

Contextual integration of visual objects necessitates attention

Nurit Gronau · Meytal Shachar

Published online: 22 January 2014
© Psychonomic Society, Inc. 2014

Abstract Objects that form a contextually coherent percept are grasped more rapidly and efficiently than objects that are contextually inconsistent with each other. The extent to which such clustering processes depend on visual attention is largely unknown. The present research examined the necessity of attention for object-to-object contextual integration processes during a brief visual glimpse. Participants performed an object classification task on associated object pairs that were presented for a short duration (59 ms). Objects were positioned either in expected relative locations (e.g., a desk lamp on a desk) or in unexpected relative locations (e.g., a desk lamp under a desk). When both stimuli were relevant to task requirements, latencies to spatially consistent object pairs were significantly shorter than those to spatially inconsistent pairs (Experiment 1). These contextual effects disappeared, however, when spatial attention was drawn to one of the two object stimuli while its counterpart object appeared outside the main focus of attention, serving as a task-irrelevant distractor (Experiment 2). Attentional modulation of contextual integration processes was shown to be independent of distractor recognition per se (Experiment 3). Finally, the role of goal-directed (endogenous) and spatial (exogenous) attention factors in contextual integration was explored (Experiment 4). Taken together, our findings suggest that contextual associations play an important role in processing multiple-object visual displays. However, regardless of whether objects are associated by active or passive relations, the construction of a coherent contextual representation strongly relies on the availability of attentional capacity. Possible implications for theories of scene and object recognition are discussed.

Keywords Visual attention · Context · Object recognition · Contextual/spatial relations · Action relations

Introduction

Recognition of everyday objects feels almost effortless. This may be surprising given the fact that objects typically appear within highly rich, cluttered scenes, where they often compete for limited processing resources. One factor that may reduce scene complexity and streamline visual recognition is the contextual setting within which objects tend to appear. A desk lamp, for instance, normally rests on a desk, by a telephone or a computer screen, within a broader office setting. Recognizing the desk lamp may, therefore, activate, or prime, a set of contextually associated objects, thereby reducing the cognitive resources required for their processing. Indeed, much research has shown that contextual information, such as a visual scene or an individual object within a scene, can facilitate recognition of an associated object (Antes, Penland, & Metzger, 1981; Bar & Ulfman, 1996; Biederman, 1972; Davenport & Potter, 2004; Friedman, 1979; Munneke, Brentari, & Peelen, 2013; Palmer, 1975; for reviews, see Bar, 2004; Oliva & Torralba, 2007). Associative relations among objects can be grasped within a short visual glimpse (e.g., Auckland, Cave, & Donnelly, 2007; Biederman, Mezzanotte, & Rabinowitz, 1982; Davenport, 2007; Roberts & Humphreys, 2011), suggesting that contextual information is extracted rapidly and efficiently (Oppermann, Hassler, Jescheniak, & Gruber, 2012).

The present research examined the role of attention in the perception of object-to-object relations during a brief glance. Given the rapid nature of associative processing, we sought to explore whether attention is necessary for clustering several objects presented simultaneously in order to form a contextually coherent scene representation. Specifically, we asked

N. Gronau (✉) · M. Shachar
Department of Psychology & Cognitive Science Studies,
The Open University of Israel, Raanana, Israel
e-mail: nuritgro@openu.ac.il

whether a meaningful, integrated visual percept can arise during a rapid visual glimpse, when important contextual information appears outside the main focus of visual attention. Much research has documented the importance of contextual knowledge in guiding visual attention toward a target object. Visual search for a target is typically faster when the target is positioned in an expected (i.e., contextually consistent) location rather than in an unexpected (i.e., contextually inconsistent) location within a scene (e.g., Chun & Jiang, 1998; Droll & Eckstein, 2008; Neider & Zelinsky, 2006; Torralba, Oliva, Castelano, & Henderson, 2006; for a review, see Wolfe, Vö, Evans, & Greene, 2011). While contextual information is clearly utilized in guiding visual search for a prespecified unattended target, the present research assessed object-to-object contextual processing when an unattended stimulus is not necessarily part of the task set (i.e., when it is a task-irrelevant distractor) and when short exposure durations limit serial scanning of the display. We posited that if visual integration processes depend on the availability of attentional resources, only a small set of objects appearing inside the main focus of attention may be bound to form a meaningful percept. If, however, an attended object can rapidly prime or preactivate other contextually associated objects appearing at unattended visual locations, partial recognition of the latter may take place. In this case, a fairly reliable contextual percept may be formed within a short visual glance, despite the underprivileged status of the unattended objects.

Traditional attention models have posited that high-level object processing requires focal attention and that the successful binding of several visual elements into a conjoined, conscious percept relies on attention allocation (e.g., Kahneman, Treisman, & Gibbs, 1992; Treisman & Gelade, 1980; Wolfe & Cave, 1999). In the absence of focal attention, only coarse information about an object can be recovered, such as its basic visual features and, to some extent, its general category (e.g., whether the object is animate or inanimate; e.g., Evans & Treisman, 2005). In addition, under unattended conditions, salient changes to an object may be overlooked, and the construction of a global scene representation may be incomplete (e.g., Mack & Rock, 1998; Rensink, O'Regan, & Clark, 1997; Simons & Levin, 1997). Several studies have shown that, in contrast to common intuition, even highly important stimuli, such as a person's own written name or an image of one's own face, may go unnoticed unless these important stimuli are relevant to online task requirements (e.g., Breska, Israel, Maoz, Cohen, & Ben-Shakhar, 2011; Devue & Brédart, 2008; Devue, Van der Stigchel, Brédart & Theeuwes, 2009; Gronau, Cohen & Ben-Shakhar, 2003, 2009; Harris & Pashler, 2004). Furthermore, classic attention models have stressed the role of attention in encoding an object's location in space (e.g., Treisman, 1996). Accordingly, attentional allocation may be important for assessing the relative spatial and/or functional relations between a set of contextually associated

objects within a scene. A desk lamp, for instance, typically rests on a desk, not under it. According to conventional attention models, perceiving the lamp's precise location along with its particular spatial relations with other objects in the scene depends on the availability of attentional resources. Although limited in its capacity, visual attention may enable several objects to be processed simultaneously and to be linked together within a meaningful global percept (see, e.g., Treisman, 2006). Accordingly, the likelihood of detecting and/or recognizing an object that is conceptually (or spatially) inconsistent with its contextual surrounding is dramatically decreased if the object is positioned outside the main focus of visual attention.

In contrast to classic attention models, accumulating evidence in the last decade has suggested that during a brief visual glance, real-world objects (as opposed to meaningless, arbitrary stimuli) form a unique class of stimuli that can be detected and recognized in the near absence of attention. In a series of studies using a dual task paradigm, participants performed a demanding central task while, at the same time, successfully detecting a predefined object category (e.g., an animal or a vehicle) appearing at an unattended (i.e., peripheral) location (Fei-Fei, VanRullen, Koch, & Perona, 2005; Li, VanRullen, Koch, & Perona, 2002). Follow-up studies have revealed that tasks requiring much finer object discriminations (e.g., Is a peripheral object a car or a noncar vehicle?, or Is a peripheral object the face of Tom Cruise or not?) can be additionally performed in the absence of attention capacity (Poncet, Reddy, & Fabre-Thorpe, 2012; Reddy, Reddy, & Koch, 2006; Reddy, Wilken, & Koch, 2004; see also VanRullen, Reddy, & Fei-Fei, 2005). On the basis of these findings, the authors concluded that high-level object representations are formed in parallel and are accessed even when stimuli appear outside the main focus of attention (but see M. A. Cohen, Alvarez, & Nakayama, 2011; Evans & Treisman, 2005; Walker, Stafford, & Davis, 2008). It appears, then, that contrary to traditional views, real-world objects may be categorized and identified even in the absence of focal attention. Despite the evidence above, however, it yet remains to be determined whether object-to-object associative relations can be readily grasped under such conditions.

Scene-object contextual integration

Research in the field of attention has provided important input concerning the necessity of attention to visual recognition and to visual integration processes. A somewhat different perspective on these issues, however, arises from studies directly associated with the influence of visual context on object processing. Common approaches to the facilitative effects of contextual information on visual processing typically posit that objects may be preactivated (or primed) by their context

to a degree that is sufficient for their partial recognition and for their successful integration with the surrounding environment (e.g., Antes et al., 1981; Bar, 2004; Biederman et al., 1982; Boyce, Pollatsek, & Rayner, 1989; Friedman, 1979; Mudrik, Lamy, & Deouell, 2010; Palmer, 1975; but see Ganis & Kutas, 2003; Hollingworth & Henderson, 1998). Contextual integration, or the ability to perceive a coherent visual setting, according to such views, may take place even if an object appears in a peripheral visual location, outside the main focus of visual attention. Support for the relative effortless nature of object recognition in the presence of contextual information comes mainly from studies that have examined the effects of scene recognition (e.g., a kitchen) on the processing of contextually consistent (e.g., a kettle) and inconsistent (e.g., a fire hydrant) objects embedded in the scene. The rationale behind such scene–object consistency paradigms is that a scene’s gist, or its broad semantic category, can be recognized with minimal effort (e.g., Munneke et al., 2013; Oliva & Schyns, 2000; Oliva & Torralba, 2001; Potter, 1975; Rousselet, Joubert, & Fabre-Thorpe, 2005) that allows rapid preactivation of object identities within the scene¹ (e.g., Bar & Ullman, 1996; Biederman et al., 1982; Boyce et al., 1989; Friedman, 1979; Palmer, 1975; but see Hollingworth & Henderson, 1998). Several researchers have proposed that contextual oddities can be computed preattentively, such that an object that is inconsistent with its surrounding environment rapidly pops out and captures attention (e.g., Loftus & Mackworth, 1978; Underwood & Foulsham, 2006; Underwood, Templeman, Lamming, & Foulsham, 2008; see also Mudrik, Breska, Lamy, & Deouell, 2011). Gordon (2004), for instance, demonstrated that within a short visual glance, attention is preferentially allocated to contextually inconsistent objects, rather than to consistent objects, within a scene (see related findings with change detection in Hollingworth & Henderson, 2000). Additional studies using eye movement measures have shown that scene–object inconsistencies attract eye fixations when a contextually inconsistent object appears at an extrafoveal location (Becker, Pashler, & Lubin, 2007; Bonitz & Gordon, 2008; Loftus & Mackworth, 1978; Underwood & Foulsham, 2006; Underwood et al., 2008). Given the tight correlation between foveal processing and visual attention in everyday vision (see, e.g., Belopolsky & Theeuwes, 2009; Hoffman & Subramaniam, 1995; Theeuwes, Kramer, Hahn, & Irwin, 1998), these findings suggest that scene–object contextual relations can be at least partially recovered when stimuli are presented outside the main focus of visual attention. Other

studies, however, have yielded contradictory results, obtaining no evidence for rapid extrafoveal detection of contextually inconsistent objects within scenes (De Graef, Christiaens, & d’Ydewelle, 1990; Gareze & Findlay, 2007; Henderson, Weeks, & Hollingworth, 1999; Rayner, Castelhana & Yang, 2009; Vö & Henderson, 2009, 2011). According to these studies, scene–object consistency effects may stem from processes occurring subsequent to, rather than prior to, the allocation of visual attention and the eyes to an object in the scene. The extent to which contextual relations in scenes, therefore, are computed in the absence of attention is still under ongoing debate.

