Inter-object grouping in visual processing: How the brain uses real-world regularities to carve up the environment

PhD Thesis

handed in by

Daniel Kaiser

28th cycle cognitive neuroscience student

Center for Mind / Brain Sciences,

University of Trento

Supervisor:

Prof. Marius V. Peelen

CiMeC
Center for Mind/Brain Sciences
UNIVERSITY
OF TRENTO - Italy

Acknowledgements

Thanks to Marius & Timo.

Thanks to Elena Aggius Vella for her help in behavioral data collection.

All studies were funded by the Autonomous Province of Trento, Call "Grandi Progetti 2012", project "Characterizing and improving brain mechanisms of attention – ATTEND". Study 3 also received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement number 329363 and the German Research Foundation (grant STE 2239/1-1), awarded to Timo Stein.

Table of Contents

1 General Introduction

The human visual system has limited resources that confine parallel processing of information (Broadbent, 1958). This limited capacity becomes apparent in tasks where multiple objects need to be processed at the same time. Visual conjunction search is a prime example of such a task: If participants have to search for a particular feature conjunction (i.e., a horizontal red bar) in an array of distracters that each share one of these features (i.e., horizontal green bars and vertical red bars), a great reaction time cost can be observed when adding further distracters (Treisman & Gelade, 1980; Wolfe, Cave, & Franzel, 1990). Similar costs of processing multiple items at a time can be seen in visual working memory, where three to four items can be stored accurately, while performance drops rapidly when adding more items (Alvarez & Cavanagh, 2004; Cowan, 2001; Luck & Vogel, 1997). Such capacity limitations have been linked to the functional architecture of the visual system, where different stimuli have to compete for representation in specific neural circuits (Desimone & Duncan, 1995; Franconeri, Alvarez, & Cavanagh, 2013): When multiple stimuli are present at the same time (Kastner, De Weerd, Desimone, & Ungerleider, 1998; Miller et al., 1993; Moran & Desimone, 1985; Rolls & Tovee, 1995), and when these stimuli are recruiting more similar neural mechanisms (Cohen, Konkle, Rhee, Nakayama, & Alvarez, 2014), processing becomes less efficient. The degree to which representations of stimuli overlap has been connected to behavioral performance in capacity-limited tasks, such as visual working memory (Cohen et al., 2014), and is predictive of how well stimuli hinder each other from entering visual awareness (Cohen, Nakayama, Konkle, Stantic, & Alvarez, 2015). In sum, capacity limitations seem to be tightly linked to brain architecture, and they severely constrain performance in a range of behavioral tasks that are relevant in our everyday lives, such as visual search and visual working memory.

To better exploit this limited capacity, the brain can partly bypass processing limits by using recurring patterns in visual input. One efficient way of taking such pattern structures in the input into account is to perceptually group display items based on Gestalt formation (Wagemans et al., 2012; Wertheimer, 1923): Driven by low-level attributes such as spatial proximity, connectedness, or common movement, stimuli can be processed together (e.g., multiple items that move together can be processed as a perceptual group). Such grouping processes can reduce the total amount of information that needs to be processed – instead of processing individual items separately groups of items become the units of representation. Gestalt grouping has been shown to enhance performance in a range of behavioral tasks. In visual search, distracters that can be grouped are easier to reject, making the search process more efficient (Banks & Prinzmetal, 1976; Bundesen & Pedersen, 1983; Donnelly, Humphreys, & Riddoch, 1992; Humphreys, Quinlan, & Riddoch, 1989; Rauschenberger & Yantis, 2006; Roelfsema & Houtkamp, 2011; Wolfe & Bennett, 1997). For example, when distracters can be grouped by color, search is more efficient and can be predicted on the basis of the number of color groups rather than the number of single items (Bundesen & Pedersen, 1983). Similarly, target detection is facilitated when distracter stimuli form predictable contours (referred to as “good continuation” in Gestalt psychology; Banks & Prinzmetal, 1976). Such findings can be interpreted as a reduction of set size resulting in reduced inter-object competition (e.g., all items belonging to a contour can be processed together). Comparable benefits of grouping have been shown for capacity-limited visual working memory, where grouped items can be maintained more accurately (Anderson, Vogel, &

