SPEED LIMITS IN NATURAL SCENE DISCRIMINATION

Speed Limits: Orientation and Semantic Context Interactions Constrain Natural Scene Discrimination Dynamics

http://www.apa.org/journals/xhp/homepage.html

Running head: SPEED LIMITS IN NATURAL SCENE DISCRIMINATION

Speed Limits: Orientation and Semantic Context Interactions Constrain Natural Scene Discrimination Dynamics

Jochem W. Rieger, Nick Köchy, Franziska Schalk, Marcus Grüschow, Hans-Jochen Heinze

Otto-von-Guericke-University

Magdeburg

Correspondence should be addressed to:

Jochem Rieger

Otto-von-Guericke University

Dept. of Neurology II

Leipziger Str. 44

39120 Magdeburg

Germany

Email: jochem.rieger@nat.uni-magdeburg.de

Phone: ++49-391-6721740

Fax: ++49-391-67290224

Abstract

The visual system rapidly extracts information about objects from the cluttered natural environment. In five experiments we quantified the influence of orientation and semantics on the classification speed of objects in natural scenes, particularly with regard to object-context interactions. Natural scene photographs were presented in an object discrimination task, and pattern-masked with various scene-to-mask stimulus onset asynchronies (SOAs). Full psychometric functions and reaction times (RTs) were measured. We found that: 1) Rotating the full scenes increased threshold SOA at intermediate rotation angles but not for inversion. 2) Rotating object or context degraded classification performance in a similar manner. 3) Semantically congruent contexts had negligible facilitatory effects on object classification compared to meaningless baseline contexts with a matching contrast structure, but incongruent contexts severely degraded performance. 4) Any object-context incongruence (orientation or semantic) increased RTs at longer SOAs, indicating dependent processing of object and context. 5) Facilitatory effects of context emerged only when the context shortly preceded the object. We conclude that the effects of natural scene context on object classification are primarily inhibitory and discuss possible reasons.

Keywords: natural scenes, context, orientation, semantics, psychophysics

Speed Limits: Orientation and Semantic Context Interactions Constrain Natural Scene Discrimination Dynamics

Natural scene processing has been shown to be extremely fast and efficient (Potter, 1975; Rieger, Braun, Bülthoff, & Gegenfurtner, 2005; Schyns & Oliva, 1994). Pattern masking has shown that 24 ms undistorted processing are sufficient to encode enough information for better than chance recognition of natural scene photographs, and after 90 ms nearly perfect recognition rates are obtained (Rieger et al., 2005). The nature and the level of complexity of the information extracted from scenes during these short intervals are currently not well understood. Many theories of scene processing (Bar, 2004; Biederman, Mezzanotte, & Rabinowitz, 1982; Ullman, 1996) suggest that there are both facilitatory and inhibitory effects of the scene context on object processing. Inhibition is assumed to occur due to violations of learned object-context relations, while for facilitatory effects to occur the object and the scene must be in concordance. But what are the important attributes?

Effects of orientation in object classification

The category classification of objects in scenes (e.g. animals, faces or man-made objects) appears to be immune to massive manipulations of relatively "low-level" perceptual scene attributes such as color (Delorme, Richard, & Fabre-Thorpe, 2000; Gegenfurtner & Rieger, 2000) and orientation (Rousselet, Mace, & Fabre-Thorpe, 2003; Vuong, Hof, Bülthoff, & Thornton, 2006). Rousselet et al. (2003) found only minimal effects of scene inversion on the classification of the objects embedded in briefly flashed natural scenes (faces and animals), and Vuong et al. (2006) found that inverting scenes had no effects on the detection of humans in still images and short movie clips.

These findings are surprising in the light of data that indicate that the recognition of isolated objects requires increasingly more time as they are rotated further away from their canonical or learned orientation (Jolicoeur, 1985; Tarr & Pinker, 1989; Yin, 1969). This orientation dependence has been taken as evidence that successful recognition requires the mapping of the bottom up processed object representation to a memory model that is stored in a canonical (upright) orientation (Tarr & Bülthoff, 1998; Ullman, 1996). Nevertheless, reports exist that inverted isolated objects are identified faster than objects at intermediate picture plane rotations (Jolicoeur, 1990a, 1990b; Lawson & Jolicoeur, 2003; Murray, 1997). To account for this, it has been hypothesized that the matching process requires slow rotation at intermediate rotation angles, but that 180° inverted objects can be identified using some faster flipping compensation process (Murray, 1997). Alternatively, it has been suggested that orientation compensation can be achieved in the population code of object specific neurons tuned to various orientations (Ashbridge, Perrett, Oram, & Jellema, 2000; Perrett, Oram, & Ashbridge, 1998). In this theory, the speed at which disoriented objects are recognized would depend on the relative number of neurons that represent an object at a given orientation. The more neurons are involved, the faster evidence accumulation would proceed, and the faster the object is recognized.

However, orientation-specific object features (e.g. the axis of elongation or the spatial relation of component features) may not be the only information that supports the recognition of disoriented objects. In addition to orientation compensation rotation-invariant features of mid-level or low-level complexity might contribute to the recognition of rotated objects (DeCaro & Reeves, 2000; Jolicoeur, 1990a, 1990b; Murray, Jolicoeur, McMullen, & Ingleton, 1993; Ullman, 1996; Ullman, Vidal-Naquet, & Sali, 2002). Fully or partly orientation invariant features (e.g. an eye, the wheel of a car, or a mouth) could be extracted quicker than it takes to match the orientation dependent configuration of the object with stored knowledge (DeCaro & Reeves, 2000; Rousselet et al., 2003; Ullman, 1996). When processing time is short the rapid detection of a few informative features in the object or the scene context could help or even suffice to accomplish tasks that require only a coarse level of specificity such as object detection and object discrimination (DeCaro & Reeves, 2000; Grill-Spector & Kanwisher, 2005). But the usefulness of orientation invariant features might depend on several variables, for example the complexity and the informativeness of the stimulus material (Dickerson & Humphreys, 1999; Hamm & McMullen, 1998) or the complexity of the task. Orientation effects are typically greater when the experimental paradigm involves judgments of the spatial configuration (e.g. the object handedness or orientation) of a disoriented object (e.g. Corballis, 1988; DeCaro & Reeves, 2000), or sets of highly similar objects (e.g. Humphreys, Riddoch, & Quinlan, 1988; Lawson & Jolicoeur, 1998), or object naming at a high level of specificity (e.g. Dickerson & Humphreys, 1999), suggesting that paradigms involving these factors put more weight on time consuming orientation compensation.

Due to differences in the stimulus material (relatively simple line drawings of objects vs. photos), and the different levels of task complexity it is currently not clear whether orientation compensation can play a role for the recognition of objects embedded in natural scenes. These stimuli provide a wealth of information beyond object shape from both the object and its scene context. This information might be sufficient for orientation-independent classification and detection as suggested by recent studies (Rousselet et al., 2003; Vuong et al., 2006). Moreover, to our knowledge, nothing is currently known about the relative importance of object and context orientation for object recognition with natural scenes.

Context influences on object classification

In many mid- and high-level theories of visual processing, scene context is assumed to have an influence on object recognition, and the receptive fields of object selective cells in macaque IT shrink drastically, when target objects are presented in a natural context or in spatial proximity to a distractor object (Rolls, Aggelopoulos, & Zheng, 2003). Besides features like position and relative size, semantics of the context may play an important role in object discrimination (Biederman et al., 1982; Boyce, Pollatsek, & Rayner, 1989). However, the nature and the level of this object-context interaction is a matter of dispute. Biederman et al. (1982) and others (Bar, 2004; Henderson & Hollingworth, 1999; Kosslyn, 1994; Ullman, 1996) have suggested a facilitatory influence of semantically congruent contexts on object recognition at several processing stages: It has been proposed that congruent contexts may facilitate bottom up processing (Biederman, 1988; Schyns & Oliva, 1994), or may reduce the search space for the memory match of the object at an intermediate processing level (Bar, 2004; Ullman, 1996). In contrast, Henderson and Hollingworth (1999) suggested that context exerts its effect rather late, when semantic knowledge is accessed after the successful match.

However, for scene-context semantics to influence bottom up object processing or the matching of perceptual representations, it must be very quickly accessible in the processing stream (Bar, 2004; Biederman et al., 1982; Boyce et al., 1989). Bar (2004) suggests that the “gist” of a scene is processed very quickly using low spatial frequencies, and serves to restrict the search space for possible objects. Hollingworth and Henderson (1998), on the other hand, found no object-scene congruency advantage when they repeated the Biederman et al. (1982) study and controlled for position cues and response bias. They concluded that object identification is isolated from scene context. In these studies, outline drawings of scenes were used to investigate context effects.

A newer study by Davenport and Potter (2004) using natural scene photographs reported that semantic scene congruency has an effect on the processing of both the target object and its context. Interestingly, these investigators found that accuracy for objects presented on a uniform white background was better than for objects presented in either a congruent or an incongruent background. This finding can be taken to imply that any contextual information has an inhibitory effect, or that clutter at the border of objects in scenes interferes with object processing. As only one single scene-mask SOA (80 ms) was used in this study nothing is known about the dynamics of these interactions. Ganis and Kutas (2003) provide evidence from EEG that the effects of object-context incongruence occur only after 300 ms, possibly at the level of semantic analysis, after bottom-up processing.

Although both studies noted above report an effect of scene context, they provide very limited information about the dynamics of the accumulation of information and the (inhibitory and/or facilitatory) nature of the object/scene-context interaction. Moreover, while "high-level" cognitive context semantics might be expected to exert their effect at relatively late object processing stages, nothing appears to be known about the effects of context orientation, a relatively "low-level" physical scene attribute.

The present approach

Here we report investigations of the dynamics of the influence of both low-level (orientation) and high-level (semantic) context properties on the discrimination of objects embedded in photographs of natural scenes. Pattern masking was used to vary the duration of information accessibility. The experiments reported determine whether scene orientation can affect object discrimination when scenes are rotated at different angles (Experiments 1 & 2) and explore the interaction between object and context processing when their orientation is manipulated independently (Experiments 3 & 4). In addition, we probed the dynamics of the influence of high-level semantic context features on object discrimination and tried to distinguish facilitatory and inhibitory context effects (Experiment 5). Because we simultaneously acquire both accuracy and reaction time data over a range of scene-mask intervals, we are able to relate these measures across several levels of task difficulty. The results of our experiments shed new light on the dynamics of object context interactions in natural scenes.

Experiment 1

In the first experiment we tested whether the discrimination of objects embedded in a natural background is immune to rotation, as suggested by results from Rousselet et al. (2003), who found no effect of scene inversions on object discrimination. Because it is known that isolated objects rotated by 180° can be faster identified than objects rotated at intermediate rotation angles (Murray, 1997), we tested for possible *rotation effects* by presenting photographs of natural scenes at three orientations: upright, rotated 90°, and rotated 180° (inverted). Subjects' discrimination of objects embedded in these scenes was measured in a two alternative forced choice (2AFC) task.

Methods

Participants.

Fifteen right-handed participants volunteered for the study (seven female). All of them received payment or course credits for their participation. Ages ranged from 19 to 30 years (average 26.5 years). All had normal or corrected to normal vision, and were naive as to the purpose of the experiment.

Stimuli and apparatus.

Colored scene photographs were taken from a commercial database (Corel Stock Photo Library). The photographs (size 256x256 pixel with 24 bits per pixels resolution) were chosen to contain a clearly identifiable object around the center. The edges of the images were vignetted to obtain a smooth transition into the background. The background was set to the mean pixel value of the full set of images. Target photos contained a readily identifiable animal (mammals, reptiles, birds etc.). Animals that had no clear canonical orientation (e.g. butterflies) were excluded. Distractor photos contained a non-animal object (rocks, mushrooms, cars etc.). Care was taken to match the spatial structure and the luminance of the two classes of stimuli. Examples of the stimuli are shown in Figure 1B.

The photos were displayed in a dark room on a 21” Sony GDM-F520 CRT-screen, connected to a Matrox Parhelia graphics board in a standard PC at 1600*1200 pixels resolution and 100 Hz refresh rate. The stimulus sequence was controlled by the Presentation stimulus software (National Institute of Neurological Disorders and Stroke, version 0.53) and synchronized to the vertical refresh of the monitor. A chin rest was used to maintain a constant viewing distance of 60 cm. The photographs subtended 9.5 deg visual angle with the centers of the two simultaneously presented photos 5.2 deg lateral to the fixation cross. The RGB-to-luminance functions of all three channels were measured with a PR650 Spectrascan spectroradiometer (Chatsworth, CA). This information was used to compare the mean luminances of the photos in the two groups. Participants gave responses on a standard keyboard. Responses were recorded by the presentation software.

Procedure.

Before the experiment started, participants received instructions in written form. After reading the instructions they received one block of practice trials to familiarize them with the paradigm and to allow them to establish their criteria.

Insert Figure 1 A-G here

Each experimental trial started with a fixation cross. After 300-850 ms two natural scenes were presented simultaneously, one to the left and one to the right of a fixation cross (Figure 1A). The scenes remained on the screen for various durations and were then replaced by colored pattern masks at the same locations. The masks were created by overlaying successively smaller rectangles with different orientations. This type of pattern mask mimics contrast distribution in natural scenes and proved to be the most effective in a previous study on the dynamics of natural scene processing (Rieger et al., 2005). Eleven different masks were used. The scene-mask stimulus onset asynchronies (SOAs) were 10 ms, 30 ms, 50 ms and 70 ms; all were multiples of the refresh rate. Both scenes contained a clearly identifiable object approximately in the scene center. On each trial one of the scenes contained an animal. The participant's task was to indicate by a button press whether the animal was in the left or the right scene (spatial 2AFC). Scenes containing animals were presented with equal probability on the left and right. Both scenes had the same orientation, and 90° rotated scenes were rotated at equal probability either clockwise or counterclockwise. Participants were instructed to respond on every trial as quickly and accurately as possible, and to guess when in doubt. The response time window (1200 ms) was indicated by a color change of the fixation cross. The next trial was automatically initiated when the response time window ended. Each SOA was presented 40 times with a different pair of photos in a random sequence. For each participant new pairs of photos were assembled and randomly associated with different SOAs. Each of the 960 images was presented only once to a participant. The 480 trials per participant (40 pairs * 4 SOAs * 3 orientations) were presented in 2 blocks.

Analysis.

The proportion of correct responses was calculated for each SOA. Weibull functions were fitted with the Psignifit software package (Wichmann & Hill, 2001a) to determine the threshold SOA, the slope of the function at the 75% correct discrimination level, and the respective two sided 95% bootstrapped confidence intervals. Significant differences between conditions for the fitted parameters are indicated by non-overlapping confidence intervals. Differences in the parameters of interest were additionally tested by pair-wise Monte-Carlo tests (Wichmann & Hill, 2001b). The Monte-Carlo tests seek to determine the probability that the observed threshold and slope differences are generated by the same underlying process instead of two different processes. In this test the data measured in two different conditions are fitted separately with a psychometric function and the difference of the slopes and thresholds are determined. The parameters of the psychometric function obtained when both data sets are lumped together are then used in a Monte-Carlo process (at least 2000 repetitions) to create the distribution of differences expected when the datasets are created by the same underlying binomial process. Threshold and slope differences that exceed 95% of the simulated differences are considered as significantly different on a 5% alpha-level. Details and validity tests of the method are given by Wichmann and Hill (2001b). The results from these tests will be denoted “Monte-Carlo test” (MCT). Mean reaction times for the different SOAs and conditions were also analyzed. Only correct responses were included in the reaction time analysis, since false responses are not informative about the conditions for successful processing of the stimulus material. We report partial eta squared values (η^2) as a measure of relative treatment effect sizes in the ANOVA-tests (Cohen, 1988) together with the mean squared error (MSE). The eta square measure reflects the proportion of the effect-plus-error-variance that is attributable to the effect.

Results

Image statistics.

The mean luminance did not differ significantly between the animal and non-animal photographs (animal: 16.31 cd/m², non-animal: 16.59 cd/m², t_{1918}= 0.359; p>0.05).

Psychophysical performance.

In Figure 2A we show the proportions of correct responses and the fitted psychometric functions for discrimination of upright, 90° rotated, and 180° rotated natural scenes. The exact threshold SOAs and slopes obtained from the psychometric functions and their confidence intervals are given in Table 1. For the upright and the 180° rotated scenes, the threshold SOAs required to obtain threshold discrimination are very short (31.9 ms and 34.6 ms) and do not differ (MCT, p>0.05). Rotating the scenes by 90° significantly prolonged this threshold SOA (to 54.9 ms) relative to both these conditions (MCT both comparisons, p<0.01).

Insert Figure 2 A/B here

The slopes of the psychometric functions differ between the upright and rotated conditions. The slope is highest for upright scenes and shallower for inverted and 90° rotated scenes (MCT, p<0.01 in both cases), indicating an increased trial-by-trial variability for both scene rotations. The slopes in the two rotated conditions do not differ significantly (MCT, $p>0.05$).

Insert Table 1 here

The reaction times are shown in Figure 2B. We used a within subject ANOVA (3 orientations x 4 SOAs) to assess the effects of SOA and scene orientation on RTs. Both scene orientation and SOA produced significant main effects (orientation: $F_{2,28}=8.2$, $p<0.01$, $\eta^2=0.37$, MSE=1674.9; SOA: $F_{3,42}=11.4$, $p<0.01$, $\eta^2=0.44$, MSE=1110.8). The fastest reaction times for correct responses were obtained with the upright scenes. A significant interaction between orientation and SOA ($F_{6,84}=2.5$, $p<0.05$, $\eta^2=0.15$, MSE=930.3) indicates that RT decreases more quickly with upright scenes than with rotated scenes (Figure 2B). The RTs for the rotated scenes are slower, do not differ significantly from each other, and do not seem to decrease with longer SOAs (SOA: $F_{3,42}=2.6$, $p>0.05$, $\eta^2=0.16$, MSE=1153.8; rotation: $F_{1,14}=1.2$, $p>0.1$, $\eta^2=0.08$, MSE=914.9; no interaction). This change of RT does not reflect a speed accuracy trade-off, since such a trade-off would result in a negative correlation between error rate and RT. For upright and inverted scenes the error rates and RTs are significantly positively correlated (upright: $r=0.6$, $p<0.01$; inverted: $r=0.29$ $p<0.05$). With the 90° rotated scenes, a positive correlation is at the margin of significance ($r=0.24$, $p=0.064$).

Discussion

Our data replicate and extend previous findings on the influence of rotation on the classification of objects embedded in natural scenes. Concordant with previous studies of Rousselet et al. (2003) and Vuong et al. (2006), who only tested upright and inverted scenes, our threshold SOAs for the discrimination of objects in the inverted scenes did not differ significantly from upright scenes. This was the case although we used a fundamentally different paradigm. However, the threshold SOAs were substantially prolonged when the scenes were picture-plane rotated by 90°. This accords with reports from studies using isolated objects (Lawson & Jolicoeur, 2003; McMullen, Hamm, & Jolicoeur, 1995; Murray, 1997) in which performance showed the greatest deterioration at intermediate rotation angles. Our results therefore demonstrate that the duration of undistorted processing (the scene-mask SOA) required for successful object discrimination is strongly dependent on scene orientation, at least for 90° rotated scenes.

An open question is the nature of the information used for the classification of the scenes (e.g. Grill-Spector & Kanwisher, 2005). Results from psychophysical studies (DeCaro & Reeves, 2000, 2002; Dickerson & Humphreys, 1999) using isolated objects and computational (Ullman et al., 2002) studies suggest that, during object processing, initially isolated low- or mid-level complexity features that allow a coarse classification are extracted, followed by more complex information that depends on the spatial configuration of the objects and allows for classification or identification at higher specificity (DeCaro & Reeves, 2000; Dickerson & Humphreys, 1999). In addition, Jolicoeur (1990a) has already suggested that such isolated features might also have some intrinsic orientation specificity, and speculated that some isolated features or categorical relations between features have a vertical symmetry axis that would lead to the reduced rotation effects found with inverted objects.

For configurational information to contribute to recognition, a time consuming normalization process may be required when the scenes are disoriented (Dickerson & Humphreys, 1999; Jolicoeur, 1990a). This would predict in our experiment that differences between upright and rotated scenes should evolve over time. In support of these theories, the RTs are similar over orientations at 10 ms and 30 ms SOA and diverge at longer SOAs (50 ms and 70 ms). The effect is mostly due to the RT-decrease for upright scenes, since no effect of SOA on RT was found for the rotated scenes. This RT reduction with upright scenes is expected at longer SOAs, because many choice models that incorporate temporal evidence accumulation (e.g. Hick, 1952; Perrett et al., 1998; Usher, Olami, & McClelland, 2002) predict shorter RTs when the decision process has more information available. Bacon-Mace, Mace, Fabre-Thorpe, and Thorpe (2005) found a similar effect using masked presentations of natural scenes. The lack of a RT-reduction with the rotated scenes at longer SOAs (50 ms and 70 ms), despite the increasing accuracy, suggests that orientation compensation prolongs the processing when information is extracted that allows for configuration dependent processing (Dickerson & Humphreys, 1999; Jolicoeur, 1990a). The slopes of the psychometric functions draw a similar picture as the RTs, with both indicating slower information uptake from the rotated photograph. A more thorough discussion of potential interpretations of the behavioral parameters will be presented in the general discussion.

The orientation effects we found in our object discrimination task with photos of natural scenes were consistent with those obtained with line drawings of isolated objects (e.g. McMullen et al., 1995; Murray, 1997). This was unexpected in the light of previous studies that suggested that the classification or identification of objects in natural scene photographs might be immune against orientation effects (Rousselet et al., 2003; Vuong et al., 2006). In the next experiment we aimed to investigate the contribution of the scene context to the discrimination of objects embedded in natural scenes at various orientations. The context is an obvious difference between our stimuli and the line drawings of isolated objects used in most previous studies, and it may provide additional information that is helpful in the classification task (e.g. Bar, 2004; Biederman et al., 1982). Furthermore, it would be desirable to have more orientations tested to obtain a better parametric description of how the scene discrimination depends on orientation.

Experiment 2

In this experiment we investigated the *orientation effects* in scene classification using a wider range of angles, and, most importantly, compared the orientation effects obtained with *isolated objects* to those obtained with *objects embedded in a scene background*. We thought that the increase in the number of investigated angles might allow for a better comparison to the orientation effects obtained in previous investigations using line drawings of isolated objects (e.g. Dickerson & Humphreys, 1999; Jolicoeur, 1985). Furthermore, we wanted to explore how the *scene-context* may contribute to the rapid discrimination of objects in natural scenes. This context could have negative, as well as positive effects. On one hand, it could provide additional orientation invariant information as well as information about the orientation of the object, thereby speeding up the orientation compensation process. On the other hand, a structured context could impede object recognition, for example by slowing down figure-ground segmentation (Geisler, Perry, Super, & Gallogly, 2001).

Methods

We used the same paradigm as in the first experiment (Figure 1A), and thresholds were determined in the same manner. Here we report only the procedural changes.

Participants.

Fifteen new participants (7 female) were recruited for this experiment. The mean age was 27 years (from 20 to 42 years).

Stimuli.

As in the first experiment, we selected colored scene photographs containing clearly identifiable animal and non-animal objects with a clear canonical orientation from a commercial database (Corel Stock Photo Library). In addition, the same two classes of isolated objects (animals and man-made objects such as cars, camping chairs, baskets, boats etc.) were selected from a database (Hemera Photo-Objects) or extracted from additional scene photographs. The size and type of the isolated and the context-embedded objects were approximately matched. The stimuli were presented at five orientations (0°, 45°, 90°, 135°, and 180°), and the rotation direction (clockwise and counterclockwise) was randomized with equal probability for both directions. Both scenes in a pair were rotated in the same direction, and each was presented in a simulated octagonal window with vignetted borders. The screen background (as well as the background of the isolated objects) was set to the mean color of all scenes. The image pairs were presented with 20 ms, 40 ms, 70 ms and 120 ms scene-mask SOA. In total we presented 1200 isolated objects and 1200 scenes. The scenes were randomly distributed over the different combinations of orientation and SOA, and every combination was repeated 30 times with different scene or isolated object pairs. In addition, we inserted a condition in which only the mask was presented (0 ms SOA) to control for possible bias effects. Examples of the scene-embedded and the isolated animal-stimuli are shown in Figure 1C.

Results

Image statistics.

The mean luminance did not differ significantly between the animal and non-animal photographs (animal: 16.31 cd/m², non-animal: 16.59 cd/m², t_{1198}= 0.359; $p>0.05$).

Psychophysical performance.

In Figure 3 we show the discrimination threshold SOAs obtained with upright, 45°, 90°, 135° and 180° rotated natural scenes (Figure 3A) and isolated objects (Figure 3C). The exact threshold SOAs and slopes obtained from the psychometric functions and their confidence intervals are listed in Table 2. The threshold SOAs for the upright and the 180° rotated scenes were short (35.6 ms and 41 ms) and did not differ (MCT, $p>0.05$). Rotating the scenes by 90° significantly prolonged this threshold SOA (to 46.3 ms) relative to the upright condition (MCT, $p<0.05$). Scene inversion and intermediate rotation angles (45° and 135° rotation) did not prolong the threshold SOA significantly (all MCT, $p>0.05$) compared to upright scenes. These results reproduce and extend the finding from the first experiment, that 90° scene rotation significantly prolongs the threshold SOA, while scene inversion has no significant effect.

When isolated objects were rotated we found no effect of orientation on the threshold SOA, as indicated by the overlapping confidence intervals (Table 2) and the non-significant Monte Carlo simulations (all MCT upright vs. rotated $p>0.05$).

A paired t-test over all orientations revealed that isolated objects required significantly shorter threshold SOAs than context-embedded objects (mean threshold isolated: 33.9 ms, and mean context embedded: 40.3 ms; $t_4=2.8$, $p<0.05$).

Insert Table 2 here

The slopes of the psychometric functions did not differ between the upright and rotated conditions for the context-embedded objects. The slopes for the upright isolated objects tended to be steeper than for the other orientations, indicating an increased trial-by-trial response variability for rotated objects. Since only one of the two tests reaches significance (all tests upright vs. rotated MCT, $p<0.05$ but confidence intervals overlap), we consider this effect as marginal. The mean slopes did not differ between the isolated and the context embedded objects (0.011 proportions correct per ms and 0.0089 proportions correct per ms, respectively). However, for the isolated upright objects the slope tended to be higher than for upright the context-embedded objects (MCT, $p<0.05$ but overlapping confidence intervals).

Insert Figure 3 A-D here

The RTs for objects in scene contexts are shown in Figure 3B, and for isolated objects in Figure 3D. We analyzed the RT data with a three factor within-subject ANOVA, with context/isolated, orientation (5 levels) and SOA (4 levels) as the factors. The RTs from the mask-only condition were omitted from this analysis. The ANOVA revealed a main effect of orientation ($F_{4,56}=5.09$, $p<0.005$, $\eta^2=0.26$, MSE=1124), and no other significant main effects (context: $F_{1,14}=1.0$, $p>0.25$, $\eta^2=0.07$, MSE=4650; SOA: $F_{3,42}=0.89$, $p>0.25$, $\eta^2=0.06$, MSE=13296). There was no significant interaction between context/isolation and orientation

($F_{4,56}$=1.9, $p>0.1$, η^2=0.12, MSE=736), indicating that orientation had a qualitatively similar effect on the RTs obtained with isolated and context-embedded objects. In contrast, context/isolation and rotation both interacted with SOA ($F_{3,42}$=3.2, $p<0.05$, η^2=0.19, MSE=1001, and $F_{12,168}$=2.9, $p<0.005$, η^2=0.17, MSE=745). Since we found no three way interaction ($F_{12,168}$=0.56, $p>0.5$, η^2=0.04, MSE=642), we analyzed the two significant interactions separately. The average RT-differences between context-embedded and the isolated objects provide information about the context/isolation interaction. These differences peak at intermediate SOAs with isolated objects leading to faster RTs (mean: Δ10.4 ms and Δ15.6ms at 40 ms and 70 ms SOA, respectively). At the longest and shortest SOAs the differences are small (mean: Δ1.2 ms and Δ-4.9 ms at 20 ms and 120 ms SOA, respectively). The interaction of rotation with SOA may be attributable to a more rapid drop in RTs with upright than with rotated stimuli as the SOA increases (see Figures 3B and D). In accordance with this hypothesis, we found no significant interaction when the RTs obtained from upright presentations were removed from the ANOVA ($F_{9,126}$=1.0, $p>0.25$, η^2=0.07, MSE=676), but the interaction persisted when any other orientation was removed (45°, 90°, 135°, 180° all at least $p<0.005$). This pattern replicates the RT-effects found in the first experiment. We found no indications for a speed-accuracy trade off (all regressions $p>0.2$). Note that the longest RTs have been found with 135° rotation angles at the longest SOA when accuracy is high.

Discussion

Experiment 2 replicated the most important effects found in Experiment 1. Compared to upright scenes the threshold SOAs for 90° rotated scenes increased, and RTs decreased faster at longer SOAs with upright than with rotated scenes. The results of this experiment also indicate that the metal rotation-like processes play only a minor role for orientation compensation in natural scene discrimination. Moreover, our results suggest that scene context has a detrimental effect on the discrimination of objects that is not only due to slowed figure-ground segmentation.

Only a weak contribution of mental rotation.

Some theories of object recognition assume that orientation compensation is required for the recognition of objects that deviate from their most likely, canonical orientation (Jolicoeur, 1985; Ullman, 1996). The time required for a compensation process like mental rotation is a monotonically increasing function of the extent of the disorientation (e.g. Jolicoeur, 1985; Shepard & Metzler, 1971). Our data provide several arguments against the assumption that a simple mental rotation account is sufficient to explain how stimulus rotation influences the discrimination of natural objects embedded in natural scenes or presented in isolation. First, the effects of rotation on object discrimination accuracy (threshold SOA) peaked at the 90° rotation. Second, the effect of rotation on RT peaked at 135° but tended to be weaker at 180° rotation when the scenes are discriminated with high accuracy (120 ms, Figure 3B and D). Such effects have been described previously, and attributed to a change in the processing strategy at 180° rotation, for example a fast flipping of the image (Murray, 1997), or the contribution from a second, rotation independent object recognition system (Jolicoeur, 1990a). Third, although the RT increases linearly with up to 135° rotation ($r=0.99$, $p<0.05$), the rotation speed derived from the regression is 2801 deg/s at 120 ms SOA which is much higher than the 1000 deg/s that were suggested as an upper bound for mental rotation (Cohen & Kubovy, 1993; Hamm & McMullen, 1998). While these results do not permit rejection of the assumption that orientation compensation can play some role in the discrimination of natural objects embedded in natural scenes, orientation compensation may contribute less in our experiment than it does with other tasks and other with other stimulus material (Dickerson & Humphreys, 1999; Jolicoeur, 1990a).

Effects of context.

In concordance with this assumption our data indicate that the processing speed of disoriented objects is modulated by the amount of scene-context presented with the object. We found the strongest orientation effect on object-discrimination accuracy when scenes were rotated 90° with the square shaped context in the first experiment. With the octagonal context we used in this experiment, we found a somewhat weaker effect with the 90° rotations. The area of the pictures with the octagonal context is approximately 20% smaller than the area of the square context, and this loss of area occurred primarily in the context. Finally, virtually no rotation effects were found when the objects were presented in isolation, and accuracy was higher with isolated than with context-embedded objects. This result suggests that the increased saliency of the object borders on the uniform background could make object shape more quickly accessible for orientation-independent object discrimination. However, the effect of context on object-discrimination accuracy is most likely not only due to prolonged figure-ground segmentation. If this was the case, one would expect longer RTs for context-embedded than for isolated objects, even at longer SOAs. By contrast, we found that context prolonged the RTs only at intermediate SOAs, when the discrimination accuracy was around threshold. Furthermore, the comparison of the discrimination accuracy obtained with isolated and with context-embedded objects implies that possible beneficial effects of the context are outweighed by the detrimental effects.

The next experiment was designed to further investigate the orientation dependent object scene-context effects. There we wanted to investigate how object and scene context interact with respect to orientation.

Experiment 3

In the third experiment we explored possible *interactions between object and context* that were implied by the results of Experiment 1 and 2. We investigated whether the slower discrimination of objects in rotated natural scenes is caused by the rotation of the object, the context, or both. We tested the following hypotheses: a) If object and context are processed independently with respect to orientation, upright objects should be processed fastest, independent of the context orientation. b) If, on the other hand, object orientation and context orientation are processed jointly, then object discrimination should be degraded even when the object is upright but the context is rotated. Furthermore, c) if context supports object processing, one might expect better object discrimination for rotated objects in upright contexts than in rotated contexts. We tested these hypotheses by rotating the objects and the backgrounds in the scenes independently.

Methods

We used the same paradigm as in the first experiment (Figure 1A). Here we report only the procedural changes.

Participants.

Fifteen new participants (seven female) were recruited for this experiment. The mean age was 25.4 years (from 20 to 33 years).

Stimuli.

Two classes of isolated objects (animals and man-made objects such as cars, chairs, boats etc.) were selected from a database (Hemera Photo-Objects) or extracted from photos. Adobe Photoshop was used to insert the objects into scene photographs that served as contexts. The size of the objects was approximately matched to the spatial scale of the backgrounds, and average object sizes were matched between the four orientation conditions. Prior to the insertion of objects in the scene backgrounds, some backgrounds and objects were rotated. We produced images with the following combinations: object upright/context upright (T0 C0), object 90° rotated/context upright (T90 C0), object upright/context 90° rotated (T0 C90), object 90° rotated/context 90° rotated (T90 C90). Examples are shown in Figure 1D.

Altogether 480 objects (240 in each category) were embedded into 960 contexts. Thus, each object was presented twice with a different context.

Insert Figure 4 A/B here

The manipulated scenes were randomly distributed over 4 different SOAs (10 ms, 30 ms, 60 ms and 120 ms). In addition we inserted a condition in which only the mask was presented (0 ms SOA) to control for possible bias effects. Each SOA was presented 30 times in random order.

Results

Image statistics.

There was no difference in the mean luminance of photos containing animals and man-made objects (animal: 18.4 cd/m², non-animal: 17.8cd/m², t_{958}= 0.2; p>0.05), and no difference in the sizes of the embedded objects (animal: 6160.5 pixels, non-animal: 6384.0 pixels, t_{478}=0.47; p>0.05).

Psychophysical performance

Insert Table 3 here

The proportion of correct responses at the different SOAs and the fitted psychometric functions are shown in Figure 4A. The threshold SOAs and slopes of the psychometric functions at 75% correct discrimination performance are given in Table 3. The shortest threshold SOA (36.2 ms) was obtained when both the background and the object were upright. A MCT confirmed that the threshold in this upright condition was significantly different from the threshold in the rotated conditions (all p<0.001) When objects, backgrounds, or both were rotated, thresholds were prolonged (T90 C0=56.9 ms; T0 C90=61 ms; T90 C90=63.7 ms), but pair wise comparisons failed to reveal any significant threshold difference between the rotated conditions (MCT: T0 C90 vs. T90 C90: $p>0.05$; T0 C90 vs. T90 C0: $p>0.05$; T90 C0 vs. T90 C90: $p>0.05$). The threshold SOAs for both the upright and the fully rotated conditions were similar to the ones obtained in the first experiment. This result indicates that the insertion of the objects into the background scenes produced no side-effects.

The slopes of the psychometric functions showed the same pattern as the thresholds: The slope was steepest when both object and background were upright, and a MCT confirmed a significant slope difference between the upright and the rotated conditions (-0.0087 prop. correct/ms, $p<0.01$). The slopes in the rotated conditions were shallower, and the pair wise comparisons of the psychometric functions failed to reveal any significant differences between them (MCTs: T0 C90 vs. T90 C90: $p>0.05$; T0 C90 vs. T90 C0: $p>0.05$; T90 C0 vs. T90 C90: $p>0.05$).

A within subject ANOVA (4 orientations x 4 SOAs) of the RTs (see Figure 4B) reveals a significant influence of the orientation ($F_{3,42}=3.9$, $p<0.05$, $\eta^2=0.22$, MSE=1317), but not of the SOA ($F_{3,42}=1.16$, $p>0.1$, $\eta^2=0.08$, MSE=3881), and a significant interaction between these two factors ($F_{9,126}=2.37$, $p<0.05$, $\eta^2=0.14$, MSE=891). There is an interesting pattern in the RT-data shown in Figure 4B. The RT appears to drop for the all-upright scenes with longer SOAs, but the RTs increase over the first three SOAs when only the background is rotated. The increase of the RTs in the latter condition appears to be the source of the statistical interaction, since none was found when this condition was excluded from the ANOVA ($F_{6,84}=1.69$, $p>0.1$, $\eta^2=0.14$, MSE=971). When participants accurately classified the scenes (120 ms SOA) we obtained the longest RTs when object and background had incongruent orientations (T90 C0, T0 C90). This increase was significant compared to the congruent orientations ($F_{1,14}=9.4$; $p< 0.01$, $\eta^2=0.40$, MSE=518). Error rate and RT were uncorrelated or positively correlated (T0 C0: $r=0.41$, $p<0.01$), so there was no indication for a speed accuracy trade off.

Discussion

The results of the third experiment reveal that the processing of objects embedded in natural scenes depends strongly on the orientation of the context. The scene-mask SOA (the interval of undistorted processing) required for successful object discrimination is prolonged if the scene background is rotated even when objects are upright. When either the object or context is rotated the slope of the psychometric function decreases and the threshold SOA increases. This argues against independent processing of object and context (Henderson & Hollingworth, 1999), at least with respect to orientation. The inhibitory effect of the rotated background on the discrimination of upright objects is incompatible with parts based recognition approaches (e.g. Biederman, 1987) and hard to explain by template matching (Ullman, 1996) since the target object is already in upright orientation. Our finding is better compatible with theories that assume that object recognition is at least partially orientation-dependent and involves orientation compensation either by rotation or a reference frame. An orientation conflict between object and context could delay object recognition, for example by invoking a competition for orientation normalization or a common reference frame (Hinton, 1981; Jolicoeur, 1990a). The prolonged RTs we found at long SOAs with incongruent compared to congruent object and scene-context orientations further support the assumption that orientation conflicts between scene parts require extra processing time to resolve the conflict.

So far we have found evidence for inhibitory effects of the scene context on object discrimination. It has been suggested that upright contexts facilitate the processing of rotated objects, because the upright context could in principle provide information to guide the orientation compensation of the object (e.g. Basri, 1993; Jolicoeur, 1990a), but we found no such advantage. Rotated objects embedded in upright scene contexts (T90 C0) produced similar SOAs and even longer RTs than all rotated scenes (T90 C90). This results is in concordance with the previous experiment that also indicated inhibitory rather than facilitatory effects of the scene context.

However, in all experiments so far, object and background appeared simultaneously and only for short time intervals. In the next experiment we investigated whether the background orientation can have facilitatory effects on object discrimination when it is presented prior to the object, and predicts the object orientation.

Experiment 4

This experiment was designed to further explore the *dynamics of the perceptual integration of objects and their scene-contexts*. We tested whether the orientation of the scene context would influence the processing of the embedded object when the scene context without the object preceded the presentation of the object within the context.

Prior work with isolated objects has shown that a preceding object can reduce the effect of rotation a second object when both objects have the same orientation, and the objects are presented in close temporal proximity (Gauthier & Tarr, 1997; Graf, Kaping, & Bülthoff, 2005; Jolicoeur, 1990a; Tarr & Gauthier, 1998). Graf et al. (2005) presented two line drawings of natural objects in rapid succession and found lower naming accuracy when the two had incongruent rather than congruent orientations. The effect could not be explained by simple shape priming, because it was found when the successive objects were from different categories, as well as when they had different elongation axes. This finding led us to suspect that a preceding scene-context may have an effect on the recognition of the object, although both are highly dissimilar. The authors suggest that the first object in the sequence adjusts the orientation of an abstract reference frame in which the second object is initially processed.

However, the preceding context may also provide semantic and structural information that could facilitate the recognition of the object (Bar, 2004; Biederman et al., 1982), in addition to orientation priming. Bar (2004) recently proposed a physiologically motivated model of object recognition in which he suggested that the scene context may facilitate the initial processing of the object, for example by limiting the search space during object recognition. However, because the processing of objects in natural scenes is very rapid, this theory would predict a very narrow time window in which the context could have a beneficial effect. Processing delays between object and context could therefore reduce any beneficial effect of the context.

To investigate the dynamics of the object-context interactions we presented the object after various delays in the scene context (Figure 1E). We used the same object-context orientation combinations as in the previous experiment. We took care that the objects were likely to occur in their scene contexts but that those contexts contained no independent information relevant for the object discrimination task. More precisely, the scene contexts presented as pairs were from the same scene class (nature or man-made scenes), and both scenes were likely to contain the target and the distractor object. In this way we sought to follow the approach of Sanocki (1999) and decouple effects due to representations invoked by the scene-context from effects due to independent information about the target picture from the context. The scenes preceded the target objects at various durations. The critical question was whether, and to what extent object discrimination accuracy and RTs would benefit from the scene-contexts in the different object-context orientation combinations. From the results of the previous experiment and the results of Graf et al. (2005), we hypothesized that congruency of object and context orientation is an important determinant of the effects of a preceding context. In particular we expected the following effects. a) Rotated objects in rotated scenes would be recognized more accurately and faster when preceded by a rotated scene-context. b) That upright objects in upright scene-contexts would be recognized faster when preceded by an upright scene-context. c) Benefits from the preceding scene-context would be reduced if that context had a different orientation than the target object.

Methods

Here we used a modified version of the paradigm we used in the first experiment (Figure 1E). We report only the procedural changes.

Participants.

Fifteen new participants (9 female) were recruited for this experiment. The mean age was 26 years (from 21 to 38 years). Six additional participants judged the semantic congruency of object and scene-context.

Stimuli.

Two classes of isolated objects (animals and man-made objects such as cars, boats, baskets etc.) were selected from a database (Hemera Photo-Objects) or extracted from photos. Adobe Photoshop was used to insert the objects into scene photographs that served as contexts. The size of the objects was approximately matched to the spatial scale of the backgrounds, and average object sizes were matched between targets and distractors. The object-background combinations were held fixed but the orientation combination was randomly assigned for each subject. We produced images with the following combinations (the same orientation combinations as in Experiment 3): object upright/context upright (T0 C0), object 90° rotated/context upright (T90 C0), object upright/context 90° rotated (T0 C90), object 90° rotated/context 90° rotated (T90 C90). Examples are shown in Figure 1F. There were two scene context classes: nature scenes and man made scenes. Scenes from the same class were presented as pairs, one containing the target object and the other containing the distractor object. The target-distractor pairings were randomized within class for each subject.

Altogether 960 objects (480 targets and distractors, respectively) were embedded into 960 contexts. The congruency of the object context combinations were rated by six judges and three authors. Pictures with object context combinations that were rated congruent by less than four judges were replaced.

Insert Figure 5 A/B here

The presentation sequence in a trial was similar to that in Experiment 3, except that the SOA between the scenes containing the objects and the mask was fixed at 40 ms, and the scenes containing the objects were preceded by the scene-contexts without the objects. At the scene-mask SOA employed we expected intermediate recognition performance. The duration of the delay between the onset of the scene context and the appearance of object in the scene served as the independent variable (Figure 1E). Each trial began with the presentation of the appropriately oriented pair of scenes-contexts without the objects. After one of four context-object SOAs (0 ms, 40 ms, 80 ms, 160 ms), the scene-contexts were replaced by the same scene photographs with an inserted object. The orientation of the scene-context was maintained through the trial: Upright scenes with an object (T0 C0 and T90 C0) were preceded by the upright scene context, and 90° rotated scenes with objects (T0 C90 and T90 C90) were preceded by the 90° rotated scene. The 0 ms context-object SOA condition served as the baseline for possible facilitatory or inhibitory effects of the preceding scene-context. In this condition the scene-contexts and objects appeared simultaneously, as in Experiment 3. In all cases, the scene with the object was on the screen for 40 ms before being replaced by a pattern mask. The individually combined scene pairs were randomly distributed over the 4 different context-object SOAs for each subject. Each combination of SOA and orientation was repeated 30 times, and consequently each object-context combination was presented once.

Results

Image statistics.

There was no difference in the mean luminance of photos containing animals and distractor objects (animal: 11.06 cd/m², non-animal: 10.94 cd/m², t_{958}= 0.503; p>0.05), and no difference in the sizes of the embedded objects (animal: 9311 pixels, non-animal: 9813 pixels, t_{958}=1.7; p>0.05). On average, over the six participants, 92.4 % (standard error 1.42 %) of the pictures were judged as having a congruent semantic object scene-context. It should be noted that, in addition to the participants, three of the authors also judged the images as congruent, when they constructed the stimuli.

Psychophysical performance.

The mean proportions of correct responses obtained with the different scene-context object SOAs are shown in Figure 5A. A two factor within subject ANOVA with the context-object SOA (4 levels), and the context-object orientation combinations (4 levels) as factors revealed significant main effects of both the SOA ($F_{3,42}$=6.7, p<0.005, η^2=0.33, MSE=0.001) and the orientation combination ($F_{3,42}$=20.9, p<0.001, η^2=0.60, MSE=0.009). These factors did not interact ($F_{9,126}$=1.0, p<0.25, η^2=0.07, MSE=0.006). This result indicates that a preceding natural scene context can improve object discrimination, however, the effect differs between orientation combinations.

In concordance with our hypothesis that orientation congruency is an important determinant of context-object priming Figure 5A indicates that the congruent orientation conditions produced the greatest accuracy improvement when SOA increased from 0 ms to 40 ms. To confirm this effect, we combined the data from the two congruent orientation conditions (T0 C0 and T90 C90) and calculated a two factor within subjects ANOVA over the two shortest SOAs (0 ms and 40 ms). The ANOVA reveals a significant effect of context-object orientation congruency and SOA ($F_{1,14}$=13.22, p<0.005, η^2=0.49, MSE=0.004, and $F_{1,14}$=9.6, p<0.01, η^2=0.41, MSE=0.006, respectively), and a significant interaction ($F_{1,14}$=7.4, p<0.05, η^2=0.35, MSE=0.002). The accuracy benefit produced by presenting the context 40 ms in advance was 9.3% when both the context and target were upright (t_{14}=3.7, p<0.005), and 9.8% when both were rotated 90 degrees (t_{14}=3.4, p<0.005). The performance peak in these congruent conditions occurred at this short context-object SOA. The two conditions with incongruent object-context orientations did not benefit significantly from the 40 ms advance presentation of the context (T90 C0: 2.8%, t_{14}=0.7, p>0.25 and T0 C90: 2.5%, t_{14}=1, p>0.25).

At 80 ms SOA the "benefit" in both incongruent conditions peaked, but still did not reach significance (T90 C0: 6.2%, t_{14}=1.5, p>0.1 and T0 C90: 4.7%, t_{14}=1.9, p>0.05).

At the longest SOA (160 ms) the effect of the preceding presentation of the scene context was reduced in all orientation conditions, although this tendency appears weaker in the all rotated scenes condition (T90 C90). On average, over all SOAs, subjects achieved the best discrimination performance with the all-upright scenes (mean T0 C0: 80.6%) and the worst with the upright backgrounds and the rotated objects (mean T90 C0: 67.5%). The means of the other two orientation conditions were similar (mean T0 C90: 73.6%, T90 C90: 72.2%), but the benefit of the preceding scene-context was higher for the all rotated scenes (T90 C90) than for the upright objects with rotated context (T0 C90). The accuracies obtained with each orientation and SOA combination are listed in Table 4, together with the respective RTs.

Insert Table 4 here

The mean RTs obtained with the different scene-context object SOAs are shown in Figure 5B. The pattern of the RTs is approximately the inverse of the pattern shown by the accuracy data. A two factor within subject ANOVA with the context-object SOA (4 levels) and the context-object orientation combinations (4 levels) as factors revealed significant main effects of both SOA ($F_{3,42}$=4.3 p<0.01, η^2=0.24, MSE=7659), and the rotation combination ($F_{3,42}$=8.3 p<0.001, η^2=0.38, MSE=2648). These factors did not interact ($F_{9,126}$=0.95 p<0.25, η^2=0.06, MSE=1318). A two factor within subject ANOVA in which we pooled the orientation conditions by the congruency of object and background (congruent: T0 C0 and T90 C90, and incongruent: T90 C0 and T0 C90) revealed that subjects responded faster when the scene and object orientations were congruent ($F_{1,14}$=7.6; p<0.05, η^2=0.35, MSE=2381). We found no significant correlation of RT with the error rate in any of the conditions, and therefore no indication for a speed-accuracy trade off.

Discussion

The results of this experiment show that the scene-context can have a facilitatory effect on the recognition of briefly presented objects in natural scenes when the onset of the scene-context precedes the onset of the object. The effect of the scene-context was transient and most pronounced at short SOAs. The facilitatory effect was orientation-congruency dependent, with congruently oriented objects and contexts producing a discrimination accuracy increase of nearly 10%.

The contribution of the preceding scene context to object discrimination.

Our finding that rotated objects in rotated scenes were better and faster discriminated when the scene context preceded the object presentation by a short interval (40 ms) is broadly consistent with the idea that the preceding scene context sets up a reference frame (a kind of a coordinate system) in which the subsequently presented object is processed (Graf et al., 2005; Jolicoeur, 1990b). However, even with a preceding presentation of the context, a rotation effect is still apparent in the accuracy and in the RT data. Subjects discriminate all upright scenes (T0 C0) faster and more accurately than disoriented scenes. As previously discussed in Experiment 1, the slower and less accurate processing of the disoriented scenes might be due to a time-consuming orientation-compensation process (Jolicoeur, 1990b), or, alternatively, to a smaller neuronal population being available to signal the presence of the disoriented target object (Perrett et al., 1998).

The data from the all upright scenes (T0 C0) provide additional information about facilitatory effects of a preceding context. The benefits of a 40 ms preceding context for object discrimination were as strong as with the all rotated scenes (T90 C90). At a first glance this result could be seen as evidence that the scene context contributed independent information to the discrimination task. Several aspects of the data argue against this assumption. First, we found only a non-significant facilitatory tendency for incongruently oriented context object orientation. We think that this was at least partially due to the fact that our stimuli were designed to minimize unwanted independent information that the scene context could convey regarding the location of the target picture (Sanocki, 1999). Second, and most importantly, the facilitatory effect of the background is weaker at longer SOAs when the context is presented longer, which is the opposite of what one would predict when the background is informative for the task. We think that the facilitation we observed with all upright scenes is most likely due to an improvement of the object recognition process, as suggested, for example, by Bar (2004). In Bar's theory of object-context interaction, the facilitation of object processing by the context occurs automatically and requires that information about the scene context is available in the matching stage of object processing. This predicts a facilitatory effect of a preceding semantically congruent context on the discrimination of objects embedded in all upright scenes. However, the theory also predicts that the facilitatory effect is time critical, because the context influence must be established when the processing of the object reaches the matching stage. Incongruent contexts would require some sort of time-consuming orientation compensation on either the object or the context and thus reduce and delay the facilitatory effect, which was the tendency that we found. However, the theory of Bar (2004) cannot explain why we found a smaller context effect with longer SOAs.

The transient nature of the facilitatory scene context effect.

The peak facilitation of object discrimination by the scene context occurs at SOAs around 40 ms to 80 ms. Graf et al. (2005) using line drawings of isolated objects found facilitatory effects in a similar time range. At the longest SOA we used (160 ms) the facilitatory effect of scene context is less or not apparent. Because the effect of scene context is initially facilitatory, pattern masking of the object by the scene context cannot account for it. Pattern masking would predict reduced discrimination accuracy at short SOAs (Bacon-Mace et al., 2005; Breitmeyer, 1984; Rieger et al., 2005), but we found an accuracy increase. An alternative explanation incorporates two different types of object-context interactions:

Perceptual object-context integration during the initial processing that supports the recognition of the object at short scene-object SOAs, and a competitive interaction when the scene context is presented further ahead in time. Competitive interactions between successively presented, meaningful stimuli have been found in a wide variety of paradigms (for reviews see Coltheart, 1999), typically when the stimulus onsets are separated by 100 ms to 400 ms. In this time window a preceding meaningful stimulus can reduce the accuracy of the report of subsequent meaningful stimulus. An idea common to most accounts of this effect (Chun & Potter, 1995; Kanwisher, Yin, & Wojciulik, 1999; Seiffert & Di Lollo, 1997) is that the detection of the second stimulus (here the target object) is impaired at a relatively high and attention-binding processing level. This level is thought to have a limited capacity that is utilized by the preceding stimulus (here the context). However, this explanation of the reduction of the facilitatory context effect at the longer SOAs we used is speculative and requires further testing.

In this experiment we found orientation-congruency-dependent facilitatory effects of scene contexts on briefly presented objects when the context was presented somewhat earlier than the object. Independent of the orientation effects, we also found that the semantically congruent contexts we used increased the accuracy and speeded up the discrimination of objects presented in all-upright scenes, compared to performance levels when the object and scene context were presented simultaneously. Conversely, the results from Experiment 2 suggest that semantically congruent contexts may have inhibitory effects, compared to uniform, meaningless backgrounds. Therefore, the next experiment was designed to further investigate the effects of context semanticity in object discrimination.

Experiment 5

After having found strong effects of orientation, a rather "low-level" stimulus attribute, on contextual object processing in natural scenes, we were interested in the dynamics of more cognitive, *semantic object-context interactions* that are based on the knowledge about the co-occurrence of objects in a meaningful natural scene-context. In addition we wanted to investigate further whether the scene context can facilitate object processing when both are presented simultaneously.

In an influential study Biederman et al., (1982) found faster detection of objects embedded in semantically congruent compared to incongruent scenes. Hollingworth and Henderson (1998), on the other hand, attributed the results of Biederman et al. (1982) to response bias. These authors did not find a congruency advantage when they controlled for bias and concluded that object identification is isolated from semantic information about the scene context. In both studies line drawings of scenes and objects were used. Davenport and Potter (2004), presented photos of natural scenes at one scene-mask SOA (80ms), and found an increased naming error rate when object and context are semantically incongruent, as compared to congruent scenes. However, as Davenport and Potter (2004) found, and we found in our second experiment, objects were reported most accurately when they were presented in isolation on a meaningless uniform background. This could be taken as evidence that any meaningful context, congruent or incongruent, hinders object identification in natural scenes. Alternatively, a relatively simple perceptual advantage in the uniform background condition, such as the increased contrast and visibility of the object contour on a uniform background, could account for this result.

Natural scenes and objects have a meaning that includes knowledge about the likelihood of the occurrence of an object in a scene-context. In this experiment we manipulated the semantic congruency of the meaningful objects and scenes by varying the likelihood of their co-occurrence, one aspect of the semantic relation between them (for a discussion of other aspects see Biederman et al., 1982). This approach is similar to that used in previous studies using line drawings of objects and contexts (Biederman et al., 1982; Hollingworth &

Henderson, 1998), and natural scene photos (Davenport & Potter, 2004). In pictures with semantically congruent object-context relation it was very likely for the participants that the object can be found in the particular scene-context (e.g. deer on grassland, Figure 1G), whereas in pictures with incongruent semantic object-context relation the objects were presented in an atypical surrounding scene-context (e.g. a bookshelf on a beach, Figure 1G). We also attempted to address the question about facilitation and inhibition of object recognition by semantic context congruency. This question requires the introduction of a baseline condition in addition to the semantic congruency manipulations of the meaningful scenes. Because the hypothesis under investigation is that the speed of the recognition of an object is influenced by the participants knowledge about the objects that are likely to occur in a scene-context, one requirement for the baseline context was that it should not activate knowledge about the probability of the occurrence of an object. Although uniform backgrounds fulfill this criterion, we hypothesized that they may not be a good choice. The object recognition performance measured with uniform backgrounds may confound effects of semantic congruency with contour extraction effects when compared to the performance obtained with meaningful contexts. Therefore, we used Fourier phase-randomized scenes as the meaningless baseline condition (Figure 1G, right column). Phase-randomized scenes are meaningless in the sense that the participants have no knowledge about objects that are more likely in to occur in them, and in this limited sense we consider them as semantically neutral. However, phase randomized scenes have some advantages over uniform backgrounds with respect to their low-level perceptual features. They are structured, and have the same orientation and RMS-contrast statistic as natural scenes. A physiological study has found that they produce the same activation levels in early visual areas as intact scenes (Rieger, Schalk, Gegenfurtner, & Heinze, 2004), so that their influence on figure-ground segmentation in early visual areas can be expected to be relatively similar to that of natural scenes.

Methods

Except for the reported changes, methods were similar to those employed in the first experiment (Figure 1A).

Participants.

Fifteen new volunteers (9 female), who were naive with respect to the experiment's purpose, participated. Their mean age was 25.3 years (range 23 to 28 years). Five additional participants judged the semantic consistency/inconsistency of object and background.

Stimuli.