Object-to-object contextual integration

While scenes form a special case in which objects may be preactivated due to rapid global processing of a scene’s gist (e.g., Oliva & Schyns, 2000; Oliva & Torralba, 2001; Potter, 1975; Rousselet et al., 2005), a different approach to the study of the effects of context on visual recognition is to examine the influence of an individual object (rather than a whole scene) on recognition of a contextually associated object. Indeed, several studies have shown that pairs of contextually related items are recognized more efficiently than pairs of unrelated items, regardless of a background scene (Auckland et al., 2007; Davenport, 2007; Henderson, Pollatsek, & Rayner, 1987; Oppermann et al., 2012). The spatial relations between the objects—that is, whether they are positioned in plausible or implausible locations with respect to each other—further affect accuracy and speed of recognition performance (e.g., Bar & Ullman, 1996; Green & Hummel, 2006; for related fMRI findings, see also Gronau, Neta, & Bar, 2008; Kim & Biederman, 2011; Roberts & Humphreys, 2010). These studies have highlighted the importance of object-to-object relations as an important source of contextual facilitation, independent of scene knowledge (Henderson et al., 1987; Hollingworth, 2007; but see Boyce et al., 1989). Notably, in most of the aforementioned studies, stimuli were presented in the main focus of visual attention, or at the least, there was no attempt to control for attentional factors. Therefore, the role of attention in gluing several objects within a global, contextually coherent percept remains largely unknown.

Perhaps the most compelling evidence in favor of an unattended processing of object-to-object relations comes from studies conducted with patients suffering from attentional deficits following parietal lobe damage (Riddoch, Humphreys, Edwards, Baker, & Willson, 2003; Riddoch et al., 2006; Riddoch et al., 2011). A typical symptom exhibited by such patients is extinction, an impaired ability to respond to a contralesional stimulus under conditions of competition from a simultaneous stimulus on the ipsilesional side. In their study, Riddoch et al. (2003) presented patients

¹ Note that according to some researchers, scenes and objects are processed in parallel. Consequently, not only can a scene rapidly activate an object, but also an object can activate a scene. Thus, scene–object facilitative effects may be bidirectional (see, e.g., Davenport and Potter 2004; Fei-Fei, Iyer, Koch, & Perona, 2007; Joubert, Fize, Rousselet, & Fabre-Thorpe, 2008; Underwood et al., 2008).

suffering from extinction with two contextually associated objects, one on each side of the screen. The pairs of objects actively interacted with each other either in a correct manner (e.g., a corkscrew directed toward the cork of a wine bottle) or in an incorrect manner (e.g., a corkscrew directed toward the base of a wine bottle). The researchers' main finding was that a contralesional (unattended) object was more likely to be detected and identified if it correctly interacted with an ipsilesional (attended) object. Namely, extinction was reduced when the two objects were positioned appropriately, relative to when they were positioned inappropriately, for their combined use. According to the authors, visual attention was tuned to the interactive relations between the objects, so that attention was spread across both members of a pair of objects being used together, allowing their explicit identification. These results strongly suggest that objects can be preattentively grouped on the basis of their interactive relations (see also Green & Hummel, 2006; Riddoch et al., 2006, 2011; Roberts & Humphreys, 2011). Specifically, object-to-object (active) relations can be perceived even when one of the two objects is perceptually and/or attentionally underprivileged.

The present study

The goal of the present study is to further investigate whether object pairs can be processed and successfully linked within a coherent visual percept when one of the two objects is positioned outside the main focus of attention. In contrast to most previous studies, we carefully manipulated visual attention while neurologically intact participants performed an unbiased object classification task on the object pairs. Each stimulus pair contained two contextually associated objects that were presented simultaneously for a brief duration. In order to assess processing of contextual associations, we manipulated the spatial relations within each pair, such that objects interacted either in a contextually correct (i.e., plausible/expected) or in a contextually incorrect (i.e., implausible/unexpected) manner. As was mentioned earlier, the spatial relations between objects were found in previous studies to play an important role in the speed and accuracy, as well as in the neural activation, associated with visual object recognition (e.g., Bar & Ullman, 1996; Green & Hummel, 2006; Gronau et al., 2008; Kim & Biederman, 2011; Riddoch et al., 2003, 2011; Roberts & Humphreys, 2010, 2011). In the present study, the spatial consistency effect (i.e., the difference in reaction time [RT] for the object classification task in the spatially correct vs. the incorrect object pairs) was assessed under conditions in which stimuli in a pair were both attended (Experiment 1) or when only one of the two stimuli was attended while its pair

stimulus appeared outside the main focus of visual attention (Experiment 2). We hypothesized that when both stimuli are attended, participants will grasp the two objects more easily if they are co-located in contextually correct (plausible) locations than if they are co-located in contextually incorrect (implausible) locations. Namely, spatially consistent object pairs will elicit shorter RTs than will spatially inconsistent object pairs (e.g., Gronau et al., 2008). Our main interest concerned participants' responses when one of the two objects was attended while its counterpart object was unattended or, at most, minimally attended. To the extent that the unattended object was associatively linked to the attended object, a spatial consistency effect should emerge. If, however, perception of contextual associative relations is attention dependent, no spatial consistency effect should be seen. In this case, a clear dissociation should be observed between the two attentional conditions. Since previous research conducted with parietal-injured patients has emphasized the importance of active relations in preattentive object grouping, a secondary goal of the study was to examine possible differences between active and passive contextual associations under the various attentional conditions.

As was mentioned above, Experiment 1 served as a baseline study, in which both objects were presented inside the main focus of attention, while in Experiment 2, we manipulated visual attention in an attempt to assess contextual processing when one of the two objects appeared outside the attentional focus. As a preview of following experiments, Experiment 3 was similar to Experiment 2, with the additional demonstration that contextual integration processes are somewhat independent of unattended stimulus processing at the individual object level. Finally, Experiment 4 further investigated the role of exogenous and endogenous attention factors in object-to-object associative binding.

Experiment 1: Both objects are attended

The goal of Experiment 1 was to determine whether object pairs that are positioned in contextually correct (i.e., spatially consistent) locations are recognized more efficiently than pairs that are positioned in contextually incorrect (i.e., spatially inconsistent) locations during a brief visual glance. Importantly, both objects within each stimulus pair were fully attended, since they were located inside the main focus of visual attention and were both relevant for task demands (see below). We hypothesized that under such conditions, objects positioned in spatially consistent co-locations would be more easily linked within a global contextual percept than objects appearing in spatially inconsistent co-locations. Therefore, the former would elicit shorter RTs than the latter.

Method

Participants

Twenty undergraduate students (16 females, 4 males) with normal or corrected-to-normal sight participated in the experiment for course credit.

Apparatus

Stimulus presentation and response collection were controlled by a PC computer, using an E-Prime 2.0 software (Schneider, Eschman, & Zuccolotto, 2002). Stimuli were presented on a 15-in. monitor with an 85-Hz refresh rate.

Experimental design

One hundred forty-four pairs of everyday objects were used. Each pair consisted of two colored objects (including, e.g., furniture, tools, household utensils, appliances, foods, clothes) that were contextually or functionally associated with each other. Sixty-four pairs involved action relations (e.g., a kettle oriented toward a mug in a pouring position, a pizza cutter oriented toward a pizza in an cutting position), while 80 object pairs involved relative passive relations (e.g., a desk lamp resting on a desk, a sandwich lying on a plate, a mirror resting on a dresser) (see details concerning the division of active and passive pair categories in the Stimuli section below). Stimuli were presented simultaneously above and below fixation: On half of the trials, the objects were positioned in spatially correct locations (spatially consistent condition), while on the other half of the trials, the objects appeared in spatially incorrect locations (spatially inconsistent condition) (see Fig. 1a and b, respectively). For each participant, object pairs were randomly assigned into one of the two spatial-consistency conditions, such that each pair appeared only once and in only one of the two conditions during the course of the experiment.

In addition to the 144 contextually associated object pairs, 96 pairs of stimuli containing a nonsense visual object were used as target stimuli in an object classification task. The nonsense objects consisted of artificial, colorful shapes that were manually designed with Adobe Photoshop software. On each trial, participants' task was to determine rapidly whether the stimulus display contained a nonsense target stimulus or not by using a two-alternative forced choice keypress. The goal of the object classification task was to allow an indirect, and unbiased, measure of participants' responses to the contextually associated stimulus pairs (note that a negative response was required for both spatially consistent and inconsistent pairs, ruling out possible response bias accounts for any contextual effects). Namely, the task was orthogonal to the main question of interest—the differences in responses

between the two spatial consistency conditions. Importantly, within the trials containing a nonsense target stimulus, the target could appear in an upper or a lower location with equal probability; thus, participants had to be attentive to both stimulus locations.

Pairs containing a nonsense target object consisted either of two nonsense stimuli (both-target condition, 32 trials) or of a nonsense stimulus appearing in an upper/lower location, paired with a real, nontarget object that matched in its category with the objects forming the contextually associated pairs (e.g., furniture, tools, appliances, food, etc.; mixed-target condition, 64 trials) (see Fig. 2). Note that the two types of target trials (i.e., both target, mixed target) were particularly important in Experiments 2 and 3, in which they served as an index for participants' sensitivity to the difference between unattended target and nontarget distractor stimuli (see detailed explanation in Experiment 2).

Finally, 48 additional trials containing a single object were used. These consisted of 24 nonsense target objects and 24 real, nontarget objects, appearing at either an upper or a lower location. As was the case with the target object pairs, the single-object trials were particularly important in Experiments 2 and 3 (for assessing cue validity effects; see below) and were inserted in order to allow similar presentation conditions across experiments. Altogether, there were 288 trials in a block (i.e., 144 contextually associated pairs, 96 pairs containing a nonsense target object, and 48 single objects—either nonsense targets or real nontarget objects).

Each block was repeated 3 times in the course of the experiment. The order of the trials within each block was determined randomly. Prior to the beginning of the experiment, there was a practice session of 120 trials. Object images in the practice session were excluded from the experiment.