Awh, 2013; Peterson & Berryhill, 2013; Woodman, Vecera, & Luck, 2003; Xu, 2006; Xu & Chun, 2007). For example, when stimuli can be grouped based on color similarity, they are better remembered (Peterson & Berryhill, 2013). Similarly, grouping induced by the formation of illusionary contours (i.e., pacman-like inducers oriented to form an illusionary shape) enhances visual working memory (Anderson, Vogel, & Awh, 2013). Interestingly, grouping based on such low-level cues can also influence how stimuli enter visual awareness in the first place: Inducers oriented to form an illusionary contour preferentially reach awareness during continuous flash suppression (Wang, Weng, & He, 2012), and extinction patients display enhanced detection for items in their bad hemifield when these items are part of a perceptual group (Driver, 1995). On a neural level, competitive interactions have been shown to be reduced for stimuli that can be grouped based on Gestalt formation. When competing stimuli can be aggregated based on similarity, suppressive interactions in visual cortex are reduced (Beck & Kastner, 2005, 2007; Knierim & van Essen, 1992). Similarly, if stimuli are forming an illusionary figure, competitive interactions are diminished as compared to when stimuli cannot be grouped based on contour formation (McMains & Kastner, 2010, 2011). Altogether, these findings provide evidence that low-level grouping based on Gestalt laws is a powerful mechanism to reduce competition and as a consequence to overcome capacity limitations.

Although low-level grouping undoubtedly has a profound impact on visual capacity, this process seems much less valuable in real-world environments. While clearly low-level attributes in natural scenes are organized in specific, recurring ways (see Purves, Wojtach, & Lotto, 2012), most of our daily life tasks require perceptual processing (and subsequent acting) on the level of meaningful objects. Crucially, relevant stimuli very often are not tied to specific low-level properties (e.g., when trying to recognize a friend’s face in a crowd of people, low-level features won’t be very diagnostic for solving this task), and they can appear all over the visual field. Nonetheless, humans can maintain striking performance levels when faced with complex natural scenes (see Peelen & Kastner, 2014). For example, observers can efficiently search for arbitrary targets in cluttered scenes, as indicated by a relatively low cost of adding further distracters (Wolfe, Alvarez, Rosenholtz, Kuzmova, & Sherman, 2011a). Strikingly, detection of categorical targets (e.g., animals) is very efficient also when images are only flashed for very brief amounts of time (Thorpe, Fize, & Marlot, 1996) and even in the near absence of attention (Li, VanRullen, Koch, & Perona, 2002). These findings are intriguing because real-world environments are highly complex, and most scenes from our everyday lives contain dozens of different objects (Wolfe et al., 2011a). How does the brain achieve this remarkable efficiency in dealing with complex natural scenes?

Here we propose a novel grouping mechanism that operates at the level of meaningful everyday objects, which can contribute to more efficient scene parsing. Real-world scenes are often highly regular in the object arrangements they contain (Bar, 2004; Chun, 2000): In a typical dining room scene, we most often see a table that is surrounded by chairs and a lamp above the table. When seeing this typical dining room arrangement, the visual system might group the objects into a “dining group”, representing the group instead of its constituent objects. Crucially, similar to low-level grouping, this process would reduce the total number of items that need to be processed and therefore enhance performance in capacity-limited visual tasks. To test this idea, we used pairs of objects that commonly appear in a specific configuration (e.g., a lamp above a table, a mirror above a sink, or a camera on top of a tripod). These stimuli allowed us to test whether regularly configured pairs (e.g., lamp above table) are processed differently from irregularly positioned pairs (e.g., table above lamp). In three studies, we provide evidence that object regularities have a profound influence on visual processing. In Study 1, we show that objects that can be grouped based on real-world regularities are interfering less with the processing of other (behaviorally relevant) objects. Using fMRI, we demonstrate that objects that can be grouped based on such regularities interfere less with the representation of unrelated stimuli (houses). We relate this reduction of competition in visual cortex to a benefit in visual search: When grouped objects appear as distracters they are easier to reject, benefiting target detection performance. In Study 2, we tested whether this reduction of competition between objects groups is similarly helpful in visual working memory, which is also severely capacity-limited. Our results show that real-world regularities help to correctly hold objects in memory, demonstrating that this grouping mechanism is not only useful in perception, but also benefits working memory performance. In Study 3, we examined whether real-world regularities already influence visual processing at the stage of visual detection, that is, whether they influence if we consciously see something in the first place. Using continuous flash suppression, we demonstrate that regularly configured object pairs break through inter-ocular suppression faster than irregularly configured ones, indicating that the visual system is preferentially detecting stimuli that follow real-world regularities.