The size of the scene context was 256 by 384 pixels and subtended 9.5° by 14.25° of visual angle. The objects (animals and man-made objects, see previous experiment) were embedded into semantically congruent, incongruent, or meaningless backgrounds. Meaningless backgrounds were generated from the congruent backgrounds by randomizing their Fourier-phase without changing the amplitude spectrum. This manipulation preserves the RMS-contrast of the images (Parseval theorem), which is a widely accepted contrast measure for complex scenes (Bex & Makous, 2002; Reinagel & Zador, 1999). Fourier-phase randomization, when appropriately applied, does not change the BOLD-response strength or response dynamics in early visual areas (Rieger et al., 2004). Matlab software to perform the phase randomization can be downloaded from http://www.uni-magdeburg.de/rieger. Note that this manipulation assumes a linear RGB-to-luminance function of the display system. We therefore measured the functions, calculated a linearization lookup-table and transformed all manipulated images accordingly.

Before we conducted the experiment, five participants judged the congruency and incongruency of object and context in our stimuli in addition to four of the authors who constructed the pictures. A scene was removed from the set when it was judged by less than four participants as originally intended by these authors. Example scenes are shown in Figure 1G.

Insert Figure 6 A/B here

Two hundred pairs of scenes were presented in each of the three conditions (congruent, incongruent, and meaningless contexts). Presentations were randomly distributed over 4 different SOAs (30 ms, 50 ms, 70 ms and 120 ms). Each of the 600 image pairs was presented only once to a participant. Additionally, we included a condition in which only the mask was presented (0 ms SOA). Each SOA was repeated 50 times, except for the mask-only condition, which was repeated 60 times. The resulting 660 trials per participant were presented in 2 blocks.

Results

Image statistics.

Mean object sizes in the two classes (animal/non-animal) in the congruent, incongruent and meaningless pictures did not differ (congruent animal: 8971 pixels, congruent man-made: 9618 pixels, $t_{398}=0.27$, $p>0.05$; incongruent animal: 8393 pixels, incongruent man-made: 8469 pixels, $t_{398}=0.76$, $p>0.05$; meaningless animal: 9105 pixels, meaningless man-made: 8615 pixels, $t_{398}=0.4$, $p>0.05$). There was also no difference between the mean luminances of the images containing animals and man-made objects in the congruent and incongruent image classes (congruent animal: 37.1 cd/m^2, congruent man-made: 37.3 cd/m^2, $t_{398}=0.22$, $p>0.05$; incongruent animal: 37.3 cd/m^2, incongruent man-made: 37.6 cd/m^2, $t_{398}=0.12$, $p>0.05$). In the meaningless-context condition there was a small but significant difference (meaningless-context animal: 36.6 cd/m^2, meaningless-context man-made: 37 cd/m^2, $t_{398}=2.67$, $p<0.05$). Although this difference is statistically significant, it amounts to a Michelson-contrast difference of only 0.5% between the two categories. A difference this small is below the detection threshold (Gegenfurtner & Rieger, 2000) and therefore extremely unlikely to have had any influence on classification performance given the short SOAs we used.

In 94.1% of the images (standard error 2.3%) the semantic object-context congruency or incongruency was judged by all five participants to be as intended by four of the authors when they constructed them.

Psychophysical performance.

The proportions of correct responses at the different SOAs and the fitted psychometric functions are shown in Figure 6A. The exact SOAs and the slopes of the psychometric functions with their 95% confidence intervals are listed in Table 5.

The threshold SOA for semantically congruent scene-contexts (36.5 ms) was very similar to the thresholds for the upright scenes obtained in the previous experiments. This result indicates that our image manipulations did not have additional, unexpected, effects. Semantic incongruency between object and context prolonged threshold SOAs by more than a factor of two in our object classification task (77.6 ms). The threshold SOA obtained with the meaningless, phase-randomized, contexts (40.4 ms) did not differ from that found with meaningful, congruent scenes (MCT, $p>0.05$). At the shortest SOA used, semantically consistent contexts slightly improved classification performance relative to the meaningless contexts (30 ms: $\Delta=10.7\%$, $t_{14}=6.0$, $p<0.001$), but this advantage vanishes at longer SOAs. The slope of the psychometric function was steeper in the meaningless condition (MCT, $p<0.01$).

Insert Table 5 here

In contrast, threshold SOAs were clearly shorter for congruent than for incongruent contexts (MCT, $p<0.01$), but the slopes did not differ significantly (MCT, $p>0.05$). The psychometric functions obtained with semantically incongruent and meaningless contexts differed with respect to both parameters. Meaningless contexts led to shorter threshold SOAs and steeper slopes (MCT both $p<0.01$).

RTs were analyzed with a two factor within subject ANOVA (3 semantic congruency levels x 4 SOAs) and are plotted in Figure 6B. Both SOA and semantic congruency had a significant main effect on the RTs (SOA: $F_{3,42}=20.2$, $p<0.01$, $\eta^2=0.59$, MSE=997; context: $F_{2,28}=21.3$, $p<0.01$, $\eta^2=0.60$, MSE=392), and these factors interacted significantly ($F_{6,84}=3.6$, $p<0.01$, $\eta^2=0.20$, MSE=210). The RTs in Figure 6B suggest that this interaction was caused by increased RTs in the incongruent conditions at short SOAs. The data indicate that the fastest RTs were obtained with a meaningless or a semantically congruent context, whereas the longest RTs were obtained when target and scene-context were incongruent, although with the 120 ms SOA the differences were very small (meaningless baseline: 425 ms; semantically congruent: 431 ms; semantically incongruent: 434 ms). In all conditions RTs were positively correlated with the error rate, and therefore give no indication of a speed-accuracy trade off.

Discussion

A major result of this experiment is that humans require longer undistorted processing (scene-mask SOAs) to discriminate objects when they are embedded in semantically incongruent contexts than when they are embedded in congruent contexts. For RTs this effect is most pronounced at short scene-mask SOAs that impose severe limitations on the information that can be acquired from the scenes. The RT-effect of the semantic object-context incongruence appears to diminish at the longest SOA (120 ms) we used. This reduction is predicted by many choice models that incorporate temporal evidence accumulation (e.g. Hick, 1952; Usher et al., 2002). If one assumes that the effect of semantic context congruency depends on the amount of information that can be extracted from the object, these models predict shorter and more similar RTs at longer scene-mask SOAs. The reduction of the RT effect with incongruent contexts suggests that context semantics may become less important for discrimination when sufficient processing time is available. We also found slightly better object discrimination accuracy with congruent contexts relative to meaningless contexts at the shortest SOA (30 ms). However, the threshold SOAs did not differ between these two conditions. Together these results imply a very rapid accumulation of object specific information from the scenes. The somewhat longer RTs for semantically congruent versus meaningless, semantically neutral, baseline contexts may indicate increased processing demands imposed by additional information from the background.

Phase randomized scenes as a baseline for facilitation/inhibition.

We used phase-randomized natural scenes as a meaningless baseline condition, to evaluate whether semantically congruent contexts may have an inhibitory or facilitatory effect on object discrimination. Phase-randomized scenes have the advantage that they match global RMS-contrast and orientation distribution of natural scenes, both important determinants for coding and activation levels in early visual areas (Rieger et al., 2004). It is unlikely that phase randomized scenes provide semantic information that interferes with object processing, in the sense that the participant can make an informed guess about likely objects in the particular context. Thus, we think that phase-randomized scenes may provide a better baseline to investigate facilitatory and inhibitory semantic context effects than uniform backgrounds do. Other attempts could be made to construct a baseline condition that additionally includes features such as lines and edges. However, great care would be necessary to avoid such stimuli accidentally forming objects, or carrying other semantic information.

General Discussion

Our data clearly show that both low-level (orientation) and high-level (semantic) context attributes affect object processing in natural scenes. The effect of the context appears to be asymmetric: Although the context can delay the classification of the target object, we found only slight evidence for a facilitatory effect when both object and context are presented simultaneously. *Rotating* either object or context lengthened the threshold SOAs and RTs, and increased the trial-by-trial variability of the response (as reflected by the shallower slopes of the psychometric function). Upright contexts did not facilitate the classification of rotated objects. Disoriented objects in meaningless uniform backgrounds were faster recognized than disoriented objects in the scene-contexts. Objects in meaningful, *semantically congruent* scene-contexts were faster discriminated than objects in semantically incongruent contexts. The comparison of performance obtained with a semantically congruent context and a meaningless baseline context, matched for early shape processing demands, revealed that the congruent contexts had only a slight facilitatory effect and only at the shortest SOA used. In contrast to the simultaneous object-context presentations, we found reliable facilitatory effects of the scene context when the onset of the context *preceded* the onset of the object shortly (40 ms) and both were presented at the same orientation.

In the following we will relate our data to models of object processing and discuss their implications.

Orientation compensation and orientation invariant recognition of objects in natural scenes.

A robust finding reported in the literature is that RT increases linearly with rotation angle when the task is to identify disoriented objects (e.g. Jolicoeur, 1985). Many models that deal with the recognition of disoriented objects assume that this effect is due to the need to compensate for the rotation of the object's perceptual representation, in order to produce a representation that matches the representation of the object held in memory (e.g. Jolicoeur, 1985; Shepard & Metzler, 1971; Tarr & Pinker, 1989). The orientation compensation process is presumed to require a fixed amount of time for each degree of rotational compensation, leading to longer RTs at larger rotation angles. We will refer to this hypothesis as the "object-rotation hypothesis". Currently it is unclear, whether a compensation mechanism such as that proposed in these models is applicable to object recognition in natural scenes. Jolicoeur (1990a), in an attempt to explain effects that are not compatible with such a mental-rotation-like process, proposed that parallel to the orientation compensation system, object recognition can be accomplished by a feature recognition system that uses orientation invariant features that are typical enough to allow recognition of an object (e.g. the stripes of a zebra). DeCaro and Reeves (2000, 2002), using line drawings of natural objects, found faster naming latencies for the naming of entry-level object identities than for naming object orientations, and suggested that orientation compensation is not required for object identification. The few recent studies that have investigated object recognition in natural scenes failed to find orientation effects (Guyonneau, Kirchner, & Thorpe, 2006; Rousselet et al., 2003; Vuong et al., 2006), and have concluded that in natural scenes object recognition is practically orientation invariant.

However, our first four experiments show that object classification in natural scenes depends on orientation. Our findings in the first two experiments confirm and extend results from previous studies that found only mild effects of inversion in static displays (Rousselet et al., 2003), or no inversion effects in dynamic displays (Vuong et al., 2006). In both these studies only upright and inverted scenes were used. Because rotation effects are often weaker when objects are inverted (e.g. Jolicoeur, 1985), we speculated that the failure to find orientation effects in these studies might be due to the use of inversion. Therefore, we used several orientations (0°, 90° and 180° in Experiment 1 and in addition 45° and 135° in Experiment 2) and found that while upright and inverted scenes produced the same threshold SOAs rotation by 90° prolonged them. In addition we found shallower slopes of the psychometric functions and longer RTs for all the rotations investigated. We also found that RTs depend strongly on the duration of stimulus availability, a parameter not controlled for in the Rousselet et al. study (2003). However, although we found a clear rotation effect on both RTs and accuracy the speed of the orientation compensation we found in Experiment 2 with objects in context is faster by a factor of approximately three than what would be expected from mental rotation (Dickerson & Humphreys, 1999; Hamm & McMullen, 1998; Jolicoeur, 1990b). Moreover, we found in Experiment 2 that the discrimination of isolated objects presented on uniform backgrounds seems to be rotation invariant.

Task and strategy affect the relative contribution of orientation dependent and invariant processing

At least two not necessarily exclusive suggestions have been made to explain the variability of rotation-speeds. Jolicoeur (1990b) argued that the information used to recognize objects (orientation dependent or orientation independent) may vary from trial to trial, producing variations in the apparent speed of mental rotation. This implies that strategy effects may account for the variation of apparent mental rotation speeds. Depending on the task requirements (speed or accuracy) subjects may favor the use of rapidly accessible orientation independent or slowly accessible orientation dependent (structural) object information. A recently published study (Guyonneau et al., 2006) reported rotation invariance for saccade reaction times when subjects used eye movements to selected target scenes in a speeded choice reaction task. The participants in this study exhibited a speed-accuracy trade off which indicates that some strategy effects were involved. Alternatively, two studies (Dickerson & Humphreys, 1999; Hamm & McMullen, 1998) provide evidence that the strength of the orientation effects varies with the task complexity. Subordinate-level object classification may produce the strongest and superordinate-level classification may produce the weakest rotation effects. Dickerson and Humphreys (1999) suggested that subordinate-level object identification may require object processing at great detail, including orientation dependent structural object properties, to distinguish between highly similar objects. Superordinate-level identification might require only quickly accessible, orientation independent object features to distinguish between highly dissimilar object categories. The results of Grill-Spector and Kanwisher (2005) indicate that similar effects could play a role in our study using classification of objects embedded in natural scenes. Grill-Spector and Kanwisher provide evidence that information sufficient for basic-level or superordinate-level object classification can be extracted as fast as the information required for object detection in natural scenes. Detection and classification success were tightly linked in this study, suggesting that they rely on similar information.

Subordinate-level classification, on the other hand, required much longer exposure durations prior to masking, indicating that the level of task specificity might have a strong influence on the information required to accomplish a task. Based on the stimulus material we used in Experiments 1 and 2 we suppose that participants could have accomplished the object discrimination task on base-level or superordinate-level. The apparently high rotation speed that we obtained in the second experiment might therefore reflect the participants' tendency to discriminate objects on the basis of orientation independent object features. Thus, even stronger orientation effects could be expected when the task required processing on a greater level of detail.

The orientation compensation hypothesis is in concordance with our data from Experiments 1 and 2. However, an alternative explanation exists that does not require an explicit orientation compensation stage to match a stored object representation for recognition (Ashbridge et al., 2000; Perrett et al., 1998). Perrett et al. (1998) suggest that apparent orientation compensation is accomplished by the distribution of the orientation tunings of object selective neurons and evidence accumulation by spike integration over time from the neuronal population. Single cell studies in macaques have shown that several neurons in object selective cortices respond to the same object but with different orientation preferences (e.g. Ashbridge et al., 2000; Logothetis, Pauls, & Poggio, 1995). Rotated objects would produce a weaker population response, if neurons with a preference for upright objects are more abundant in the neuronal population. As a consequence, it would take longer to accumulate enough spikes from the neuronal population to reach a criterion level of "evidence" for an object. This would explain the reduced performance with rotated objects without the involvement of an additional mental rotation mechanism. Perrett et al. (1998) also suggest that object recognition by parts could be accomplished by the neuronal population, since many neurons respond selectively to a part of an object. However, single-cell studies by definition cannot look simultaneously at the combined action of multiple neurons or even brain systems. A decision on the question whether an explicit mental rotation stage plays a role in the discrimination of disoriented natural scenes therefore has to await further physiological studies that involve the investigation of population responses in multiple brain areas. Furthermore, to our knowledge none of these theories make explicit predictions about the effects of a scene-context on the discrimination of disoriented objects.

Scene context orientation: Reference frames vs. object rotation

Another alternative to the object-rotation hypothesis is the assumption that there is a reference frame (coordinate system) that is brought in alignment with the rotated object to allow the processing of the object within this oriented reference frame (e.g. Graf et al., 2005; Hinton, 1981). A test case for the "object-rotation" theories is the case in which both the object and context are rotated independently with congruent and incongruent rotation angles. According to the object-rotation theories the matching of an object to a stored representation should be rapid when the objects are in upright position, independent of the orientation of the background. Theories assuming the alignment of a reference frame would predict that the alignment to the object would be impaired when object and context orientations differ. Previous experiments using letters (Jolicoeur, 1990b) or sequentially presented line drawings of objects (Graf et al., 2005) provide evidence that two or more objects presented in close temporal or spatial proximity can be better recognized if they have the same (congruent) orientation, and are recognized worse when they have different (incongruent) orientations. However, two other studies, using unfamiliar objects, suggest that this congruency effect might be limited to objects from the same object class (Gauthier & Tarr, 1997; Tarr & Gauthier, 1998). It was therefore not clear whether an orientation congruency effect would occur between objects and their scene-context, because these are highly dissimilar. If we had found no congruency effect our results would have favored the object-rotation hypothesis. However, the results of our third experiment do not support this hypothesis. When we compared the effects of rotating the object and context (Experiment 3), we found no primacy of one over the other. Disoriented scene-contexts had a similar effect on discrimination accuracy of upright objects as the rotation of the full scene or the rotation of the object alone. This result is in concordance with the reference frame theory. This theory is also supported by our finding that longest RTs were obtained with scenes where the orientation of the object and context were incongruent (Figure 4B, 60 and 120 ms). The predictions of the reference-frame theory hold not only for simultaneous presentations of scene-context and object, but also when both are presented sequentially (Experiment 4). When we presented the rotated context only 40 ms before the target appeared in it at the same orientation we found a clear accuracy increase. No such increase was found when object and context had incongruent orientation. This supports the view that orientation compensation is achieved by the rotation of a perceptual reference frame instead of individual object axes (Graf et al., 2005; Jolicoeur, 1990b). According to this view, any orientation contrast would produce an orientation ambiguity between object and context that has to be resolved before the object is recognized. The time required for this process may add to the total processing time required, leading, as we found, to longer RTs even at long SOAs when the accuracy is high.

Semantic congruency

Some models predict facilitation of object identification by semantically congruent backgrounds (Bar, 2004; Biederman et al., 1982; Ullman, 1996). One might expect any such advantage to be biggest with longer scene-mask SOAs, which permit more semantic information to become available. In contrast, we found RTs in our congruent and incongruent conditions to become increasingly similar with longer SOAs. It was only under the most restrictive conditions we employed (a 30ms scene-mask SOA) that we found better classification performance for semantically congruent than meaningless scene-contexts. The information available from the scene within such short intervals is likely to be relatively coarse and unspecific, but may suffice occasionally to activate learned object context-relations that bias the participant’s decisions (Bar, 2004). Possible costs of this discrimination advantage may be indicated by the somewhat longer RTs for congruent contexts. However, previously described naming accuracy advantages (Davenport & Potter, 2004), and the discrimination accuracy advantages we found for objects in uniform contexts relative to objects in structured, semantically congruent contexts (Experiment 2), seem likely to be due to an advantage at relatively early, sensory processing levels where figure ground segmentation occurs. This process might be activated in early visual areas (von der Heydt, 1994) before semantic knowledge is accessed. In accordance with this view, our data indicate that phase-randomized contexts can minimize this advantage.

Properties of information extracted from the scenes in short masked presentations

Results from physiological studies using single-cell recordings in macaques, (Kovacs, Vogels, & Orban, 1995; Rolls & Tovee, 1994), human functional Magnetic-Resonance-Imaging (Grill-Spector, Kushnir, Hendler, & Malach, 2000), and magnetoencephalography (Rieger et al., 2005) indicate that the pattern-mask limits the information extracted from a scene by interrupting the processing in shape and object selective visual areas. Therefore, analyzing the dynamics of the interaction between objects and their scene-context may help to reveal the properties of the information extracted from the scenes during the short scene-mask SOAs we used.

Insert Figure 7 here

Clues regarding the properties of the information extracted at different scene-mask SOAs are provided by the RTs obtained in Experiments 3 and 5 which included conditions with object-context conflicts. In Figure 7 we plot the RT differences obtained with the incongruent object-context orientation conditions and the congruent upright condition (T90 C0-T0 C0 and T0 C90-T0 C0) in the third experiment, and the RT differences between the semantically incongruent and congruent condition in the fifth experiment. These RT differences indicate that at least 50 ms undistorted information accumulation from the scenes was required to acquire enough information to produce a significant RT effect of the incongruent scene context. This duration was similar whether orientation or semantic congruency was manipulated (Experiments 3 and 5, respectively), suggesting that the stimulus representations must build to some level of complexity for the conflicting information from object and context to interact. We deem it likely that these object-context interactions occur at the processing levels at which structural information about orientation (DeCaro & Reeves, 2000, 2002; Dickerson & Humphreys, 1999; Hamm & McMullen, 1998) or information about scene semantics (Ganis & Kutas, 2003; Hollingworth & Henderson, 1998) first becomes available. It is highly probable that both the target-scene and the distractor-scene in our experiments are processed in parallel to the level of information complexity required for object discrimination. This is indicated by a study of Rousselet, Fabre-Thorpe, and Thorpe (2002) who asked participants to detect an animal in a scene. RT, d-prime, and the simultaneously recorded EEG were independent of whether one or two simultaneously presented scenes could contain the target object. As others (e.g. Li, Van Rullen, Koch, & Perona, 2002), Rousselet et al. (2002) concluded, that object categorizations can be achieved in parallel in natural scenes, without the need for sequential focal attention.

Furthermore, we found significant rotation effects on accuracy and RTs as long as a scene context was available. Isolated objects, however, were better discriminated than scene-embedded objects, and with isolated objects we found orientation independence, at least in the accuracy data. Because the major difference between the two presentation conditions is the accessibility of the shape, we speculate that object shape may have provided additional orientation independent information to improve superordinate object classification. Overall, the data suggest that while scene-contexts could provide information about an objects superordinate classification, any benefit derived from this information is offset by the costs imposed by the need to segregate the object from the clutter that the scene places at its borders.

Interpretation of the behavioral measures

To our knowledge only few studies have reported both accuracy and RT data over several SOA levels (but see also Bacon-Mace et al., 2005; Grill-Spector & Kanwisher, 2005). There is reason to think that these different measures can provide complementary information. The results of physiological studies on scene or object processing imply that in pattern-masking paradigms SOAs limit accuracy by limiting the amount of information that can be accumulated from the scenes (Grill-Spector et al., 2000; Kovacs et al., 1995; Rieger et al., 2005; Rolls & Tovee, 1994). Shallower slopes of the psychometric functions, on the other hand, indicate higher trial-to-trial variability of the responses (DeCaro & Reeves, 2002; Gescheider, 1997; Green & Swets, 1989; Strasburger, 2001). Green and Swets (1989) point out that RTs can provide information complementary to accuracy measures, as they are related to the complexity, or the noisiness, of processing from the sensory to the decision stage. An alternative, although not exclusive, interpretation of the slopes obtained with our masked presentations comes from mathematics. If we assume that the proportion of correct responses reflects the amount of information accumulated from the scenes in a given SOA, then the slope reflects the maximum information accumulation rate as it is determined at the point where the psychometric function has maximum steepness (c.f. DeCaro & Reeves, 2002).

In our data, both the slopes of the psychometric functions and the RTs are equal for the scenes rotated by 90° and 180° but differ from the upright condition. We speculate that the longer RTs and the shallower slopes we observed for rotated scenes indicate increased noise due to reduced efficiency of the coding of the scenes (Perrett et al., 1998), or in matching the processed input to a stored representation (Jolicoeur, 1990b; Tarr & Pinker, 1989; Ullman, 1996). Perrett et al. (1998) suggested that fewer neurons are activated when the object orientation deviates from the commonly experienced view, leading to slower accumulation of information about the object. Alternatively, additional orientation compensation required to solve the discrimination task (Graf et al., 2005; Jolicoeur, 1990a) may add noise to the discrimination process (e.g. Green and Swets, 1988 p. 324ff; Gescheider, 1997, p. 105ff). The RT effects presented in Figure 7 and the RT effects we found in Experiments 1 and 2 (Figure 2B and Figure 3B and D) could be seen as support for the notion that orientation compensation requires additional time for the processing of the object. The RTs were longer for rotated than for upright scenes even when the accuracy levels were high at long SOAs. This interpretation is further supported by rotation effect studies in which accuracy is high and unchanged but rotation effects exist in the RTs (e.g. McMullen et al., 1995; Murray et al., 1993), and by physiological studies that show that the processing of rotated objects additionally activates cortical areas involved in spatial processing (e.g. Vanrie, Beatse, Wagemans, Sunaert, & Van Hecke, 2002).

In summary we have shown that object discrimination in naturalistic photographs depends on orientation and contextual information. Both, cognitive (semantic) and perceptual (orientation) attributes play a role and affect the processing speed. When object and context appear simultaneously the effects of the context on object discrimination tended to be inhibitory. However, congruently oriented contexts that precede the object in time can improve object recognition. Our results suggest that the human visual system is well adapted to cope with the complexity of the highly cluttered natural environment. However, with the paradigm we used, we found only little indication that object-context is used profitably during rapid visual natural scene processing.

References

Ashbridge, E., Perrett, D. I., Oram, M. W., & Jellema, T. (2000). Effect of image orientation and size on object recognition: Responses of single units in the macaque monkey temporal cortex. *Cognitive Neuropsychology, 17*(1-3), 13-34.

Bacon-Mace, N., Mace, M. J., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course of visual processing: backward masking and natural scene categorisation. *Vision Research, 45*, 1459–1469.

Bar, M. (2004). Visual objects in context. *Nature Review Neuroscience, 5*(8), 617-629.

Basri, R. (1993). Viewer-centered representations in object recognition: A computational approach. In C. H. Chen, L. F. Pau & P. S. P. Wang (Eds.), *Handbook of Pattern Recognition and Computer Vision*. Singapore: World Scientific Publishing Company.

Bex, P. J., & Makous, W. (2002). Spatial frequency, phase, and the contrast of natural images. *Journal of the Optical Society of America A: Optics, Image Science & Vision, 19*(6), 1096-1106.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*, 115-147.

Biederman, I. (1988). Aspects and extensions of a theory of human image understanding. In Z. Pylyshyn (Ed.), *Computational processes in human vision: An interdisciplinary perspective* (pp. 370-428). Norwood, NJ: Ablex.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: detecting and judging objects undergoing relational violations. *Cognitive Psychology, 14*(2), 143-177.

Boyce, S. J., Pollatsek, A., & Rayner, K. (1989). Effect of background information on object identification. *Journal of Experimental Psychology: Human Perception and Performance, 15*(3), 556-566.

Breitmeyer, B. G. (1984). *Visual masking: An integrative approach*. Oxford: Oxford University Press.

Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception and Performance, 21*(1), 109-127.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, D. J., & Kubovy, M. (1993). Mental rotation, mental representation, and flat slopes. *Cognitive Psychology, 25*, 251–382.

Coltheart, V. (1999). *Fleeting Memories: Cognition of brief visual stimuli*. Cambridge Mass: MIT Press.

Corballis, M. C. (1988). Recognition of disoriented shapes. *Psychol Rev, 95*(1), 115-123.

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science, 15*(8), 559-564.

DeCaro, S. A., & Reeves, A. (2000). Rotating objects to determine orientation, not identity: evidence from a backward-masking/dual-task procedure. *Percept Psychophys, 62*(7), 1356-1366.

DeCaro, S. A., & Reeves, A. (2002). The use of word-picture verification to study entry-level object recognition: further support for view-invariant mechanisms. *Memory & Cognition, 30*(5), 811-821.

Delorme, A., Richard, G., & Fabre-Thorpe, M. (2000). Ultra-rapid categorisation of natural scenes does not rely on colour cues: a study in monkeys and humans. *Vision Research, 40*(16), 2187-2200.

Dickerson, J., & Humphreys, G. W. (1999). On the identification of misoriented objects: Effects of task and level of stimulus description. *The European Journal of Cognitive Psychology, 11*(2), 145-166.

Ganis, G., & Kutas, M. (2003). An electrophysiological study of scene effects on object identification. *Cognitive Brain Research, 16*(2), 123-144.

Gauthier, I., & Tarr, M. J. (1997). Orientation priming of novel shapes in the context of viewpoint-dependent recognition. *Perception, 26*(1), 51-73.

Gegenfurtner, K. R., & Rieger, J. (2000). Sensory and cognitive contributions of color to the recognition of natural scenes. *Current Biology, 10*(13), 805-808.

Geisler, W. S., Perry, J. S., Super, B. J., & Gallogly, D. P. (2001). Edge cooccurrence in natural images predicts contour grouping performance. *Vision Research, 41*(6), 711-724.

Gescheider, G. A. (1997). *Psychophysics the Fundamentals* (3rd Edition ed.). Mahwah, NJ: Lawrence Ehrlbaum Associates.

Graf, M., Kaping, D., & Bülthoff, H. H. (2005). Orientation congruency effects for familiar objects: coordinate transformations in object recognition. *Psychological Science, 16*(3), 214-221.

Green, D. M., & Swets, J. A. (1989). *Signal detection theory and psychophysics*. New York: Wiley.

Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition: as soon as you know it is there, you know what it is. *Psychological Science, 16*(2), 152-160.

Grill-Spector, K., Kushnir, T., Hendler, T., & Malach, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. *Nature Neuroscience, 3*(8), 837-843.

Guyonneau, R., Kirchner, H., & Thorpe, S. J. (2006). Animals roll around the clock: The rotation invariance of ultrarapid visual processing. *Journal of Vision, 6*(10), 1008-1017.