Stimuli

Object stimuli were presented on a gray background square subtending about 22×22 cm (corresponding to a visual angle of approximately $20^\circ \times 20^\circ$ from a viewing distance of 60 cm). In trials containing pair objects, stimuli appeared above and below fixation, centered 3.25 cm (approximately 3°) from fixation (the center-to-center distance between objects was approximately 6°). Objects' size varied, ranging from 3.0 to 6.5 cm (visual angle of approximately 2.8° – 6°), in both height and length dimensions. Importantly, among the contextually associated object pairs, object images were collected individually, rather than taken from scenes comprising both objects. The images were gathered from various sources, including commercially available CDs and the Internet. Contextually associated object pairs were carefully constructed from the individual objects, while avoiding physical contact between the stimuli. In cases in which an object would normally rest on its counterpart pair object (e.g., a desk lamp

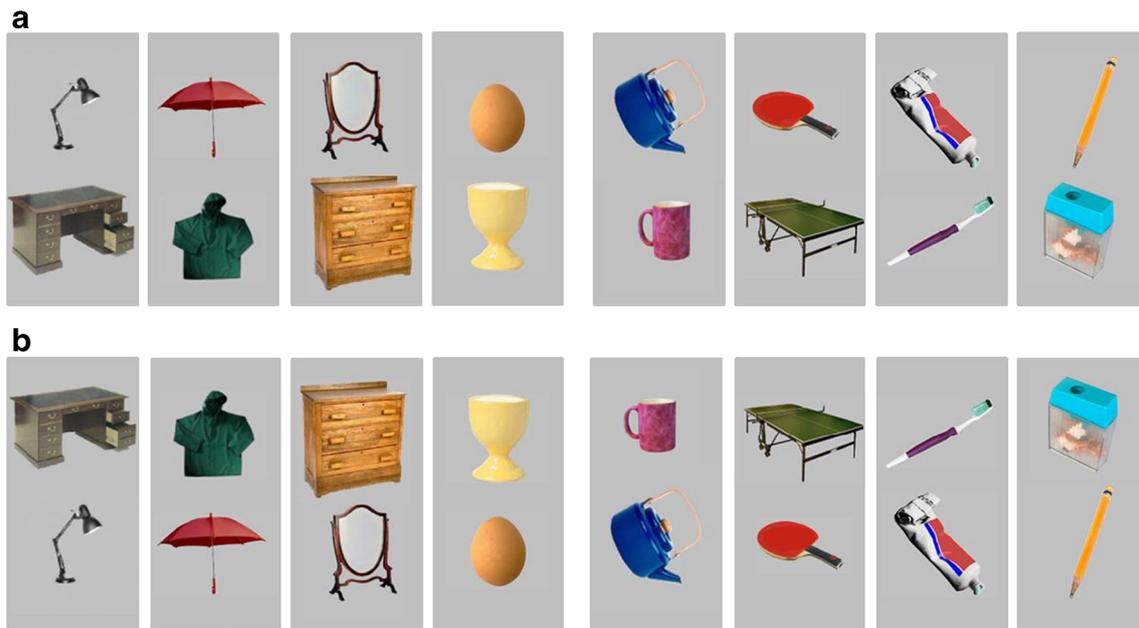


Fig. 1 Examples of the two pair conditions in Experiment 1. **a** Spatially consistent condition. **b** Spatially inconsistent condition. Pairs that depicted action relations (i.e., active pairs) are presented on the right,

and pairs that did not depict action relations (i.e., passive pairs) are presented on the left. All images were presented in color

normally rests on a desk), the upper stimulus often seemed to be slightly floating above the lower stimulus. Although somewhat unnatural, the existence of a spatial gap between the two objects ensured that stimuli were visually dissociated from each other and could not be grouped on the basis of perceptual factors, such as physical support or continuity. Thus, we ensured that any spatial consistency effect obtained among the contextually associated pairs would solely rely on semantic/conceptual factors.

In order to determine the nature of object-to-object relations among the stimulus pairs (i.e., whether a pair involved action relations or not), we ran an independent activity survey on all pair stimuli. Twenty-seven participants were asked to determine whether each object pair, presented at spatially correct locations, depicted action relations or not. Pair stimuli were presented for 80 ms (masked), with an interstimulus interval of 3 s. For each pair, a mean activity score was computed, based on participants' responses (1= contains

action relations, 0= does not contain action relations). Of the 144 pair stimuli, pairs obtaining an average score larger than .5 were defined as active ($n = 64$; mean activity score = .79, $SD = .1$), while other pairs were defined as passive ($n = 80$; mean activity score = .22, $SD = .13$). Note that the difference in the mean activity score for the two action categories was highly significant, $t(26) = 17.7, p < .001$.

Procedure

Each trial in Experiment 1 lasted 2 s (see Fig. 3). The trial began with a fixation cross appearing at the screen center for 494 ms, followed by a blank interval of 59 ms. Subsequently, a central cue (identical to the fixation cross) appeared for 59 ms, followed by a short blank interval (47 ms) and by the object stimuli presented for an additional 59 ms. The insertion of the central cue immediately before the appearance of the object stimuli ensured that participants fixated on the screen



Fig. 2 Examples of the nonsense object targets in Experiment 1: Both-target condition (left) and mixed-target condition (right). In the mixed-target condition, the target stimulus appeared in an upper or a lower location with equal probability

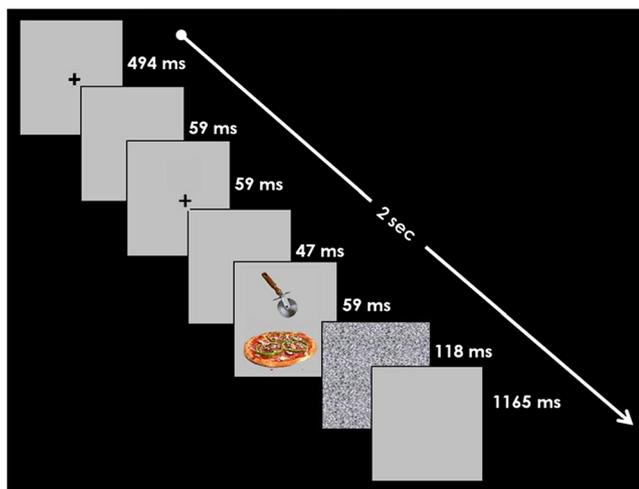


Fig. 3 An example of a trial sequence in Experiment 1

center. In addition, it created sequence presentation conditions similar to these of Experiment 2, in which a peripheral spatial-cueing manipulation was used (see below).

Stimuli were masked by a black and white random noise pattern for 118 ms, followed by a 1,165-ms blank lasting until the end of the trial. Participants were instructed to maintain fixation during the course of all trials. Their task was to press a key (J) marked 1, using their index finger, if there was a nonsense target stimulus in the display and to press a key (K) marked 2, using their middle finger, if there was no nonsense target stimulus in the display. As was noted earlier, the object classification task required participants to be attentive to both stimulus locations, since the nonsense target could appear in either an upper or a lower location (or in both).

Results and discussion

Mean RTs for each condition were computed separately within each experimental block. Trials on which participants made errors (8%) or trials yielding extreme outlier responses (i.e., RTs that deviated from the participant's mean RT by more than three standard deviations; less than 1%) were excluded from the RT analysis. In this experiment, as well as in the following experiments, there was a significant effect of experimental block due to an overall improvement (i.e., decrease) in participants' RTs with accumulating experience and stimulus exposure. However, since there was no interaction effect between the block factor and the spatial consistency factor in any of the experiments, and to simplify the result presentation, we computed mean RTs across the three blocks. We, therefore, report only the mean RTs collapsed across the blocks (see Table 1). We focus on the spatial consistency effect among the contextually related object pairs.

As was expected, spatially consistent items elicited statistically significant shorter RTs than did spatially inconsistent items, $t(19) = 3.26, p < .005$, effect size (ES) = 0.73² (all t -tests reported are two-tailed). These results confirm our hypothesis that objects that form a contextually coherent percept are grasped more efficiently (i.e., faster) than objects that do not form a contextually coherent percept. There was no difference in the error rate between the two spatial consistency conditions, $t(19) = 1.07, p > .1$, ES = 0.24. In order to assess potential latency differences between pairs that involved active versus passive associative relations, we conducted a two-way repeated measures analysis of variance (ANOVA) of spatial consistency (consistent, inconsistent) by pair activity (active, passive) factors. This analysis revealed a significant spatial consistency effect, $F(1, 19) = 11.54, p < .005$, yet there were no significant effects of pair activity, $F(1, 19) < 1, p > .1$, or of the interaction between the two factors, $F(1, 19) = 1.2, p > .1$ (see Table 2). These results suggest that active and passive contextual relations were equally effective in facilitation of response latencies, when both objects were presented inside the main focus of attention.

Overall, our RT results are in accordance with previous findings demonstrating enhanced recognition of objects appearing in contextually correct, relative to incorrect, spatial positions (e.g., Bar & Ullman, 1996; Green & Hummel, 2006; Gronau et al., 2008). Recall that the spatial consistency factor within the contextually associated pairs was irrelevant to task requirements, suggesting that processing of object-to-object contextual relations was rather automatic.

We now turn to our main question of interest—to investigating visual contextual processing when one of the two objects is located outside the main focus of attention.

Experiment 2: One of the two objects is unattended

Experiment 1 has established that object pairs that are positioned in spatially consistent (expected) co-locations are grasped more readily than objects appearing in spatially inconsistent (unexpected) co-locations. Importantly, stimuli in Experiment 1 were fully attended and were both relevant to task requirements. In Experiment 2, we sought to investigate whether a spatial consistency effect may be obtained when one of the two objects is located outside the main focus of visual attention and when it serves as a task-irrelevant distractor. Using a visual display similar to that of Experiment 1, we now added a spatial-cueing manipulation that summoned participants' visual attention, prior to stimuli

² The effect size (ES) was defined as the standardized mean difference between the two conditions. We use a standard norm of ES equaling 0.2, 0.5, and 0.8 for small, medium, and large effects, respectively (see J. E. Cohen, 1988).

Table 1 Mean reaction times (in milliseconds) and proportions of errors in the object classification task in Experiment 1 (standard errors in parentheses)

	Contextually Related Pairs		Target Pairs		Single Objects
	Spatially Consistent	Spatially Inconsistent	Both Target	Mixed Target	
Reaction time	523 (13)	533 (12)	506 (11)	556 (12)	520 (11)
Proportion of errors	.06 (.01)	.07 (.01)	.05 (.01)	.16 (.02)	.06 (.01)

Note. Spatially consistent = pairs containing objects positioned in contextually correct (i.e., plausible) locations; spatially inconsistent = pairs containing objects positioned in contextually incorrect (i.e., implausible) locations; both target = pairs containing two nonsense target stimuli; mixed target = pairs containing a nonsense target stimulus and a nontarget (real) object; single objects = single stimuli containing either a target or a nontarget object.

appearance, to one of the two object stimuli in a pair. In addition, task instructions were changed such that the object classification task was now performed on the cued object only, instead of on both objects within a pair. The cued object, therefore, was task relevant, while the uncued (i.e., unattended) object became a task-irrelevant distractor. We asked whether a spatial consistency effect would be obtained under such conditions.

To the extent that a spatial consistency effect was observed, we could conclude that object-to-object contextual relations were processed despite the underprivileged attention conditions. Obtaining no spatial consistency effect, in contrast, would strongly suggest that linking several objects within a contextually coherent percept necessitates attentional capacity.

Method

Participants

Twenty undergraduate students (18 females, 2 males) participated in the experiment for course credit. All participants had normal or corrected-to-normal sight.

Experimental design and stimuli

The experimental design and the stimuli were identical to those in Experiment 1, except for expanding the number of single-object trials from 48 to 72 (36 target object trials and 36 nontarget object trials in each block). This was done in order to allow a more reliable assessment of the spatial-cuing manipulation (see below).

Procedure

The procedure was identical to that of Experiment 1, except for the following changes. First, a spatial-cuing manipulation was added, by replacing the central cue that preceded objects' appearance in Experiment 1 with a peripheral cue. The peripheral cue was visually identical to the original central cue (i.e., a fixation cross), but instead of appearing at the screen center, it appeared at an upper or a lower screen location (3° above or below fixation). Within each experimental condition, there were equal proportions of upper and lower cues. In all other aspects, the trial sequence was identical to the trial sequence in Experiment 1.

Second, task instructions were changed. While in Experiment 1 participants were required to determine, on each trial, whether there was a nonsense target object in the display, in the present experiment they were asked to determine whether there was a nonsense target object in the cued location only. The cued object, then, became a task-relevant stimulus on which participants performed the object classification task, while the uncued stimulus effectively became a task-irrelevant distractor. Note that within the mixed-target condition, the target (i.e., nonsense object) always appeared in the cued location, whereas the nontarget (i.e., real object) always appeared in the uncued location. In order to simplify the experimental design, we excluded the reversed condition, in which a nontarget appeared in a cued location and a target appeared in an uncued location. Table 3 presents the total number of targets and nontargets presented in the cued and the uncued locations in Experiment 2.