2 Study 1: Object grouping based on real-world regularities facilitates perception by reducing competitive interactions in visual cortex[1]

2.1 Introduction

In daily life, humans are confronted with complex and cluttered visual environments that contain a large amount of visual information. Because of the limited capacity of the visual system, not all of this information can be processed concurrently. Consequently, elements within a visual scene are competing for neural representation and cognitive processing resources (Desimone & Duncan, 1995; Kastner & Ungerleider, 2001). Such competitive interactions can be observed in neural responses when multiple stimuli are presented at the same time. Single-cell recordings in monkey visual cortex revealed that activity evoked by a neuron's preferred stimulus is suppressed when a non-preferred stimulus is simultaneously present in the neuron's receptive field (Miller, Gochin, & Gross, 1993; Moran & Desimone, 1985; Rolls & Tovee, 1995). Corresponding evidence for mutually suppressive interactions among competing stimuli has been obtained from human visual cortex using functional magnetic resonance imaging (fMRI) (Kastner, De Weerd, Desimone, & Ungerleider, 1998).

According to biased competition theory, these competitive interactions occur between objects rather than between the parts of a single object (Desimone & Duncan, 1995). This idea of object-based competition is supported by behavioral studies showing that judgments on two properties of one object are more accurate than judgments on the same properties distributed over two objects (Duncan, 1984). However, the degree

---

[1] This work has been published elsewhere: Stein, T., Kaiser, D., & Peelen, M. V. (2015). Object grouping based on real-world regularities facilitates perception by reducing competitive interactions in visual cortex. *Proceedings of the National Academy of Sciences USA, 111*, 11217–11222.

of competition among objects is strongly influenced by contextual factors, such as stimulus similarity (Beck & Kastner, 2005, 2007; Knierim & van Essen, 1992), geometric relationships between stimuli (Kapadia, Ito, Gilbert, & Westheimer, 1995), and perceptual grouping (McMains & Kastner, 2010, 2011). For example, competitive interactions in human visual cortex are greatly reduced when multiple single stimuli form an illusory contour and hence can be perceptually grouped into a single gestalt (McMains & Kastner, 2010).

Whereas the attentional benefit of grouping based on low-level cues is well established, much less is known about object grouping at more conceptual levels. Many objects in real-world scenes occupy regular and predictable locations relative to other objects. For example, a bathroom sink is typically seen together with a mirror in a highly regular spatial arrangement. When considering highly regular object pairs like these, it becomes clear that the world can be carved up at different levels: based on low-level cues such as those specified by gestalt laws, but also based on conceptual knowledge and long-term visual experience; a plate flanked by a fork and a knife is both a dinner plate set and three separate objects.

In the present fMRI and behavioral studies, we asked whether grouping based on real-world regularities modulates attentional competition. We hypothesized that objects that appear in frequently experienced configurations are, to some extent, grouped, resulting in reduced competition between these objects. To test this prediction, we presented pairs of common everyday objects either in their typical, regular configuration (e.g., a lamp above a table) or in an irregular configuration (e.g., a lamp below a table). Our findings indicate that grouping of objects based on real-world regularities effectively reduces the number of competing objects, leading to reduced neural competition and more efficient visual perception.

2.2 Results

2.2.1 fMRI Experiment

To measure competitive interactions between objects in human visual cortex, we followed the rationale of classical single-cell recording studies that indexed competition as the difference between neural activity evoked by a neuron's preferred stimulus presented in isolation and neural activity evoked by a neuron's preferred stimulus presented together with non-preferred stimuli (Miller et al., 1993; Moran & Desimone, 1985; Rolls & Tovee, 1995). The stronger the non-preferred stimuli compete for representation, the more the neuron’s response will be reduced. For example, an increase in the number of non-preferred stimuli would lead to a decrease in the response to (and representation of) the neuron’s preferred stimulus.

Because of the relatively poor spatial resolution of fMRI, the preferred stimulus in our study was the category houses, capitalizing on the finding that a region in the parahippocampal cortex (the parahippocampal place area, PPA; Figure 1.1c) responds preferentially to houses relative to other objects (Epstein & Kanwisher, 1998; Aguirre, Zarahn, & D’Esposito, 1998). To induce competition, the house stimuli were presented together with pairs of common everyday objects — non-preferred stimuli for the PPA. The pairs were presented either in their regular, commonly experienced configuration or in an irregular configuration, where pairs were vertically reversed (Figure 1.1a). Thus, displays with regular and irregular object pairs differed only with regard to the relative spatial position of the single objects within pairs, whereas all other stimulus aspects were identical. This allowed us to test for differences in neural competition as a function of the relative spatial positions of the objects: If regularly positioned objects are grouped, effectively reducing the number of competing non-preferred elements, they should compete less with houses than irregularly positioned objects. This would predict stronger PPA responses to houses presented together with the regular than with the irregular object pairs.