Hamm, J. P., & McMullen, P. A. (1998). Effects of orientation on the identification of rotated objects depend on the level of identity. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(2), 413-426.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology, 50*, 243-271.

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology, 4,* 11-26.

Hinton, G. E. (1981). A parallel computation that assigns canonical object-based frames of reference. In *In Proceedings of the Seventh International Joint Conference on Artificial Intelligence* (Vol. 2, pp. 683–685). Los Altos, CA: Kaufman.

Hollingworth, A., & Henderson, J. M. (1998). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology: General, 127*(4), 398-415.

Humphreys, G. W., Riddoch, M. J., & Quinlan, P. T. (1988). Cascade processes in picture identification. *Cognitive Neuropsychology, 5*(1), 67-103.

Jolicoeur, P. (1985). The time to name disoriented natural objects. *Memory Cognition, 13*(4), 289-303.

Jolicoeur, P. (1990a). Identification of disoriented objects: A dual-systems theory. *Mind and Language, 5,* 387-410.

Jolicoeur, P. (1990b). Orientation congruency effects on the identification of disoriented shapes. *Journal of Experimental Psychology: Human Perception and Performance, 16*(2), 351-364.

Kanwisher, N., Yin, C., & Wojciulik, N. (1999). Repetition blindness for pictures: Evidence for the rapid computation of abstract visual descriptions. In V. Coltheart (Ed.), *Fleeting Memories: Cognition of brief visual stimuli* (pp. 119-150). Cambridge Mass: MIT Press.

Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate*. Cambridge, MA: MIT Press.

Kovacs, G., Vogels, R., & Orban, G. A. (1995). Cortical correlate of pattern backward masking. *Proceedings of the National Academy of Sciences of the United States of America, 92*(12), 5587-5591.

Lawson, R., & Jolicoeur, P. (1998). The effects of plane rotation on the recognition of brief masked pictures of familiar objects. *Memory & Cognition, 26*(4), 791-803.

Lawson, R., & Jolicoeur, P. (2003). Recognition thresholds for plane-rotated pictures of familiar objects. *Acta Psychologica, 112*(1), 17-41.

Li, F. F., Van Rullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences of the United States of America, 99*, 9596–9601.

Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology, 5*(5), 552-563.

McMullen, P. A., Hamm, J., & Jolicoeur, P. (1995). Rotated object identification with and without orientation cues. *Canadian Journal of Experimental Psychology-Revue Canadienne De Psychologie Experimentale, 49*(2), 133-149.

Murray, J. E. (1997). Flipping and spinning: Spatial transformation procedures in the identification of rotated natural objects. *Memory & Cognition, 25*(1), 96-105.

Murray, J. E., Jolicoeur, P., McMullen, P. A., & Ingleton, M. (1993). Orientation-invariant transfer of training in the identification of rotated natural objects. *Memory & Cognition, 21*(5), 604-610.

Perrett, D. I., Oram, M. W., & Ashbridge, E. (1998). Evidence accumulation in cell populations responsive to faces: an account of generalisation of recognition without mental transformations. *Cognition, 67*(1-2), 111-145.

Potter, M. C. (1975). Meaning in visual search. *Science, 187*(4180), 965-966.

Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network-computation in neural systems, 10*(4), 341-350.

Rieger, J. W., Braun, C., Bülthoff, H. H., & Gegenfurtner, K. R. (2005). The dynamics of visual pattern masking in natural scene processing: a magnetoencephalography study. *Journal of Vision, 5*(3), 275-286.

Rieger, J. W., Schalk, F., Gegenfurtner, K. R., & Heinze, H. J. (2004). BOLD response in V1 to phase-noise degraded photographs of natural scenes. *Perception Supplement, 33,* 115.

Rolls, E. T., Aggelopoulos, N. C., & Zheng, F. S. (2003). The receptive fields of inferior temporal cortex neurons in natural scenes. *Journal of Neuroscience, 23*(1), 339-348.

Rolls, E. T., & Tovee, M. J. (1994). Processing speed in the cerebral-cortex and the neurophysiology of visual masking. *Proceedings of the Royal Society of London Series B: Biological Sciences, 257*(1348), 9-15.

Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nature Neuroscience, 5*, 629-630.

Rousselet, G. A., Mace, M. J., & Fabre-Thorpe, M. (2003). Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *Journal of Vision, 3*(6), 440-455.

Sanocki, T. (1999). Constructing structural descriptions. *Visual Cognition, 6,* 299-318.

Schyns, P. G., & Oliva, A. (1994). From Blobs to boundary edges: Evidence for time- and space-dependent scene recogntion. *Psychological Science, 5*(4).

Seiffert, A. E., & Di Lollo, V. (1997). Low-level masking in the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance, 23*(4), 1061-1073.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171,* 701-703.

Strasburger, H. (2001). Converting between measures of slope of the psychometric function. *Perception & Psychophysics, 63*(8), 1348-1355.

Tarr, M. J., & Bülthoff, H. H. (1998). Image-based object recognition in man, monkey and machine. *Cognition, 67*(1-2), 1-20.

Tarr, M. J., & Gauthier, I. (1998). Do viewpoint-dependent mechanisms generalize across members of a class? *Cognition, Special issue on "Image-based Recognition in Man, Monkey, and Machine", 67,* 71-108.

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology, 21*(2), 233-282.

Ullman, S. (1996). *High-level vision: Object recognition and visual cognition*. Cambridge, MA: MIT Press.

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience, 5*(7), 682-687.

Usher, M., Olami, Z., & McClelland, J. L. (2002). Hick's law in a stochastic race model with speed-accuracy tradeoff. *Journal of Mathematical Psychology, 46*, 704-715.

Vanrie, J., Beatse, E., Wagemans, J., Sunaert, S., & Van Hecke, P. (2002). Mental rotation versus invariant features in object perception from different viewpoints: an fMRI study. *Neuropsychologia, 40*(7), 917-930.

von der Heydt, R. (1994). Form analysis in visual cortex. In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences* (pp. 365-382). Cambridge, Mass.: MIT Press.

Vuong, Q. C., Hof, A. F., Bülthoff, H. H., & Thornton, I. M. (2006). An advantage for detecting dynamic targets in natural scenes. *Journal of Vision, 6*(1), 87-96.

Wichmann, F. A., & Hill, N. J. (2001a). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics, 63*(8), 1293-1313.

Wichmann, F. A., & Hill, N. J. (2001b). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics, 63*(8), 1314-1329.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*, 841-145.

Author Note

We want to thank Robert Fendrich for helpful comments on the manuscript. Supplementary material is available online from the Journal Web page.

JWR and MG were supported by Land Sachsen-Anhalt Grant FKZ0017IF0000 to the Magdeburg Leibniz program and FS, NK, and JWR by grants from the Deutsche Forschungsgemeinschaft (RI 1511/1-1 and RI 1511/1-3).

Correspondence concerning this article should be addressed to Jochem Rieger, Department of Neurology II, Otto-von-Guericke University, 39120 Magdeburg, Leipzigerstr. 44, Germany; email: jochem.rieger@nat.uni-magdeburg.de

Table 1

Threshold SOAs and Slopes of the Psychometric Functions for Experiment 1:

Rotations	Threshold (in ms)	Upper/lower 95% confidence limit (in ms)	Slope (prop. correct/ms)	Upper/lower 95% confidence limit (prop. correct/ms)
0 deg	31.9	34.6/28.9	0.013	0.017/0.011
90 deg	54.85	59.7/50.1	0.006	0.007/0.004
180 deg	34.57	39.01/28.73	0.007	0.009/0.006

Table 2

Threshold SOAs and Slopes of the Psychometric Functions for Experiment 2 (Objects in Scenes and in Isolation):

Rotations: Objects in scenes	Threshold (in ms)	Upper/lower 95% confidence limit (in ms)	Slope (prop. correct/ms)	Upper/lower 95% confidence limit (prop. correct/ms)
0 deg	35.61	39.03/31.81	0.0102	0.0138/0.0086
45 deg	37.32	41.67/33.02	0.0083	0.0104/0.0063
90 deg	46.33	51.93/40.63	0.0069	0.0094/0.0052
135 deg	41.62	46.12/36.32	0.0076	0.0101/0.0059
180 deg	41.07	45.17/36.67	0.0089	0.0114/0.0071

Rotations: Isolated objects	Threshold (in ms)	Upper/lower 95% confidence limit (in ms)	Slope (prop. correct/ms)	Upper/lower 95% confidence limit (prop. correct/ms)
0 deg	32.57	35.42/29.64	0.0167	0.0235/0.0119
45 deg	35.19	38.68/31.50	0.0102	0.0130/0.0083
90 deg	34.23	37.89/29.94	0.0096	0.0124/0.0077
135 deg	34.96	39.50/30.52	0.0081	0.0101/0.0058
180 deg	32.77	36.24/28.94	0.0106	0.0151/0.0086

Table 3

Threshold SOAs and Slopes of the Psychometric Functions for Experiment 3:

Rotations	Threshold (in ms)	Upper/lower 95% confidence limit (in ms)	Slope (prop. correct/ms)	Upper/lower 95% confidence limit (prop. correct/ms)
T0 C0	36.26	40.7/31.5	0.0084	0.0108/0.0063
T90 C0	56.85	66.7/46.6	0.0039	0.0052/0.003
T0 C90	60.98	71.1/51.6	0.004	0.0052/0.003
T90 C90	63.73	72.8/53.2	0.0045	0.006/0.0035

Table 4

Proportions of Correct Responses and Mean RT for Experiment 4:

Rotations:	SOA scene-object (in ms)	Mean corr. resp. (in %)	SE corr. resp. (in %)	Mean RT (in ms)	SE RT (in ms)
T0 C0	0	76.00	2.21	641.28	31.04
	40	83.77	2.81	594.39	38.40
	80	82.44	2.67	588.62	37.77
	160	76.00	3.43	625.89	39.70
T90 C0	0	66.00	2.89	671.00	39.18
	40	68.44	3.31	648.28	50.19
	80	70.00	2.76	654.91	50.88
	160	64.44	2.84	669.70	48.26
T0 C90	0	70.66	3.36	658.74	33.92
	40	73.55	2.69	628.77	41.20
	80	74.88	3.08	618.67	49.96
	160	70.66	3.27	637.58	44.37
T90 C90	0	65.77	2.48	662.89	36.41
	40	73.33	2.70	631.84	38.46
	80	73.33	3.59	620.34	41.47
	160	71.77	2.62	623.40	39.45

Table 5

Threshold SOAs and Slopes of the Psychometric Functions for Experiment 5:

Background	Threshold (in ms)	Upper/lower 95% confidence limit (in ms)	Slope (prop. correct/ms)	Upper/lower 95% confidence limit (prop.correct/ms)
Congruent	36.5	40.0/32.2	0.007	0.0084/0.005
Incongruent	77.1	83.03/70.7	0.0074	0.0142/0.006
Neutral	40.4	42.9/37.2	0.011	0.0134/0.0094

Figure Captions

Figure 1. A) The experimental paradigm used in Experiments 1, 2, 3, and 5. Each trial started with a fixation cross that was presented for variable durations (300-850 ms). Then two scenes were briefly presented left and right from the fixation cross. One of them contained the target object, an animal. A pattern mask followed both scenes at one out of several stimulus-onset-asynchronies (SOA). Finally, a green cross indicated the responses window and the next trial was automatically initiated. We analyzed the reaction time and the accuracy of the response. In the first experiment (B) the scenes were presented upright, 90°, and 180° rotated. In the second experiment (C) we compared discrimination performance for isolated (bottom row) and scene-embedded objects (top row) and extended the number of orientations (upright, 45°, 90°, 135°, 180°). In this experiment animal and non-animal pictures were presented as pairs of isolated objects, or as pairs of objects embedded in scenes. Here we show only examples of the animal-objects. In the third experiment (D) we rotated scene and context independently, but only by 90°. In the fourth experiment (E and F) the onset of the scene-context preceded the onset of the object by one out of several scene-object SOAs (E). The object-mask SOA was fixed at 40ms. As in the third experiment, object and background were rotated independently by 90° (F). The scene contexts were chosen from two classes (nature and man-made scenes). Animal (top-row) and non-animal objects (bottom-row) with the same orientation and with contexts from the same semantic category were presented as pairs. Each column represents an example of a pair of the four orientation combinations that were used. Animal and non-animal objects were selected to be semantically congruent with the context in order to minimize orientation independent information about the target picture from context. In the fifth experiment (G) the contexts were semantically congruent (left) or incongruent (center) with the objects, or were meaningless (right). Meaningless, structured backgrounds were obtained by phase randomizing the congruent scene-contexts.

Figure 2. A) The proportion of correct responses obtained with different rotations and SOAs in Experiment 1. The threshold SOAs do not differ between upright and inverted scenes, but are lengthened by the 90° rotation. B) The reaction times (RT) obtained for correct responses at the different SOAs. The RTs for upright scenes fall faster than the RTs for the rotated scenes. This fall-off does not reflect a speed-accuracy trade off.

Figure 3. A) The proportion of correct responses obtained with objects in a scene-context in Experiment 2. The threshold SOAs are longer for the 90° than for the upright condition. Thresholds for the other conditions (45°, 135°, and 180°) do not differ from the upright condition. B) The RTs obtained for correct responses with objects in a scene-context at the different SOAs. C) The proportion of correct responses obtained with isolated objects. The threshold SOAs do not differ between orientations. D) The RTs obtained for correct responses with isolated objects at the different SOAs.

The RTs for upright scenes fall off faster than the RTs for the rotated scenes (B and D). This fall off does not reflect a speed-accuracy trade off. At intermediate SOAs RTs are longer for objects in scenes than for isolated objects. Note that the longest RTs have been found with 135° rotation angles at the longest SOA when accuracy is high.

Figure 4. A) The proportion of correct responses obtained at different SOAs with different object-context rotations in Experiment 3. Rotation of the whole scene or any part of it lengthens the threshold SOA by a similar extent. B) The reaction times (RT) obtained for correct responses at each SOA. The RT for upright scenes (T0 C0) falls as SOA increases. In contrast, the RTs increase over the first three SOAs when only the background is rotated (T0 C90).

Figure 5. A) The proportion of correct responses obtained at different *context-object* SOAs with different object-context rotations in Experiment 4. The object-mask SOA was fixed at 40

ms (see Figure 1E). The proportion of correct responses was significantly increased for congruently oriented object-context orientation combinations (T0 C0 and T90 C90), when the context preceded the object by 40 ms. However, for incongruent object-context orientation combinations (T90 C0 and T0 C90) we found only a non-significant tendency. The increase in discrimination performance is transient and is less with 160 ms than with 40 ms context-scene SOA. B) The RTs obtained for correct responses at each context-object SOA. The fastest RTs were obtained with all-upright scene-object combinations (T0 C0) and the slowest with rotated objects (T90 C0). The RTs approximately exhibit an inverse relationship to the proportion of correct responses.

Figure 6. A) The proportion of correct responses obtained at different SOAs with semantically congruent, incongruent and meaningless, semantically neutral baseline contexts in Experiment 5. RMS-contrast structures in the meaningless baseline contexts and the congruent contexts were matched. The threshold SOA did not differ between the meaningless and congruent context conditions, but was longer in the incongruent condition. B) RTs for correct responses in the three conditions as a function of SOA. RTs tend to drop with longer SOAs, but the largest differences between the congruent and incongruent RTs occur at the intermediate SOAs (50 ms and 70 ms). The fastest RTs were found with meaningless contexts.

Figure 7. Mean RT differences between the incongruent object-context orientation conditions and the congruent upright condition (T90 C0-T0 C0 and T0 C90-T0 C0) obtained in Experiment 3, and the RT differences between the semantically incongruent and congruent condition obtained in Experiment 5. The error bars represent the two sided 95% confidence intervals. These RT differences indicate that at least 50 ms undistorted information accumulation from the scenes was required to gather enough information to produce a significant RT effect of the incongruent scene context.

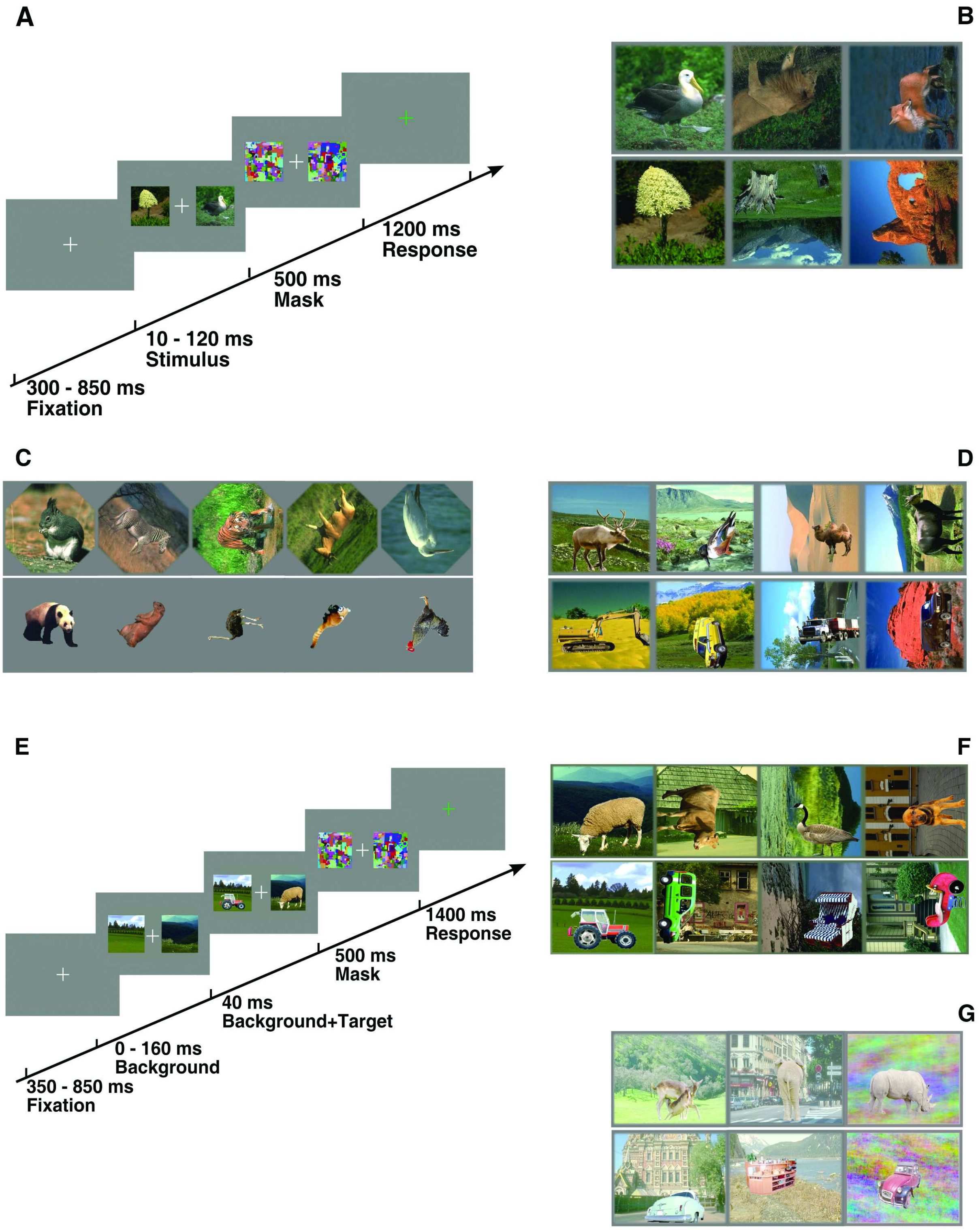
A
B
1200 ms
Response
500 ms
Mask
10 - 120 ms
Stimulus
300 - 850 ms
Fixation
C
D
E
F
1400 ms
Response
500 ms
Mask
40 ms
Background+Target
0 - 160 ms
Background
350 - 850 ms
Fixation
G

Figure 1.

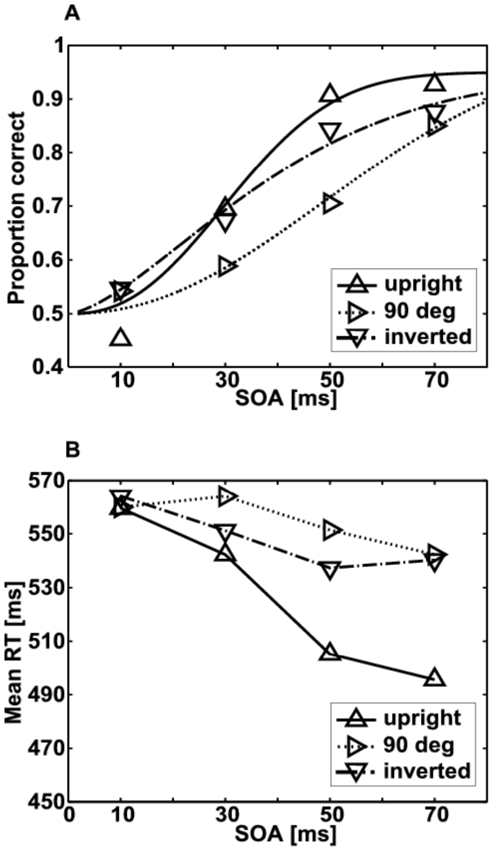

Figure 2

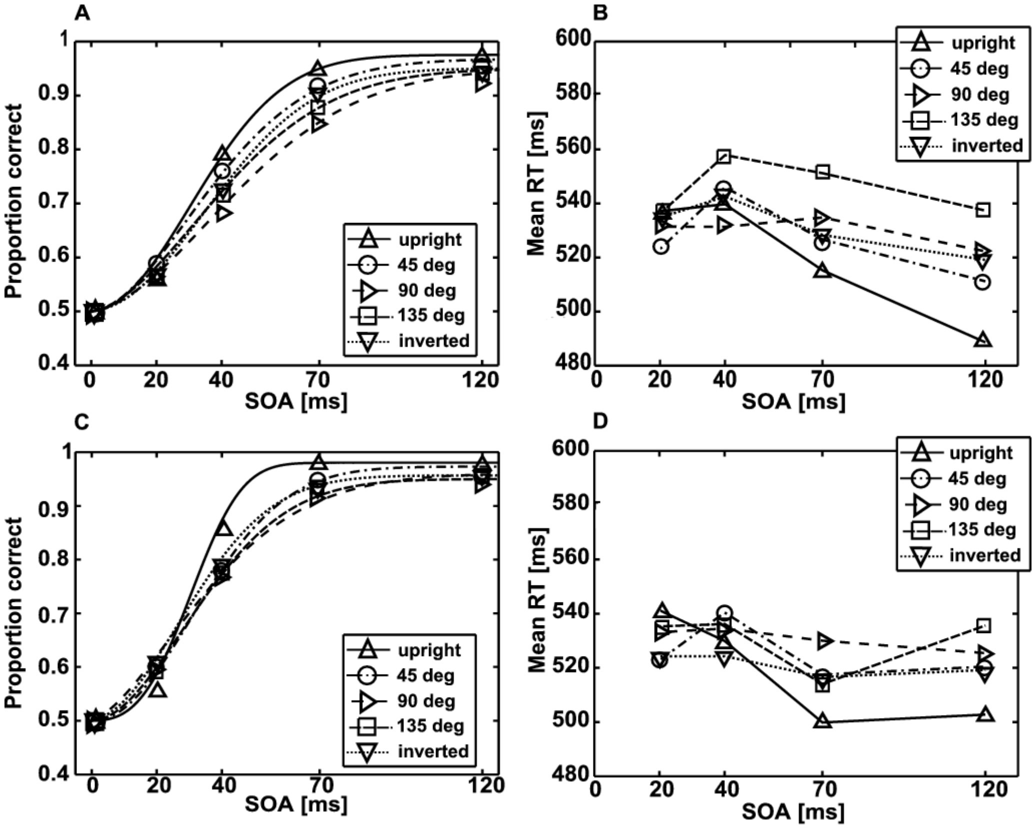

Figure 3

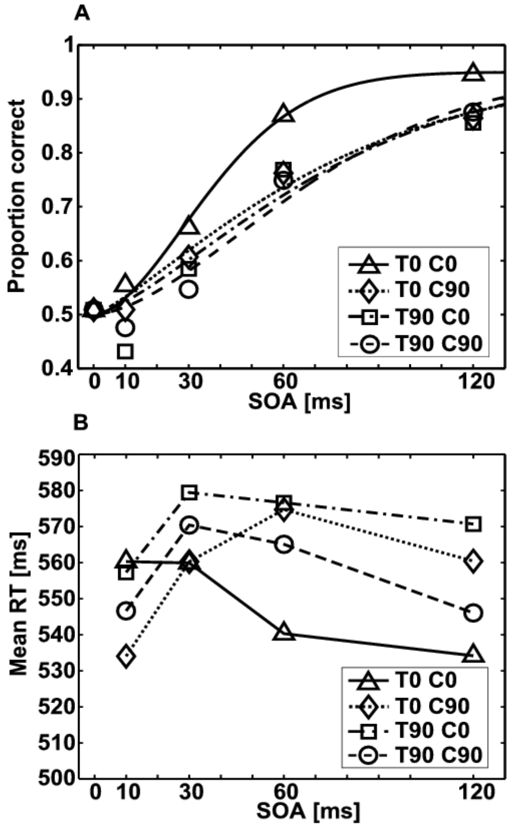

Figure 4

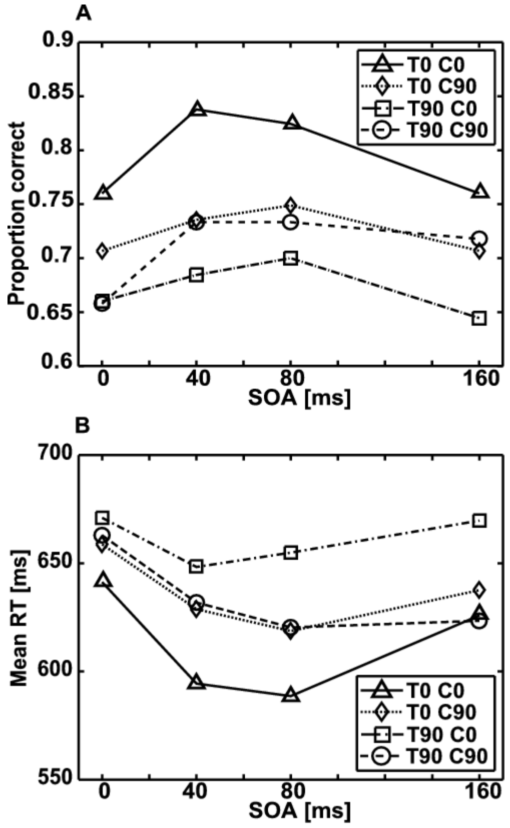

Figure 5

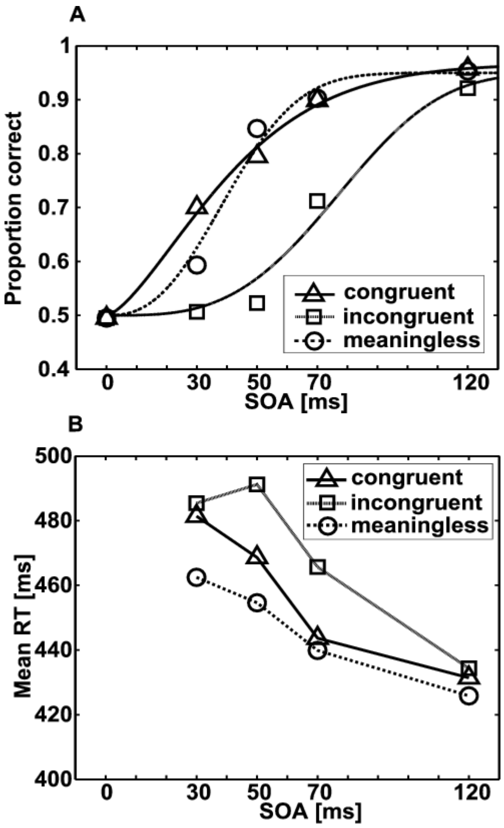

Figure 6

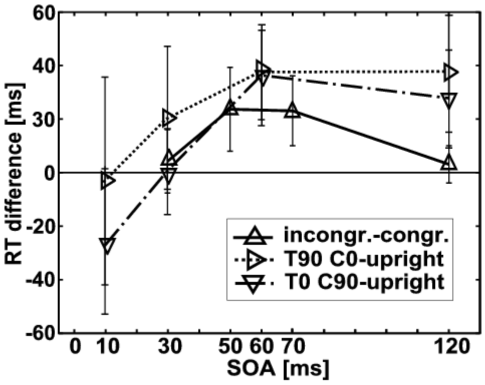

Figure 7