The single object trials served in Experiment 2 for validating the spatial-cuing manipulation—that is, ensuring that

Table 2 Mean reaction times (in milliseconds) and proportions of errors for contextually associated active and passive object pairs in Experiment 1 (standard errors in parentheses)

	Active Pairs		Passive Pairs	
	Spatially Consistent	Spatially Inconsistent	Spatially Consistent	Spatially Inconsistent
Reaction time	523 (12)	531 (12)	523 (13)	536 (12)
Proportion of errors	.06 (.01)	.06 (.01)	.07 (.01)	.08 (.01)

Table 3 Amount of targets and nontargets appearing in a cued and an uncued location (across pair-object and single object trials) in each block of Experiment 2

	Target	Nontarget
Cued location	114	162
Uncued location	50	226

participants responded to the spatial cue and shifted their visual attention to the cued location. On half of the trials, a single object (either a target or a nontarget item) appeared at a cued location (i.e., valid condition), while on the other half, it appeared at an uncued location (i.e., invalid condition). Note that participants were instructed to respond only to an object appearing at the cued location: They were asked to respond 1 if the cued stimulus was a nonsense target object and to press 2 if it was not a nonsense target object. The single object trials, however, formed a special case in which the cued location contained no stimulus on half of the trials (i.e., in the Invalid cue condition). The participants were therefore additionally informed that in a small proportion of the trials, the cued location would contain no object. In such cases, the participants should ignore the cue and respond to the stimulus appearing in the uncued location. The insertion of the single-trial condition, despite its oddity, allowed us to assess the difference in RTs between the valid and the invalid cue trials (while maintaining identical responses for the two conditions). We hypothesized that shorter RTs would be obtained for the valid than for the invalid cue trials, indicating that visual attention was indeed drawn to the cued location. Note that the overall cue validity was very high, since the invalid single-object trials formed only 11.5% of all trials in the experiment.

Each experimental block was repeated 3 times in course of the experiment. Within each block, trial order was determined randomly. As in Experiment 1, we report only the mean RTs collapsed across the blocks.

Results and discussion

Trials on which participants made errors (9%) or trials on which there were extreme outlier responses (less than 1%), were excluded from the RT analysis. Table 4 presents the mean RTs and the proportions of errors for the different conditions in Experiment 2. Overall, performance level was high and similar to that in Experiment 1, suggesting that participants obeyed task instructions and responded to the cued location. In addition, to ensure that our spatial-cuing manipulation was effective, we examined the differences among the single items between the valid and the invalid cue conditions. As can be seen in Table 3, trials in the valid

cue condition yielded significantly shorter RTs than trials in the invalid cue condition, $t(19) = 3.33$, $p < .005$, $ES = 0.75$. The difference in the proportion of errors between the two conditions was statistically significant as well, $t(19) = 4.41$, $p < .001$, $ES = 0.99$. These findings confirm that the spatial-cuing manipulation was indeed effective in summoning attention to the cued location. Next, we focused on our main question of interest—that is, on the spatial consistency effect among the contextually associated pairs, when using the cuing manipulation. As can be seen in Table 4, in contrast to the results of Experiment 1, when one of the two objects was attended while its counterpart object served as an irrelevant distractor, there was no difference in RT between the two spatial consistency conditions [a nonsignificant effect was obtained in the opposite direction; $t(19) < 1$, $p > .1$, $ES = 0.18$]. Namely, the spatial consistency effect was eliminated under the cuing paradigm. The difference in the proportion of errors between the two spatial conditions was numerically small, yet reached statistical significance, $t(19) = 3.41$, $p < .005$, $ES = 0.76$. It appears, then, that at least with respect to the response latency measure, narrowing the focus of visual attention to one of the two stimuli impaired object-to-object contextual integration processes.

Recall that Riddoch and colleagues (2003, 2006, 2011) have demonstrated reduced rates of extinction when objects correctly interacted with each other, relative to when they incorrectly interacted with each other. The authors emphasized the importance of appropriate action relations between objects (e.g., a corkscrew oriented toward a cork of a wine bottle) in the modulation of extinction. Their findings suggested that, in contrast to the results of the present study, objects can be preattentively linked, at least when action relations are involved. Note that our stimulus pool contained a smaller proportion of active than passive object pairs. On the face of it, the relatively larger proportion of passive pairs may have accounted for the lack of a spatial consistency effect in the present experiment, since such an effect among the active pairs may have been diluted by the dominance of the passive pairs. However, an inspection of the results suggests that this was not the case. In fact, there was no hint for a spatial consistency effect among the active pairs, when one of the two objects was presented outside the main focus of attention (see Table 5). A 2×2 repeated measures ANOVA for the spatial consistency and the pair activity factors further confirmed that none of the main effects or the interaction effect approached statistical significance level (all F s < 1.1 , all p s > 1.0). A nonsignificant interaction effect was also obtained with the error rate measure ($F < 1$, $p > 1.0$).

Overall, then, a dissociation has emerged between the latency findings of Experiments 1 and 2: While a clear spatial consistency effect was seen when both objects were attended (Experiment 1), no such effect was observed when one of the two objects was task irrelevant and was located outside the

Table 4 Mean reaction times (in milliseconds) and proportions of errors in the object classification task in Experiment 2 (standard errors in parentheses)

	Contextually Related Pairs		Target Pairs		Single Objects	
	Spatially Consistent	Spatially Inconsistent	Both Target	Mixed Target	Valid	Invalid
Reaction time	530 (15)	528 (15)	550 (13)	579 (14)	531 (12)	568 (19)
Proportion of errors	.05 (.01)	.06 (.01)	.08 (.01)	.17 (.02)	.05 (.01)	.12 (.02)

Note. Valid = a single object appearing in a cued location; invalid = a single object appearing in an uncued location. Both Valid and Invalid conditions are averaged across target and nontarget stimuli.

focus of attention (Experiment 2). To further examine the effects of attention on object-to-object contextual processing, we performed an additional mixed ANOVA, containing 2 (spatial consistency: consistent, inconsistent) \times 2 (experiments: 1, 2) factors. This analysis revealed a statistically significant interaction of spatial consistency and experiment, $F(1, 38) = 8.2, p < .01$. Overall, the combined latency results of Experiments 1 and 2 suggest that contextual effects can be modified by attentional allocation. Specifically, our results imply that linking several objects within a visual associative percept requires attention.

What is the nature of the attentional effect found with the stimuli pairs? While narrowing the attentional spotlight to one of two objects has eliminated object-to-object contextual effects, it is still unclear whether the attention manipulation has directly affected visual associative processing or whether it more generally affected stimulus recognition at the unattended (or minimally attended) location. Namely, the lack of a contextual effect in Experiment 2 may have stemmed from poor stimulus perception outside the focus of attention (at the individual object level) or from a specific impairment to the perception of a distractor's contextual associations with a cued (attended) object. In order to estimate the level of stimulus processing outside the focus of attention, at the individual object level, we conducted an additional analysis on the target object stimulus pairs. Note that in trials containing target object stimuli, the cued stimulus was always a nonsense target object, while the uncued distractor was either an additional nonsense target (i.e., bothtarget condition) or a nontarget (real-world) object (i.e., mixed-target condition). This created a flanker-like situation, in which the irrelevant distractor either matched or conflicted with the response to the cued stimulus, respectively. Specifically, when the distractor was

a nonsense target object, it was congruent with the response to the cued stimulus, whereas when it was a real, nontarget object, it was incongruent with the response to the cued stimulus. If participants were able to differentiate nonsense and real-object distractors appearing at the uncued (unattended) location, a response congruency effect should be seen. Namely, longer RTs should be obtained for the incongruent distractors (i.e., mixed-target condition) than for the congruent distractors (i.e., both-target condition). Recall that nontarget stimuli consisted of real-world objects from categories similar to the ones forming the contextually associated pairs (e.g., furniture, tools, household utensils, appliances, clothes, etc.). Thus, a response congruency effect in the present experiment would indicate that participants dissociated these everyday objects from the nonsense stimuli distractors.

An inspection of Table 4 reveals that indeed this was the case. A statistically significant response congruency effect was found with RTs, $t(19) = 4.01, p < .001, ES = 0.9$, as well as with error rates, $t(19) = 4.52, p < .001, ES = 1.01$. These findings imply that recognition processes took place outside the focus of attention, at least to an extent that participants were sensitive to a distractor's categorical identity (real or nonsense object). Despite these visual processing abilities, however, the integration of an unattended distractor object with a contextually related (attended) object was impaired.

Note, however, that real and nonsense objects differed in various visual aspects, including color, texture, brightness, and shape information. It is possible that dissociating these two stimulus categories at the unattended (i.e., uncued) location relied mainly on low-level feature analysis, rather than on high-level processing of the objects' identities. Namely, our results may suggest that coarse visual processing indeed took

Table 5 Mean reaction times (in milliseconds) and proportions of errors for contextually associated active and passive object pairs in Experiment 2 (standard errors in parentheses)

	Active Pairs		Passive Pairs	
	Spatially Consistent	Spatially Inconsistent	Spatially Consistent	Spatially Inconsistent
Reaction time	533 (14)	529 (14)	528 (15)	527 (15)
Proportion of errors	.04 (.01)	.06 (0.1)	.05 (.01)	.07 (.01)

place in the unattended location; however, the extent to which higher-level recognition processes occurred under these conditions yet remains an open question.

To further dissociate object-to-object contextual integration processes from visual object processing per se, we ran an additional experiment that was similar to Experiment 2, yet we now used real-world objects as target stimuli instead of the nonsense stimuli utilized in the present experiment. Specifically, we chose car images as the to-be-detected targets, since the latter are nonanimate in their nature, like the objects forming the contextually associated object pairs, yet they do not overlap with any of the stimulus categories used within the object pairs. The use of a large variety of car exemplars, of different colors, sizes, and orientations, was aimed to prevent participants from tuning themselves to the detection of specific visual features or specific stimulus exemplars. We posited that dissociating these target images from other nonanimate distractor objects relies on more fine-tuned discrimination processes than the ones required in the present experiment with the nonsense stimuli. If a response congruency effect is eliminated with the real-world target objects, one may assume that our attention manipulation mainly affects visual recognition processes occurring outside the focus of attention. Namely, the lack of a contextual effect in Experiment 2 has most likely reflected poor visual processing at the individual object level. If, however, a response congruency effect is observed under these conditions, while at the same time no contextual effect is seen, one may assume that stimulus processing takes place outside the focus of attention, at least to the extent that participants are able to perform basic-level categorical discrimination (e.g., dissociating a car from a sofa). Despite this processing at the individual level, object-to-object associative integration is impaired, suggesting that the attentional manipulation mainly taps contextual-binding processes.

Experiment 3: One of the two objects is unattended (using real-world object targets)

Method

Participants

Eighteen undergraduate students (11 females, 7 males) participated in the experiment for course credit. All participants had normal or corrected-to-normal sight.

Experimental design, stimuli, and procedure

The experimental design and the stimuli were identical to those in Experiment 2, except for replacing the nonsense target objects with car images of different brands, colors,

sizes, and orientations (see stimulus examples in Fig. 4). The procedure was identical to that in Experiment 2.

Results and discussion

Trials on which participants made errors (4%) or trials on which there were extreme outlier responses (less than 1%) were excluded from the RT analysis. Table 6 presents the mean RTs and the proportions of errors for the different conditions in Experiment 3. As can be seen from the table, the spatial-cuing manipulation was effective in summoning spatial attention: A statistically significant difference was found with the RTs between the valid and the invalid cue conditions,³ $t(17) = 2.24, p < .05, ES = 0.53$. The corresponding difference did not approach significance level with the proportion of errors measure, $t(17) = 1.1, p > .1, ES = 0.26$. As in Experiment 2, and in contrast to the results of Experiment 1, no difference was found in the RTs and in the error rate measure between the spatially consistent and the spatially inconsistent conditions, $t(17) = 0.18, p > .1, ES = 0.04$, and $t(17) = 0.65, p > .1, ES = 0.15$, respectively. Overall, then, our results replicated the findings of Experiment 2, in which the contextual effect was eliminated when one of two object

³ Recall that task instructions required participants to respond to a stimulus appearing in the cued location. The invalid cue condition, however, formed a special case in which a single object actually appeared in the *uncued* location. In order to allow similar responses for valid and invalid trials, participants in Experiment 2 were further requested to respond to stimulus identity, regardless of its appearance outside the cue, when there was no stimulus in the cued location (note that these are the typical instructions in cue validity paradigms using a single stimulus target). To account for the oddity of the invalid cue condition and to ensure that the cue validity effect obtained in Experiment 2 was not an artifact of the unique instructions in this condition, we manipulated task instructions in Experiment 3: Approximately half of the participants ($n = 10$) received the original instructions (as in Experiment 2), while the rest of the participants ($n = 8$) received modified instructions, according to which they were requested to respond to a target stimulus (i.e., a car image) *only* if it appeared within the cued location. Namely, participants were requested to press 1 if a target stimulus appeared in the cued location and to press 2 if no target stimulus appeared in the cued location (e.g., in the invalid cue condition). Thus, instructions were now equated for all conditions, yet valid and invalid target stimuli required different responses (1 and 2, respectively). We wished to examine whether manipulating the instructions would modify the cue validity effect, and specifically, we attempted to determine whether the cue validity effect obtained in Experiment 2 was a mere artifact of task instructions. Results showed that both instruction versions yielded similar cue validity effects: Mean RTs for the valid and invalid cue conditions in the original instructions version were 468 and 489 ms, respectively, and the corresponding RTs for the new instructions version were 466 and 480 ms. A mixed ANOVA containing 2 (cue validity: valid, invalid) \times 2 (experimental instructions: old, new) factors revealed a statistically significant cue validity effect, $F(1, 16) = 4.52, p < .05$, with no instructions effect or validity \times instructions interaction effect (all F s < 1.0 , all p s $> .1$). We therefore concluded that the cuing manipulation was equally effective in the two instruction versions. Notably, our results suggested that the cue validity effect obtained in Experiment 2 was not a mere artifact of the unique task instructions for the invalid cue condition.

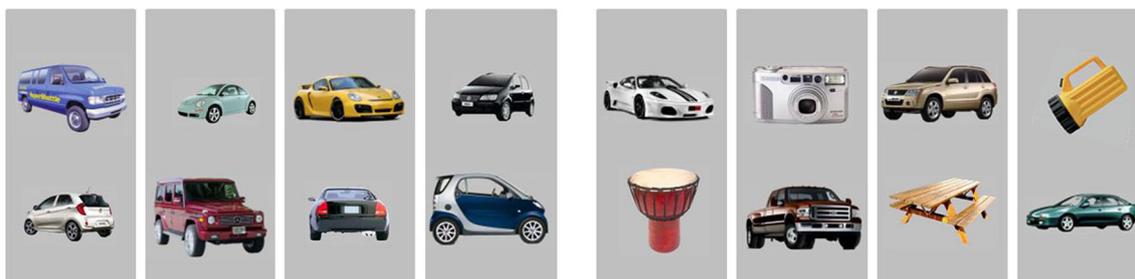


Fig. 4 Examples of the car target images in Experiment 3: Both-target condition (*left*) and mixed-target condition (*right*)

stimuli were unattended (or, at most, minimally attended). Importantly, in spite of the use of real-world stimuli as to-be-detected targets in the present experiment, participants successfully dissociated the latter from other nonanimate object distractors, as reflected by the robust response congruency effect obtained with RTs, $t(17) = 6.27$, $p < .001$, $ES = 1.48$, and with the error rate measure, $t(17) = 5.26$, $p < .001$, $ES = 1.24$. These findings imply that stimuli were processed to a rather high level at the unattended location, yet linkage of such stimuli with their counterpart (attended) pair objects was nevertheless impaired. Note that in order to link an unattended (e.g., kettle) object with its pair object (e.g., mug), one need not identify the unattended object at the exemplar (i.e., subordinate) level (e.g., a blue metal kettle). Rather, identification of the unattended object at a basic categorical level typically suffices in order to create a conjoined, meaningful contextual percept (a kettle pouring into a mug). Thus, while discriminating car distractors from other nonanimate object distractors most likely relied on basic-level categorization, such categorization may have, potentially, served as a basis for object-to-object contextual integration. Contextual associative processes, however, were nevertheless compromised when one of two objects was presented outside the main focus of attention.

As for the differences between active and passive object pairs, our results replicated the results of Experiment 2 in suggesting that neither type of associative relations are processed in the absence of spatial attention (see Table 7). A 2×2 repeated measures ANOVA for the spatial consistency and the pair activity factors revealed a main effect of the activity type factor

in both RT and error rate measures, $F(1, 17) = 5.0$, $p < .05$, and $F(1, 17) = 26.8$, $p < .001$, respectively. Namely, participants were overall faster and more accurate in responding to stimuli belonging to active rather than to passive pairs. Critically, however, there was no effect of the spatial consistency factor or of the interaction between the spatial consistency and the activity factors in either measure (all F s < 1.1 , all p s > 1.0), suggesting that active and passive associative relations behaved similarly under the unattended conditions. Taken together, the findings of Experiment 3 suggest that object-to-object contextual integration processes necessitate attention and that this necessity is at least partially independent of object processing at the individual level.

Experiment 4: Contribution of exogenous and endogenous attention factors to contextual integration

Experiments 2 and 3 suggested that object-to-object contextual-binding processes require attentional capacity. Control over participants' attention in these experiments was achieved by drawing visual attention to a cued location and by prioritizing the cued stimulus in terms of task-requirements (i.e., turning the cued object into a task-relevant stimulus while the uncued object became a task-irrelevant distractor). Note that these two attention manipulations—shifting the locus of visual attention and altering the relevance of a stimulus to the task at hand—may have each contributed independently to the dissociation obtained in latency performance between Experiment 1 and Experiments 2 and 3. Namely, the

Table 6 Mean reaction times (in milliseconds) and proportions of errors in the object classification task in Experiment 3 (standard errors in parentheses)

	Contextually Related Pairs		Target Pairs		Single Objects	
	Spatially Consistent	Spatially Inconsistent	Both Target	Mixed Target	Valid	Invalid
Reaction time	453 (12)	454 (12)	456 (10)	475 (10)	467 (11)	485 (15)
Proportion of errors	.03 (.01)	.03 (.01)	.03 (.01)	.07 (.01)	.03 (.01)	.04 (.01)

Table 7 Mean reaction times (in milliseconds) and proportions of errors for contextually associated active and passive object pairs in Experiment 3 (standard errors in parentheses)

	Active Pairs		Passive Pairs	
	Spatially Consistent	Spatially Inconsistent	Spatially Consistent	Spatially Inconsistent
Reaction time	450 (12)	452 (12)	456 (12)	456 (12)
Proportion of errors	.02 (.01)	.02 (.01)	.03 (.01)	.04 (.01)

lack of a spatial consistency effect in the latter may have resulted from the allocation of visual attention to a salient location, from altering task instructions, or from both. The goal of Experiment 4 was to uncouple these potentially dissociated factors and to assess their relative contribution to contextual integration processes. In Experiment 4A, we therefore used a display similar to that in Experiments 2 and 3, in which visual attention was summoned to one of two objects prior to stimuli appearance, yet now both stimuli within a pair (rather than the cued object only) were turned into task-relevant items. We asked whether an uncued object was associatively linked to a cued object when it was part of the task set.

In Experiment 4B, we examined a reversed situation and asked whether an object that is irrelevant to task demands (i.e., a distractor), yet positioned in a visually attended location, could be associatively linked to a task-relevant object. We hypothesized that in both cases (i.e., Experiments 4A and 4B), contextual binding would be observed, since both objects within a pair received attentional resources (whether by manipulating task demands or spatial location). To allow a direct comparison of the present results with the findings of Experiment 1, in which both objects appeared inside the main focus of attention, we used the same set of stimuli as in Experiment 1. We focus our results analyses on the spatial consistency effect within each of these experiments.

Experiment 4A: One of two objects is cued, yet both objects are task relevant

Method

Participants

Seventeen undergraduate students (16 females, one male), with normal or corrected-to-normal sight, participated in the experiment for course credit.

Stimuli

The stimuli were identical to those used in Experiment 1.

Experimental design and procedure

The experimental design and the procedure were identical to those in Experiments 2 and 3, except that participants were required to perform an object classification task on both stimuli appearing in the display (Is there a target object in the display?), rather than on the cued stimulus only (Is there a target object in the cued location?). By altering task instructions, the uncued object became an important part of the task set and was now highly relevant to task requirements (recall that similar instructions were given in Experiment 1). The spatial cue, in turn, became irrelevant or negligible, and participants were explicitly instructed to ignore it altogether. We assumed that despite the new instructions, participants would nevertheless shift their attention toward the peripheral cue, allowing enhanced attention allocation at the cued location (see, e.g., Jonides, 1981; Posner & Cohen, 1984). This allowed us to examine the spatial consistency effect under conditions in which the locations of both stimuli were part of the task set, yet one of the stimuli was underprivileged due to its appearance in an uncued location (for a similar manipulation, see, e.g., Roberts & Humphreys, 2011). We hypothesized that in spite of a shift of attention toward the cued object, participants would nevertheless cluster the two associated objects within a coherent contextual representation, due to an endogenous allocation of attention to both objects.

To ensure that the cue was sufficiently salient and allowed an efficient shift of attention, under conditions in which both object locations were relevant to task requirements, the cue (i.e., a fixation cross) was slightly enlarged and was changed from black to red (see Fig. 5). All other aspects of the experimental procedure were identical to those in Experiments 2 and 3.

Results and discussion

As in previous experiments, trials in which participants made errors (9.7%) or trials in which there were extreme outlier responses (less than 1%) were excluded from the RT analysis. Table 8 presents the mean RTs and the proportion of errors for the different conditions in Experiment 4A. As a first step, we examined the differences among the single items between the

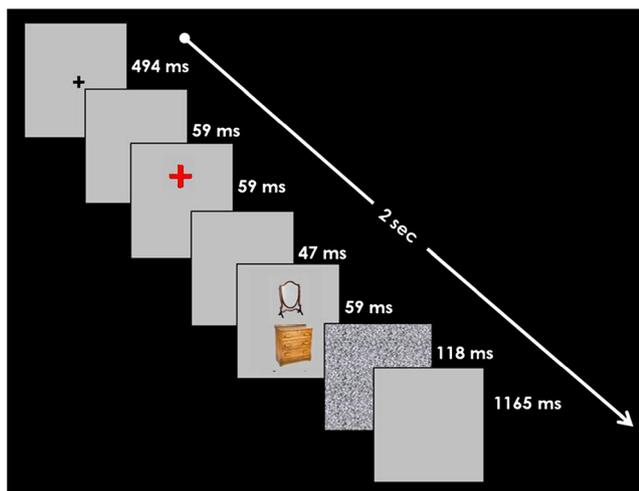


Fig. 5 An example of a trial sequence in Experiment 4A

valid and the invalid cue conditions. Note that the nonsense target stimuli were excluded from the single object results in Table 8.⁴ As can be seen, despite the irrelevance of the cue to task requirements, trials in the valid cue condition yielded significantly shorter RTs than did trials in the invalid cue condition among the real-world objects, $t(16) = 3.43$, $p < .005$, $ES = 0.83$. The difference in the proportion of errors between the two conditions was statistically significant as well, $t(16) = 2.13$, $p < .05$, $ES = 0.52$. This cue validity effect suggests that the cue indeed induced a shift of attention to the cued location. When examining the spatial consistency effect we found that, as hypothesized, significantly shorter RTs were obtained for the spatially consistent than for the spatially

⁴ As was mentioned above, in order to increase its visual saliency, the spatial cue was enlarged, and its color was turned from black to red. Surprisingly, despite the cue's relative saliency, a small, nonsignificant validity effect was obtained under these conditions [mean RT for the valid condition, 553 ms; mean RT for the invalid condition, 559 ms; $t(16) = 1.02$, $p > .1$]. Recall that the validity effect was computed across both real-world (nontarget) and nonsense (target) objects. When these were analyzed separately, however, a significant cue validity effect was found among the real objects (see Table 8), but not among the nonsense stimuli [a nonsignificant trend in the opposite direction was observed with the latter; mean RT for the valid nonsense stimuli, 529 ms; for the invalid nonsense stimuli, 516 ms; $t(16) = 1.65$, $p > .1$]. We believe that an interaction between the colored cue and the colorful nonsense stimuli caused this unexpected result with the nonsense single objects. Note that approximately 40% of nonsense stimuli contained unicolored shapes, which may have been perceived as somewhat similar in nature to the red cue. It appears, then, that altering the color of the cue from black to red caused participants to confuse it with the colorful nonsense objects (particularly when both cue and nonsense stimulus appeared in the same location—i.e., in the valid condition), resulting in the elimination of the cue validity effect for the nonsense single stimuli. Importantly, the significant validity effect found among the real-world single objects suggests that, within the spatially consistent/inconsistent (real-world) object pairs, the cued object was indeed prioritized, relative to the uncued object. We therefore present in Table 8 the results for the valid and invalid cue conditions among the real-world items only.

inconsistent object pairs, $t(16) = 3.16$, $p < .01$, $ES = 0.77$. A significant difference between consistent and inconsistent trials was also obtained with the proportion of errors, $t(16) = 2.15$, $p < .05$, $ES = 0.52$. These findings indicate that despite its underprivileged status, the uncued object was successfully linked to the cued object when both stimuli were highly relevant to participants' goals (note that the effect size for the spatial consistency effect with the RTs was equivalent to that obtained in Experiment 1 and was higher than the latter with the error rate measure). The results of Experiment 4A therefore suggest that when both object locations are relevant to task requirements, object-to-object contextual-binding processes can take place even if attention is spatially oriented to only one of the two stimuli.

Experiment 4B: A task-irrelevant distractor appears at a visually attended location

Experiment 4A revealed the importance of endogenous attention factors to contextual processing. We found that an uncued object was associatively linked to a cued object when task instructions encouraged participants to intentionally allocate attention to both stimuli. In Experiment 4B, we wished to explore an alternative situation in which an object is located in a visually attended location yet serves as a task-irrelevant distractor. Note that a contextual effect was previously obtained when both object locations were relevant to task demands (Experiments 1 and 4A). In Experiments 2 and 3, in contrast, an irrelevant distractor appeared outside the main focus of visual attention (i.e., in an uncued location), resulting in the elimination of latency differences between spatially consistent and inconsistent conditions. It is possible that the relevance of both pair stimuli to the object classification task is a necessary condition for contextual integration processes. Thus, once an object turns into a task-irrelevant distractor, processing of its relations with a contextually associated object is avoided. In the present experiment, we asked whether an irrelevant distractor may nevertheless be associatively linked with a task-relevant object if it is positioned in a visually attended location. We assumed that despite its irrelevance to task demands, the distractor receives at least some attention when it appears in a highly prioritized location. Our goal was to determine whether, under such conditions, it is automatically bound with a task-relevant object.

Method

Participants

Twenty two undergraduate students (16 females, 6 males), with normal or corrected-to-normal sight, participated in the experiment for course credit.

Table 8 Mean reaction times (in milliseconds) and proportions of errors in the object classification task in Experiment 4A (standard errors in parentheses)

	Contextually Related Pairs		Target Pairs		Single Objects	
	Spatially Consistent	Spatially Inconsistent	Both Target	Mixed Target	Valid	Invalid
Reaction time	567 (17)	576 (17)	522 (17)	573 (19)	577 (20)	603 (19)
Proportion of errors	.09 (.01)	.11 (.01)	.04 (.01)	.14 (.02)	.11 (.02)	.15 (.03)

Note. Nonsense target stimuli were excluded from the single object results (see Footnote 4).

Experimental design and stimuli

The experimental conditions were identical to the one used in the previous experiment, aside from the exclusion of the single-object conditions, due to the fact that no cue validity effect was measured.

Procedure

The procedure was similar to that in the previous experiment, yet it was slightly altered in order to accommodate for the experiment's goals. On each trial, two stimuli appeared simultaneously, one of which was always centered at fixation and the other of which was centered approximately 6° above (50%) or below (50%) fixation. Note that the relative position of stimuli on the screen was changed, yet the center-to-center distance between pair stimuli remained identical to that in previous experiments. The trial sequence was identical to previous experiments, aside from the exclusion of the spatial cue appearing prior to the object images. Namely, object stimuli appeared immediately after the disappearance of the fixation cross. As for task instructions, participants were requested to ignore the fixated object and to perform an object classification task only on the upper/lower object, while maintaining fixation on the central stimulus. The peripheral object was, therefore, task relevant, while the fixated object served as a task-irrelevant distractor. Since, in everyday vision, spatial attention is tightly correlated with eye fixations (see, e.g., Belopolsky & Theeuwes, 2009; Hoffman & Subramaniam, 1995), we assumed that the fixated distractor receives some attentional resources, despite its irrelevance to task demands. In particular, we assumed that it would be difficult to ignore a distractor at fixation when the target's locus (i.e., upper/lower location) is unknown in advance. Accordingly, we hypothesized that spatial attention would be endogenously deployed to the peripheral, task-relevant target, yet one's attentional spotlight will effectively span both target and distractor locations due to the appearance of the distractor at fixation.

Results and discussion

Trials on which participants made errors (11.3%) or trials on which there were extreme outlier responses (approximately

1.5%) were excluded from the RT analysis. Table 9 presents the mean RTs and the proportions of errors for the different conditions in Experiment 4B. As can be seen from the table, significantly shorter RTs were obtained for spatially consistent than for spatially inconsistent object pairs, $t(21) = 2.11$, $p < .05$, $ES = 0.45$. Accuracy levels did not differ significantly between the two conditions, $t(21) = 1.16$, $p > .05$, $ES = 0.25$. These results suggest that despite their irrelevance for task requirements, the fixated distractor objects affected response latencies to the contextually associated, peripheral (task-relevant) objects. It appears, then, that object-to-object associative processes may take place regardless of task relevance as long as stimuli are positioned at a highly prioritized location (e.g., fixation). Taken together, the results of Experiments 4A and 4B suggest that once objects are attended, by virtue of their relevance to task requirements or their spatial location, they are automatically clustered into a meaningful visual percept. Note, however, that contextual effects were larger in magnitude in Experiment 4A than in Experiment 4B, further implying that endogenous attention factors may play a more dominant role in contextual binding of objects than exogenous factors.

General discussion

The present research examined the role of attention in the perception of object-to-object contextual relations during a brief visual glimpse. This issue has attracted growing attention in the last decade, leading to inconsistent findings. Our basic paradigm included pairs of associated objects, which were positioned in either contextually plausible or implausible locations. We measured participants' performance in an unbiased object classification task, while carefully manipulating participants' locus of visual attention. In Experiment 1, both pair objects were relevant to task demands, and thus, attention was presumably drawn to both object locations. As was expected, shorter RTs were obtained for contextually consistent than for inconsistent stimuli. In Experiment 2, attention was drawn to one of two objects, while its pair object served as an irrelevant distractor appearing outside the main focus of attention (i.e., in an uncued location). Under these conditions, latency differences between the two contextual conditions

Table 9 Mean reaction times (in milliseconds) and proportions of errors in the object classification task in Experiment 4B (standard errors in parentheses)

	Contextually Related Pairs		Target Pairs	
	Spatially Consistent	Spatially Inconsistent	Both Target	Mixed Target
Reaction time	567 (11)	573 (12)	554 (13)	584 (15)
Proportion of errors	.12 (.01)	.14 (.01)	.07 (.01)	.12 (.01)

were eliminated. Experiment 3 further replicated the results of Experiment 2, with the additional demonstration that object-to-object contextual effects are somewhat independent from individual stimulus processing outside the focus of visual attention. In Experiment 4, we examined contextual processing when one of two objects was uncued but was task relevant or when it was irrelevant to task requirements yet was positioned in a visually attended location. Contextual binding was once again observed under these conditions. Note that in all experiments, the contextual relations between object pairs were irrelevant to the object classification task; thus, contextual processing presumably occurred in an involuntary fashion. Taken together, our results strongly suggest that when stimuli are attended objects, they are successfully linked within a coherent contextual representation. If, however, objects appear outside the main focus of visual attention and they are strictly task irrelevant, contextual processing is dramatically impaired. Note that under such conditions, participants nevertheless continue to process individual unattended stimuli, at least to an extent that basic-level categorization is enabled (e.g., dissociating a car from a table).

Previous studies examining visual contextual processing have typically used whole-scene displays, in which detecting a particular object within the scene, or shifting one's gaze toward a peripheral object, served as an index for contextual processing. Interestingly, there seems to be an unresolved paradox in the visual scene literature concerning the effects of context on object recognition. On the one hand, classic contextual models have argued for a rapid perceptual priming mechanism that enhances contextually consistent objects within a scene (e.g., Bar, 2004; Biederman et al., 1982; Davenport & Potter, 2004; Palmer, 1975). Contextually inconsistent objects, according to these models, suffer from degraded recognition, relative to consistent ones. On the other hand, as was reviewed in the Introduction, several researchers have argued for an early prioritization of contextually inconsistent objects, with the latter capturing attention earlier than contextually consistent objects (e.g., Becker et al., 2007; Gordon, 2004; Loftus & Mackworth, 1978; Underwood & Foulsham, 2006; Underwood et al., 2008; see also Mudrik et al., 2011). Findings of an early shift of attention or an eye movement toward contextually inconsistent objects within scenes have strongly supported attention-free views of visual contextual processing. Our findings have revealed no

difference between contextually consistent and inconsistent objects appearing outside the main focus of attention, in contrast to the difference observed when stimuli are presented inside the attentional focus, suggesting that contextual integration is not an attention-free process.

Several differences between our study and the studies mentioned above, however, may have accounted for the differences in the results. First, the present study used a focal attention paradigm, in which participants' attention was drawn to a specific stimulus, while its pair stimulus was presented outside the focus of visual attention (Experiments 2 and 3). In scene-object consistency paradigms, in contrast, participants are typically allowed to inspect the scene freely in a memory encoding or a visual search task. Participants may therefore engage in a rather diffused, or spread-attention, mode, which enables an early allocation of attention across a greater spatial region within the scene. Second, the scene-object consistency effects mentioned above were observed following relatively long scene exposure durations (150–200 ms or longer) but, interestingly, not during short exposure durations (40–130 ms) (Gordon, 2004). In some cases, several seconds (during which several eye movements were made) were required in order to observe an eye movement toward contextually inconsistent objects in a scene (e.g., Loftus & Mackworth, 1978; Underwood & Foulsham, 2006; Underwood et al., 2008). These relatively long exposure durations, together with the rather weak control over the locus of participants' attentional focus, strongly suggest that attention may have shifted toward contextually inconsistent objects prior to their identification as odd items. Note that several studies have failed to replicate the finding of an early impact of scene inconsistency on eye movements (e.g., De Graef et al., 1990; Gareze & Findlay, 2007; Henderson et al., 1999; Rayner et al., 2009; Vö & Henderson, 2009, 2011). The main conclusion from the latter studies, in accordance with the results of the present research, is that contextually inconsistent objects do not capture attention but, rather, engage attention (and eye gaze) once attention has already been drawn toward the objects.

Perhaps more relevant to the present study are the findings of unattended grouping of object pairs among patients suffering from parietal lobe lesions. Recall that Riddoch and colleagues (2003, 2006; 2011) demonstrated reduced rates of extinction when stimuli pairs were associated by action relations (e.g., a corkscrew oriented toward a cork of a wine

bottle). Their findings suggested that objects are preattentively linked when involving action relations. The results of the present study do not support the findings of Riddoch and colleagues, since active and passive object pairs behaved similarly under unattended, as well as attended, conditions. Specifically, when one of two objects in a pair was unattended (or, at most, minimally attended), no spatial effect was seen regardless of the type of contextual association (Experiments 2 and 3), whereas when both object stimuli were attended, a spatial consistency effect was evident in both activity conditions (Experiment 1). Thus, while several previous studies have stressed the importance of active object-to-object relations in perceptual processing (e.g., Green & Hummel, 2006; Roberts & Humphreys, 2010, 2011), our findings imply that passive contextual associations (i.e., associations that do not involve implied action) may, in fact, be as efficient as active associations in gluing objects within a coherent visual setting, given that objects are attended.

Note, however, that when accounting for the lack of a contextual effect under the unattended condition, one needs to take into account the fact that our pool of contextually related stimuli contained a smaller proportion of active pairs, in comparison with the passive pairs. Thus, although there seemed to be no hint for the contextual integration of active pairs under the unattended condition, one might need to run additional follow-up experiments and to explicitly manipulate the activity-type factor (e.g., using equal proportions of active and passive pairs). Additionally, and more critically, note that an important difference exists between the present study and the studies of Riddoch et al. (2003, 2006, 2011). In Experiments 2 and 3, participants' attention was manipulated by shifting the locus of visual attention, as well as manipulating the relevance of stimuli to the task at hand. Namely, the unattended stimulus appeared outside the main focus of visual attention (by virtue of the cuing manipulation), and it additionally served as a task-irrelevant distractor (by virtue of task requirements). In the Riddoch et al. study, in contrast, while the contralesional stimulus clearly appeared in an underprivileged (i.e., neglected) location, it was nevertheless relevant to task requirements. Namely, patients were explicitly asked to report all objects in the display. Both ipsilesional and contralesional objects, therefore, served as to-be-identified stimuli. It is possible that despite the presentation of objects in the neglected field, patients were oriented to search for object pairs and were tuned to detect the potential relations between the stimuli. The explicit search of a task-relevant object, as opposed to the irrelevance of the unattended distractor in the present study, may have accounted for the prioritization of active pair objects and the reduced extinction rates obtained with these pairs in the Riddoch et al. study. While this suggestion is mere speculation, further research is required to clarify the precise role of active and passive relations in object-to-object preattentive

binding, when one of the two objects is strictly irrelevant to task requirements.

Finally, our study taps a more general question concerning object recognition in the absence of visual attention. This question has a long history and has triggered renewed interest in recent years. Traditional views of early-selection attention have argued that visual recognition necessitates focal attention (e.g., Broadbent, 1958; Mack & Rock, 1998; Treisman & Gelade, 1980; see also Lachter, Forster, & Ruthruff, 2004). Late-selection models, in contrast, have contended that stimuli may be processed up to a rather high level in a preattentive fashion (e.g., Deutsch & Deutsch, 1963). Supporting evidence for the latter view has emerged from recent studies demonstrating efficient categorization and recognition of real-world objects (as opposed to meaningless, arbitrary stimuli), in the near absence of focal attention (Li et al., 2002; Poncet et al., 2012; Reddy, Reddy, & Koch, 2006; Reddy et al., 2004; see also VanRullen et al., 2005). Yet other recent studies have failed to replicate the findings of attention-free object categorization and recognition (e.g., M. A. Cohen et al., 2011; Walker et al., 2008; see also Evans & Triesman, 2005; Scharff, Palmer, & Moore, 2011).

The results of the present study suggest that stimulus discrimination based on categorical identity may persist even when stimuli are unattended (e.g., Experiment 3). Namely, basic-level categorical information may be available despite the appearance of stimuli outside the main focus of attention. One reservation, however, is in order with regard to the latter conclusion. Note that car images were used as targets in the present study (Experiment 3) and were successfully discriminated from other nonanimate object distractors when presented at an uncued location. Special effort was made to present a wide variety of car images of different orientations, colors, and types, in order to prevent participants from tuning themselves to the detection of specific visual features or specific stimulus exemplars. Car stimuli, however, are typically longer on the horizontal dimension than are other everyday nonanimate objects (especially when viewed in profile). Despite our attempts to prevent basic-level categorization based on simple feature detection, and although the noncar distractor objects contained a considerable number of elongated stimuli as well (e.g., tables of various types, sofas, rectangular objects), cars tend to be, on average, more elongated than most of the other everyday objects used. Thus, stimulus orientation may have possibly served as a low-level visual cue contributing to the discrimination of target and nontarget stimuli outside the focus of attention. In addition, the discrimination of car stimuli from other nonanimate objects was measured by a response congruency effect that may have been influenced by factors other than an object's categorical identity per se. For instance, searching for a specific target category may have prioritized its features' representation and may

have facilitated its processing relative to other nontarget (i.e., task-irrelevant) categories. Furthermore, response competition factors may have contributed to the difference in RTs between response-congruent and response-incongruent trials, resulting in an overestimation of object discrimination performance when distractor stimuli did not compete with responses to cued objects. Finally, the object discrimination task in the present study may have not been sufficiently difficult to create a high attentional load; thus, spare resources may have contributed to the distractor processing outside the focus of attention (see, e.g., M. A. Cohen et al., 2011). Taking these limitations in mind, our findings nevertheless imply that at least under certain circumstances, some categorical processing likely takes place at unattended (or minimally attended) locations, yet this processing does not suffice for object-to-object contextual integration. Further research is clearly required in order to better understand the exact conditions under which categorization and identification of individual objects may persist in the absence of visual attention.

Acknowledgments We thank Yifat Rosenberg, Rotem Amar, Anna Izoutcheev, Ram Kalendarev, and Tsafnat Nave for their assistance in experiment programming and data collection. We also thank Kyle Cave and an anonymous reviewer for their helpful comments on a previous version of the manuscript. This research was supported by the National Institute for Psychobiology in Israel (NIPi) and by the Israel Science Foundation (grant No. 338/11 to Nurit Gronau).

References

- Antes, J. R., Penland, J. G., & Metzger, R. L. (1981). Processing global information in briefly presented pictures. *Psychological Research*, 43(3), 277–292. doi:10.1007/BF00308452
- Auckland, M. E., Cave, K. R., & Donnelly, N. (2007). Nontarget object can influence perceptual processes during object recognition. *Psychonomic Bulletin & Review*, 14(2), 332–337. doi:10.3758/BF03194073
- Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, 5(8), 617–629. doi:10.1038/nrn1476
- Bar, M., & Ullman, S. (1996). Spatial context in recognition. *Perception*, 25(3), 343–352. doi:10.1068/p250343
- Becker, M. W., Pashler, H., & Lubin, J. (2007). Object-intrinsic oddities draw early saccades. *Journal of Experimental Psychology: Human Perception and Performance*, 33(1), 20–30. doi:10.1037/0096-1523.33.1.20
- Belopolsky, A. V., & Theeuwes, J. (2009). When are attention and saccade preparation dissociated? *Psychological Science*, 20(11), 1340–1347. doi:10.1111/j.1467-9280.2009.02445.x
- Biederman, I. (1972). Perceiving real-world scenes. *Science*, 177(4043), 77–80. doi:10.1126/science.177.4043.77
- Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, 14(2), 143–177. doi:10.1016/0010-0285(82)90007-X
- Bonitz, V. S., & Gordon, R. D. (2008). Attention to smoking-related and incongruous objects during scene viewing. *Acta Psychologica*, 129(2), 255–263. doi:10.1016/j.actpsy.2008.08.006
- 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. doi:10.1037/0096-1523.15.3.556
- Breska, A., Israel, M., Maoz, K., Cohen, A., & Ben-Shakhar, G. (2011). Personally-significant information affects performance only within the focus of attention: A direct manipulation of attention. *Attention, Perception, & Psychophysics*, 73(6), 1754–1767. doi:10.3758/s13414-011-0134-6
- Broadbent, D. E. (1958). *Perception and communication*. Elmsford: Pergamon Press. doi:10.1037/10037-000
- Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, 36(1), 28–71. doi:10.1006/cogp.1998.0681
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: Lawrence Erlbaum Associates, Inc.
- Cohen, M. A., Alvarez, G. A., & Nakayama, K. (2011). Natural-scene perception requires attention. *Psychological Science*, 22(9), 1165–1172. doi:10.1177/0956797611419168
- Davenport, J. L. (2007). Consistency effects between objects in scenes. *Memory & Cognition*, 35(3), 393–401. doi:10.3758/BF03193280
- Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, 15(8), 559–564. doi:10.1111/j.0956-7976.2004.00719.x
- De Graef, P., Christiaens, D., & d'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research*, 52(4), 317–329. doi:10.1007/BF00868064
- Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, 70(1), 80–90. doi:10.1037/h0039515
- Devue, C., & Brédart, S. (2008). Attention to self-referential stimuli: Can I ignore my own face? *Acta Psychologica*, 128(2), 290–297. doi:10.1016/j.actpsy.2008.02.004
- Devue, C., Van, d. S., Brédart, S., & Theeuwes, J. (2009). You do not find your own face faster; you just look at it longer. *Cognition*, 111(1), 114–122. doi:10.1016/j.cognition.2009.01.003
- Droll, J., & Eckstein, M. (2008). Expected object position of two hundred fifty observers predicts first fixations of seventy seven separate observers during search. *Journal of Vision*, 8(6), 320. doi:10.1167/8.6.320
- Evans, K. K., & Treisman, A. (2005). Perception of objects in natural scenes: Is it really attention free? *Journal of Experimental Psychology: Human Perception and Performance*, 31(6), 1476–1492. doi:10.1037/0096-1523.31.6.1476
- Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision*, 7(1), 10–10. doi:10.1167/7.1.10
- Fei-Fei, L., VanRullen, R., Koch, C., & Perona, P. (2005). Why does natural scene categorization require little attention? exploring attentional requirements for natural and synthetic stimuli. *Visual Cognition*, 12(6), 893–924. doi:10.1080/13506280444000571
- Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General*, 108(3), 316–355. doi:10.1037/0096-3445.108.3.316
- Ganis, G., & Kutas, M. (2003). An electrophysiological study of scene effects on object identification. *Cognitive Brain Research*, 16(2), 123–144. doi:10.1016/S0926-6410(02)00244-6
- Gareze, L., & Findlay, J. M. (2007). Absence of scene context effects in object detection and eye gaze capture. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray & R. L. Hill (Eds.), (pp. 617–637). Amsterdam, Netherlands: Elsevier.
- Gordon, R. D. (2004). Attentional allocation during the perception of scenes. *Journal of Experimental Psychology: Human Perception and Performance*, 30(4), 760–777. doi:10.1037/0096-1523.30.4.760

- Green, C., & Hummel, J. E. (2006). Familiar interacting object pairs are perceptually grouped. *Journal of Experimental Psychology: Human Perception and Performance*, 32(5), 1107–1119. doi:10.1037/0096-1523.32.5.1107
- Gronau, N., Cohen, A., & Ben-Shakhar, G. (2003). Dissociations of personally significant and task-relevant distractors inside and outside the focus of attention: A combined behavioral and psychophysiological study. *Journal of Experimental Psychology: General*, 132(4), 512–529. doi:10.1037/0096-3445.132.4.512
- Gronau, N., Cohen, A., & Ben-Shakhar, G. (2009). Distractor interference in focused attention tasks is not mediated by attention capture. *The Quarterly Journal of Experimental Psychology*, 62(9), 1685–1695. doi:10.1080/17470210902811223
- Gronau, N., Neta, M., & Bar, M. (2008). Integrated contextual representation for objects identities and their locations. *Journal of Cognitive Neuroscience*, 20(3), 371–388. doi:10.1162/jocn.2008.20027
- Harris, C. R., & Pashler, H. (2004). Attention and the processing of emotional words and names: Not so special after all. *Psychological Science*, 15(3), 171–178. doi:10.1111/j.0956-7976.2004.01503005.x
- Henderson, J. M., Pollatsek, A., & Rayner, K. (1987). Effects of foveal priming and extrafoveal preview on object identification. *Journal of Experimental Psychology: Human Perception and Performance*, 13(3), 449–463. doi:10.1037/0096-1523.13.3.449
- Henderson, J. M., Weeks, P. A., Jr., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, 25(1), 210–228. doi:10.1037/0096-1523.25.1.210
- Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception & Psychophysics*, 57(6), 787–795. doi:10.3758/BF03206794
- Hollingworth, A. (2007). Object-position binding in visual memory for natural scenes and object arrays. *Journal of Experimental Psychology: Human Perception and Performance*, 33(1), 31–47. doi:10.1037/0096-1523.33.1.31
- Hollingworth, A., & Henderson, J. M. (1998). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology: General*, 127(4), 398–415. doi:10.1037/0096-3445.127.4.398
- Hollingworth, A., & Henderson, J. M. (2000). Semantic informativeness mediates the detection of changes in natural scenes. *Visual Cognition*, 7(1–3), 213–235. doi:10.1080/135062800394775
- Jonides, J. (1981). Voluntary versus automatic control over the mind's eye's movement. In J. B. Long & A. D. Baddeley (Eds.), *Attention and performance IX* (pp. 187–203). Hillsdale: Lawrence Erlbaum Associates.
- Joubert, O. R., Fize, D., Rousselet, G. A., & Fabre-Thorpe, M. (2008). Early interference of context congruence on object processing in rapid visual categorization of natural scenes. *Journal of Vision*, 8(13), 1–18. doi:10.1167/8.13.11
- Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, 24(2), 175–219. doi:10.1016/0010-0285(92)90007-O
- Kim, J. G., & Biederman, I. (2011). Where do objects become scenes? *Cerebral Cortex*, 21(8), 1737–1746. doi:10.1093/cercor/bhq240
- Lachter, J., Forster, K. I., & Ruthruff, E. (2004). Forty-five years after broadbent (1958): Still no identification without attention. *Psychological Review*, 111(4), 880–913. doi:10.1037/0033-295X.111.4.880
- Fei-Fei, L., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences*, 99(14), 9596–9601. doi:10.1073/pnas.092277599
- Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance*, 4(4), 565–572. doi:10.1037/0096-1523.4.4.565
- Mack, A., & Rock, I. (1998). *Inattention blindness*. Cambridge: The MIT Press.
- Mudrik, L., Breska, A., Lamy, D., & Deouell, L. Y. (2011). Integration without awareness: Expanding the limits of unconscious processing. *Psychological Science*, 22(6), 764–770. doi:10.1177/0956797611408736
- Mudrik, L., Lamy, D., & Deouell, L. Y. (2010). ERP evidence for context congruity effects during simultaneous object–scene processing. *Neuropsychologia*, 48(2), 507–517. doi:10.1016/j.neuropsychologia.2009.10.011
- Munneke, J., Brentari, V., & Peelen, M. V. (2013). The influence of scene context on object recognition is independent of attentional focus. *Frontiers in Psychology*, 4(552), 1–10. doi:10.3389/fpsyg.2013.00552
- Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research*, 46(5), 614–621. doi:10.1016/j.visres.2005.08.025
- Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology*, 41, 176–210.
- Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3), 145–175. doi:10.1023/A:1011139631724
- Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, 11(12), 520–527. doi:10.1016/j.tics.2007.09.009
- Oppermann, F., Hassler, U., Jescheniak, J., & Gruber, T. (2012). The rapid extraction of gist—early neural correlates of high-level visual processing. *Journal of Cognitive Neuroscience*, 24(2), 521–529. doi:10.1162/jocn_a_00100
- Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition*, 3(5), 519–526. doi:10.3758/BF03197524
- Poncet, M., Reddy, L., & Fabre-Thorpe, M. (2012). A need for more information uptake but not focused attention to access basic-level representations. *Journal of Vision*, 12(1):15, 1–16. doi:10.1167/12.1.15
- Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 531–556). Hillsdale: Lawrence Erlbaum Associates.
- Potter, M. C. (1975). Meaning in visual search. *Science*, 187, 965–966.
- Rayner, K., Castelano, M. S., & Yang, J. (2009). Eye movements when looking at unusual/weird scenes: Are there cultural differences? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(1), 254–259. doi:10.1037/a0013508
- Reddy, L., Reddy, L., & Koch, C. (2006). Face identification in the near-absence of focal attention. *Vision Research*, 46(15), 2336–2343. doi:10.1016/j.visres.2006.01.020
- Reddy, L., Wilken, P., & Koch, C. (2004). Face-gender discrimination is possible in the near-absence of attention. *Journal of Vision*, 4(2), 106–117. doi:10.1167/4.2.4
- Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, 8(5), 368–373. doi:10.1111/j.1467-9280.1997.tb00427.x
- Riddoch, M. J., Humphreys, G. W., Hickman, M., Clift, J., Daly, A., & Colin, J. (2006). I can see what you are doing: Action familiarity and affordance promote recovery from extinction. *Cognitive Neuropsychology*, 23(4), 583–605. doi:10.1080/02643290500310962
- Riddoch, M. J., Humphreys, G. W., Edwards, S., Baker, T., & Willson, K. (2003). Seeing the action: Neuropsychological evidence for action-

- based effects on object selection. *Nature Neuroscience*, 6(1), 82–89. doi:10.1038/nn984
- Riddoch, M. J., Pippard, B., Booth, L., Rickell, J., Summers, J., Brownson, A., & Humphreys, G. W. (2011). Effects of action relations on the configural coding between objects. *Journal of Experimental Psychology: Human Perception and Performance*, 37(2), 580–587. doi:10.1037/a0020745
- Roberts, K. L., & Humphreys, G. W. (2010). Action relationships concatenate representations of separate objects in the ventral visual system. *NeuroImage*, 52(4), 1541–1548. doi:10.1016/j.neuroimage.2010.05.044
- Roberts, K. L., & Humphreys, G. W. (2011). Action relations facilitate the identification of briefly-presented objects. *Attention, Perception, & Psychophysics*, 73(2), 597–612. doi:10.3758/s13414-010-0043-0
- Rousselet, G. A., Joubert, O. R., & Fabre-Thorpe, M. (2005). How long to get to the gist of real-world natural scenes? *Visual Cognition*, 12(6), 852–877. doi:10.1080/1350628044000553
- Scharff, A., Palmer, J., & Moore, C. M. (2011). Evidence of fixed capacity in visual object categorization. *Psychonomic Bulletin & Review*, 18(4), 713–721. doi:10.3758/s13423-011-0101-1
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime, version 2.0*. Pittsburgh: Psychology Software Tools, Inc.
- Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences*, 1(7), 261–267. doi:10.1016/S1364-6613(97)01080-2
- Theeuwes, J., Kramer, A. F., Hahn, S., & Irwin, D. E. (1998). Our eyes do not always go where we want them to go: Capture of the eyes by new objects. *Psychological Science*, 9(5), 379–385. doi:10.1111/1467-9280.00071
- Torralba, A., Oliva, A., Castelano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, 113(4), 766–786. doi:10.1037/0033-295X.113.4.766
- Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, 6(2), 171–178. doi:10.1016/S0959-4388(96)80070-5
- Treisman, A. (2006). How the deployment of attention determines what we see. *Visual Cognition*, 14(4–8), 411–443. doi:10.1080/13506280500195250
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97–136. doi:10.1016/0010-0285(80)90005-5
- Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruity influence eye movements when inspecting pictures. *The Quarterly Journal of Experimental Psychology*, 59(11), 1931–1949. doi:10.1080/17470210500416342
- Underwood, G., Templeman, E., Lamming, L., & Foulsham, T. (2008). Is attention necessary for object identification? Evidence from eye movements during the inspection of real-world scenes. *Consciousness and Cognition: An International Journal*, 17(1), 159–170. doi:10.1016/j.concog.2006.11.008
- VanRullen, R., Reddy, L., & Fei-Fei, L. (2005). Binding is a local problem for natural objects and scenes. *Vision Research*, 45(25–26), 3133–3144. doi:10.1016/j.visres.2005.05.012
- Võ, M. L.-H., & Henderson, J. M. (2009). Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *Journal of Vision*, 9(3), 1–15. doi:10.1167/9.3.24
- Võ, M. L.-H., & Henderson, J. M. (2011). Object-scene inconsistencies do not capture gaze: Evidence from the flash-preview moving-window paradigm. *Attention, Perception, & Psychophysics*, 73, 1742–1753. doi:10.3758/s13414-011-0150-6
- Walker, S., Stafford, P., & Davis, G. (2008). Ultra-rapid categorization requires visual attention: Scenes with multiple foreground objects. *Journal of Vision*, 8(4), 1–12. doi:10.1167/8.4.21
- Wolfe, J. M., & Cave, K. R. (1999). The psychophysical evidence for a binding problem in human vision. *Neuron*, 24(1), 11–17. doi:10.1016/S0896-6273(00)80818-1
- Wolfe, J. M., Võ, M. L. H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Sciences*, 15(2), 77–84. doi:10.1016/j.tics.2010.12.001

Authors Notes

Nurit Gronau and Meytal Shachar, Department of Psychology & Cognitive Science Studies, The Open University of Israel, Raanana, Israel.