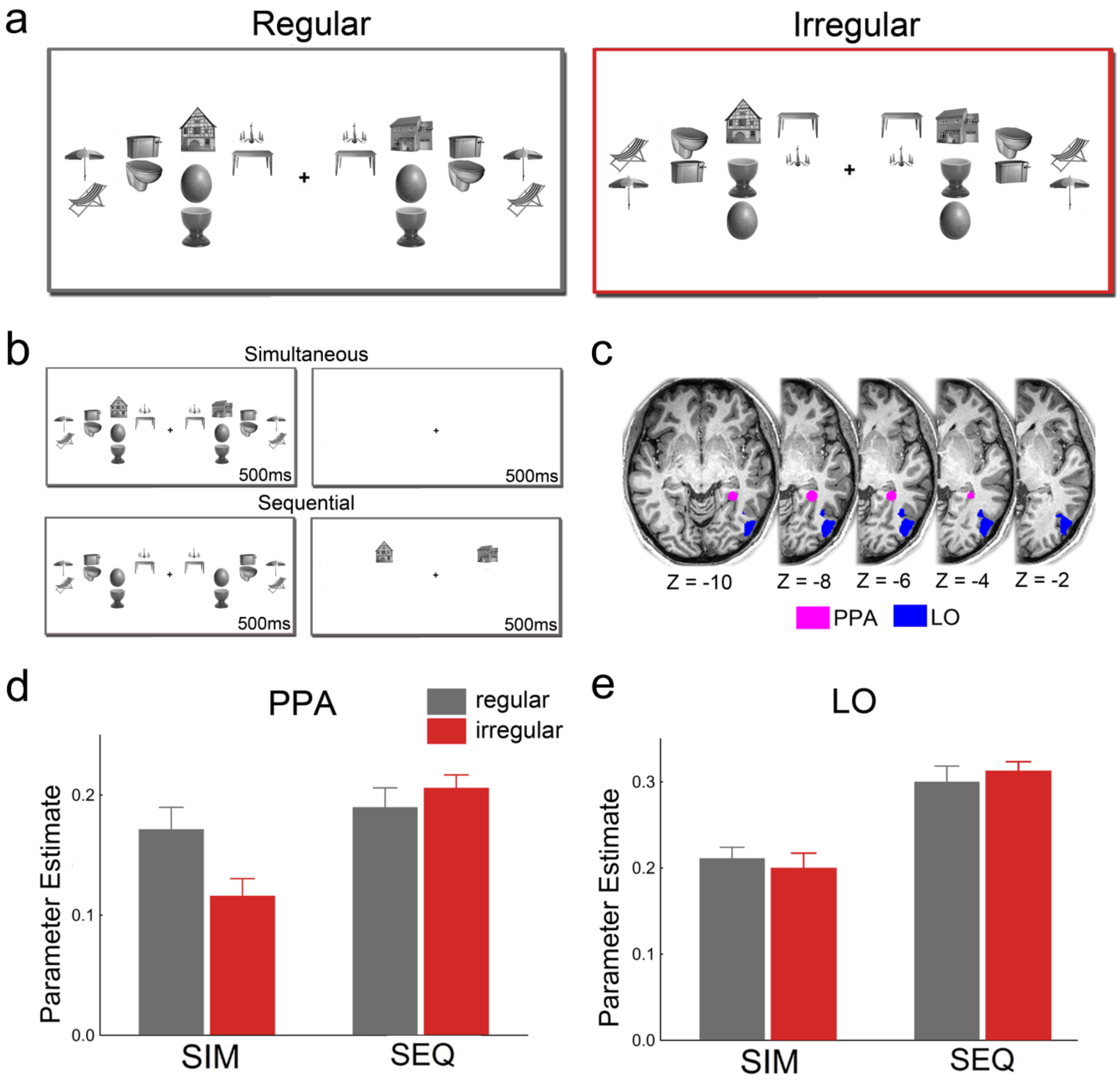

Figure 1.1. Increased house-evoked activity in PPA when simultaneously presented object distracters are positioned according to real-world regularities. (a) In each display, two house stimuli were surrounded by a total of eight object pairs. The configurations of the

objects were either regular (upper) or irregular (lower) relative to their real-world configurations. (b) Attentional competition was manipulated by either presenting the houses and the surrounding pairs simultaneously for 500 ms, followed by a 500-ms blank screen, or sequentially, for 500 ms each. (c) Location of right-hemispheric PPA and LO in a representative participant. (d) When houses and object pairs were presented simultaneously (SIM), house-selective PPA showed stronger responses when the object pairs were positioned according to real-world regularities than when they were not, indicating reduced attentional competition. No such difference was observed in the absence of competition between houses and objects — when houses and object pairs were presented sequentially (SEQ). (e) In contrast to house-selective PPA, responses in object-selective LO were not modulated by pair configuration.

Importantly, to ensure that response differences between the regular and irregular conditions reflected differences in attentional competition rather than differential responses to the regular and irregular object pairs themselves, we additionally included conditions in which the house and object stimuli were presented sequentially (Beck & Kastner, 2005; McMains & Kastner, 2010, 2011; Kastner et al., 1998). Competitive interactions among houses and object pairs are expected to occur in the simultaneous condition but not in the sequential condition (Beck & Kastner, 2005; McMains & Kastner, 2010, 2011; Kastner et al., 1998). By including the sequential condition, we controlled for possible differences in the responses evoked by the regular and irregular object pairs themselves: The object arrays presented in the simultaneous and sequential conditions are identical. We designed the experiment in this way because our interest was in the competition between the object arrays (non-preferred stimuli for PPA) and the houses (preferred stimuli for PPA), rather than in differences between regular and irregular object pairs themselves. The critical test, therefore, is the interaction between presentation order (simultaneous, sequential) and pair configuration (regular, irregular). While viewing the displays, participants were engaged in a fixation task that was unrelated to the house and object stimuli.

Results showed that activity in functionally defined PPA was stronger to houses presented together with regularly positioned objects than to houses presented together with irregularly positioned objects, $t(22) = 2.24$, $p = .035$, indicating reduced competition from regularly positioned objects (Figure 1.1d). Importantly, in the absence of competitive interactions among houses and objects — when the house and object displays were presented sequentially (Figure 1.1b) — no difference between the object conditions was observed, $t(22) = 0.96$, $p = .35$; presentation order × pair configuration interaction, $F(1,22) = 6.35$, $p = .019$ (Figure 1.1d). This indicates that the differential PPA responses in the simultaneous presentation condition reflected differences in competitive interactions between houses and objects rather than differential responses to regular and irregular object arrays themselves. These results generalized to alternative PPA definition procedures (see Supplementary Material) and were also obtained in analyses of event-related time course data (see Supplementary Material).

These results were specific to the PPA. Responses in object-selective lateral occipital cortex (LO; