natural contexts (NC1 and NC2). C1 and NC1 were identical stimulus sets to those from Experiment 2 whereas C2 and NC2 were new context switched stimuli.

**Task and procedure**

Task and procedure were identical to those in Experiment 1. Subjects performed the animal categorization task for 8 series of 96 trials. All series contained 50%
animal targets and 50% man-made objects randomly interleaved, with an equal proportion of each context condition. Each subject saw each object pasted on the 4 different backgrounds, but a given object was never seen twice in the same series.

Results

The aim of this third experiment was to check whether the performance impairment induced by contextual incongruence observed in Experiment 2 was robust, independently of the specific congruent (C) and non-congruent (NC) contexts used to paste a given object. Here we used non-parametric repeated measure Friedman test with 2 factors: the congruence and the context subset. All global and individual results are summarized in Table 3.

Accuracy

With objects pasted on congruent contexts, subjects performed the task with high global accuracy for subsets C1 (92.7%) and C2 (92.1%). They showed similar drops of accuracy when objects were shown on non-congruent context, scoring 88.3% and 87.5% correct respectively for subsets NC1 and NC2 (Figure 5A). Friedman tests revealed a significant effect of congruence after adjusting for possible context subset effects ($\chi^2 = 12.14, p = 0.0005$) while there was no effect of context subset after adjusting for possible congruence effects (n.s. $\chi^2 = 0.35, p = 0.5523$).

It is interesting to notice that, regardless of the subset of stimuli (C1 vs. NC1 or C2 vs. NC2) subjects reached similar accuracies. Moreover, their impairment was similar to the drop of performance displayed by the different group of subjects tested in Experiment 2 (92.9% and 85.2% correct respectively on congruent and non-congruent contexts).

Performance was evaluated separately for target and non-target trials, for set 1 and set 2 (Figure 5). We observed no difference ($\chi^2 = 0.15, p = 0.6972$), on correct go responses toward targets between C1 and C2 (92.7% and 91.9% respectively), nor between NC1 and NC2 (85.9% and 83.7%). As in Experiment 2, subjects tended to respond more to targets when pasted on any congruent context and to withhold their response on any non-congruent context ($\chi^2 = 21.04, p < 0.0001$). This effect was present for all subjects and replicated the results of Experiment 2.

On distractor trials featuring man-made objects, the context influence was not as clear as in Experiment 2. When objects were presented in congruent contexts, the percentage of incorrect go responses reached 7.3% and 7.8% (respectively for C1 and C2); it increased up to 9.2% and 8.8% with non-congruent contexts (respectively NC1 and NC2). This slight (around 1.5%) impairment did not reach significance (Friedman test: n.s. $\chi^2 = 0.61, p = 0.4351$) and independent permutation tests confirmed this result (1000 samples, between C1 and NC1: $p = 0.09$).

Table 3. Individual accuracy and reaction time with the 2 conditions and 4 image subsets in Experiment 3: Objects pasted on two different congruent contexts (C1 and C2) and pasted on two different non-congruent contexts (NC1 and NC2; see examples C1, C2, NC1 and NC2 in Figure 1). The four bottom lines indicate mean, standard deviation, minimal and maximal scores computed from individual scores.
C2 and NC2: $p = 0.5523$). This result was at odds with those obtained on distractor trials in Experiment 2. As Experiment 3 used only half of the stimuli of Experiment 2, we reanalyzed results from Experiment 2 on this restricted subset of stimuli. No difference in accuracy or RT was found on this image subset in Experiments 2 and 3.

Reaction times

Animals in congruent contexts were categorized with median RT of 396 ms and 402 ms, respectively for C1 and C2 images. When animal targets were embedded in man-made contexts, median RT increased to 412 ms and 411 ms (for NC1 and NC2, respectively). Two-way Friedman tests revealed no congruence effect after adjusting for possible context subset bias (n.s. $\chi^2 = 2.28$, $p = 0.1308$), and no context subset effect after adjusting for possible congruence bias (n.s. $\chi^2 = 0.12$, $p = 0.7285$). However, more sensitive permutation tests contrasting conditions two by two did show a congruence effect (1000 samples, C1 vs. NC1: $p = 0.0025$; C2 vs. NC2: $p = 0.0285$), while it confirmed the lack of context subsets bias (1000 samples, C1 vs. C2: $p = 0.1182$; NC1 vs. NC2: $p = 0.7784$). A small RT increase of about 10–15 ms is thus systematically associated with the decreased accuracy observed when animal targets are presented on non-congruent contexts.

This observation is strengthened by the $d'$ results showing that the information accumulation rate in congruent and non-congruent contextual conditions differed.
Although the $d'$ curves were superimposed for identical context conditions, the $d'$ curves associated with either congruent or non-congruent contexts diverged very early (around 280 ms for context subsets 1 and 300 ms for context subsets 2), showing that context incongruence can have in most cases a deleterious effect on object processing from very early on.

**General discussion**

The aim of the present study was to evaluate the temporal dynamics of contextual influences on fast object categorization. **Experiments 2 and 3** clearly demonstrated a performance impairment in object categorization performance due to incongruent contextual information. This effect is so fast that it affects even on the earliest responses produced by the subjects; it is very robust as it was reproduced regardless of the group of subjects and regardless of the particular contexts in which the objects were pasted.

**Experiment 1** provided two additional important results. First, manipulating natural stimuli is not a trivial operation and induces small performance alterations despite all the effort and care involved in making the stimuli. Second, it shows a surprising result, namely that animal categorization is not easier when animal targets are isolated on gray backgrounds rather than embedded in natural scenes. This latter result has to be discussed in relation to the debate about whether figure-ground segregation must precede object recognition.

**No performance improvement for isolated targets**

Data from **Experiment 1** replicated the high human performance in ultra-rapid categorization of animals in natural scenes (Fabre-Thorpe et al., 2001; Thorpe et al., 1996). When targets are embedded in their original context, animal categorization is performed with high accuracy and fast RTs (96.2% with 376 ms median RT). Such scores were expected to be improved by the use of isolated animals on gray backgrounds; however, performance showed little if any improvement. With isolated objects, accuracy was similar (96.1% correct on isolated objects versus 96.2% correct using natural scenes) and the overall median RT showed only a mild decrease of 5 ms (371 ms vs. 376 ms). This decrease reached statistical significance, and analysis of individual performances (**Figure 2B**) also showed that this tendency was present in most subjects (11 out of 12) but was indeed very small.

Although the use of natural scenes in addressing object categorization makes use of contextual guidance. Alternatively, subjects could tend to adjust their response speed across trials as a function of the time necessary to respond optimally to stimuli belonging to the hardest condition. According to this latest hypothesis, we thus might observe an advantage for the isolated condition when subjects perform the same task in a blocked design. Further experiments are currently being run using a block design to tackle the effect of object isolation in a task requiring a more detailed level of categorization (dog/non-dog).

**The “pasting” effect**

**Experiment 1** also allowed the evaluation of the performance impairment due to stimulus manipulations. Despite the extreme care taken in pasting the isolated objects in new congruent backgrounds, rapid categorization performance was impaired with these manipulated stimuli.

This “pasting effect” was observed on accuracy and response speed. A global decrease of 2.1% of correct
responses did not reach significance, but the tendency was observed in most subjects and affected the accuracy on targets more than the accuracy on distractors. This accuracy deficit on target trials was associated with a significant median RT increase of about 10 ms. This effect also was observed even for the earliest responses as illustrated by the shift of the d’ curves between original and pasted conditions. This performance impairment could be due to the decreased saliency of the foreground object when pasted on a new context as shown by the saliency analysis. If object saliency could bias early processing, performance should be improved with isolated objects that are of course salient in 100% of the cases. This was not the case although this result might be due to a ceiling effect.

Alternatively, the “pasting” effect could be due to other alterations. We can consider the local physical alteration at the object/context boundary possibly due to techniques used to introduce the object in a new context (see Methods section). Despite all the careful precautions taken during stimulus manipulation to prevent from reported violation effects (equating spatial layouts, relative scales, supports, object interpositions, Biederman et al., 1982), not all the physical features of realistic photographs could be preserved: object illuminations and shadows might not be coherent, thus violating usual co-occurrence of certain visual features. Pasting effects might result from the violation of such expectations.

Finally, another explanation of the pasting effect might be related to our definition of a “congruent” context. The present experiment considers as equivalent all natural contexts versus all urban contexts, but subjects might have clear expectations about where to find a given animal. For example, giraffes tend to live in dry, open wooded areas in the savannah and might be incongruent in a mountain scene; the mountain scene would be considered as congruent in the present study but might not be congruent for the giraffe. Although plausible, this explanation does not take into account the fact that recognition of a detailed context such as “sea scenes” or “mountains scenes” takes longer than the recognition of such scenes as “natural context” (Joubert et al., 2007), and that the impairment observed here was significant from the earliest responses.

The “congruence” effect

Experiments 2 and 3 used only manipulated stimuli, thus enabling the evaluation of contextual congruence on its own, preventing interference from other visual regularity violations revealed by Experiment 1. Both experiments clearly showed an effect of context congruence, an effect that was present regardless of the object saliency in the scene. When objects were pasted in a non-congruent context, categorization performance on targets was always worse both in terms of accuracy (10% drop) and response speed (15 ms slower) performance dropped by 10% and about 15 ms on targets. Animals were clearly less easily categorized when presented in an urban context than in a natural context. The results were less robust for the false alarms, induced by manufactured objects presented in a natural (non-congruent) context; further investigations are needed to better understand the pattern of false alarms triggered by the man-made objects.

The contextual effect on target categorization replicated in two experiments provides evidence that context processing strongly interacts with object categorization; this result is strengthened by the fact that subjects had no a priori information about briefly flashed scenes, and were thus free from top-down influences that can be primed when presentations of visual scenes precede the processing of target objects (Palmer, 1975). Obviously, in daily life, our environment is usually stable allowing predictions to be made on the most likely objects to appear. In such cases, context effects on object perception are not time constrained and probably strengthened. On the other hand, the early interactions demonstrated in our experiment might be fully used in circumstances where the surrounding context changes suddenly, as when opening a door and entering a new scene, when making large head movements, when driving a car in a city, zapping from one TV channel to another, or watching family photographs. The second aim of our study was to determine the temporal dynamics of object and context interactions when presented simultaneously. Here the results were very surprising since no minimum delay was necessary to observe a context influence on object categorization. The conflict between objects and their surrounding contexts induced an additional processing time of about 10–20 ms when considering the minimal input–output processing time (Figures 3 and 5) and increased even more for longer response latencies.

This result clearly argues against the functional isolation model proposed by Henderson and Hollingworth in which objects and contexts are processed independently without interfering (Henderson & Hollingworth, 1999; Hollingworth & Henderson, 1998). Furthermore, these object–context interactions could be bidirectional, since Davenport and Potter (2004) and Joubert et al. (2007) provided evidence that salient foreground objects can also influence context processing. Notably, Joubert et al. (2007) have recently shown that the time course of context and object processing are very similar. This implies that in some cases the context might be categorized faster than the foreground object. In that case, the rapidly processed contexts might interfere with object categorization at a pre-decisional stage. Conversely, for other natural scenes in which salient objects might be categorized faster than context, object processing would also influence context recognition at a pre-decisional stage. In most intermediate cases, one can postulate bidirectional interactions before any decision has been reached on either object or context category. Our results thus support perceptual schema models, which propose that the flow of object and context processing can interact early on during perceptual processing.
In a model proposed by Bar and his group (Bar, 2004; Bar et al., 2006), a coarse processing of the context performed through the magnocellular dorsal visual pathway can influence object recognition (Bullier, 2001). This model could well account for interaction of context on object recognition but would have more difficulty for the opposite case. Macé et al. (2005) also emphasized the guidance help that can be provided by the fast magnocellular pathway but, rather than setting this influence through a control exerted by the dorsal visual stream on the ventral visual stream, such interactions were proposed to take place mostly within the ventral visual pathway. Following these views, at each processing stage of the ventral visual stream, the fast magnocellular pathway could feedback information to guide the processing of the slowest parvocellular information in the preceding stages.

Within extrastriate areas in the ventral visual pathway, populations of neurons would tend to strengthen their connections when co-activated by groups of objects that tend to appear simultaneously. In performing our task that requires to respond only to an animal target, top-down preparation of the visual system is presumably maximal and this preparation would extend to contextual scenes in which animals are commonly seen. Through parallel processing, an animal in a natural scene would be the expected optimal stimulus and would co-activate multiple populations of neurons (Figure 6A) that are usually co-activated. On the other hand, when the animal appears in an urban scene, a conflict would arise between populations of neurons that respond to elements of the scene that are not usually co-activated (animal and urban man-made features). Such conflict might range from moderate (when some expected natural features are present in an urban background, e.g., Figure 6B) to extreme (Figure 6C). The more incongruent features in the background of the scene, the greater the competition between the neuronal responses to the background and the neuronal responses to the animal target. Facilitation would arise between populations of neurons that have reinforced mutual connections because they tend to fire simultaneously (Hebb, 1949); such learning of visual covariations has been shown to be implicit (Chun & Jiang, 1999; Jiang & Chun, 2001). Otherwise, interference would take place. Hence, with strong interference, the conflict between go and no-go responses would take longer to resolve or might lead to an incorrect motor decision at the level of the prefrontal cortex (Rousselet et al., 2002; Rousselet, Thorpe, & Fabre-Thorpe, 2004). It has recently been postulated that the parahippocampal cortex (PHC) could mediate the representation of familiar object associations (Aminoff, Gronau, & Bar, 2007; Bar, 2004; Bar & Aminoff, 2003). Thus, the conflict might be present all along the visual stream; it might be maximal in the PHC that receives information directly from the ventral visual stream (Suzuki, 1996; Suzuki & Amaral, 1994) and would encode recurrent regularities or associations in our surrounding world. The conflict could thus arise in the first feed-forward sweep of the earliest available visual information and might explain why the interaction between object and background can be observed even on the earliest behavioral responses that have been suggested to depend mostly on feed-forward processing (Fabre-Thorpe et al., 2001; Macé et al., 2005; VanRullen & Thorpe, 2002).
Implications and conclusion

In the light of these new results, it appears that objects and scene context are processed in parallel and engaged in bidirectional interactions. Among the three possible models proposed by Henderson and Hollingworth (1999), our results rule out the functional isolation model in which object identification is not influenced by expectations from scene context, and the priming model in which contextual influences occur during decisional stage “when a structural description of an object token is matched against long-term memory representation”. On the other hand they support the perceptual schema model, which describes object and context processing interacting at perceptual stages. The immediate effect of context congruence on object categorization using briefly flashed scenes is also compatible with a feed-forward wave of processing in which facilitation and interference between neuronal populations depends on whether they are usually co-activated or not.

Acknowledgments

We thank two reviewers for their valuable comments on the manuscript. This work was supported by the CNRS (Centre National de la Recherche Scientifique), by the French government (Ministère de la recherche et de l’Enseignement supérieur) and by the Fondation pour la Recherche Médicale.

Commercial relationships: none.
Corresponding author: Michèle Fabre-Thorpe. Email: michele.fabre-thorpe@cerco.ups-tlse.fr. Address: Centre de Recherche Cerveau et Cognition, UMR 5549 (CNRS-Université Paul Sabatier Toulouse 3), 133 route de Narbonne, Faculté de Médecine de Rangueil, 31062 Toulouse CEDEX9, France.

References


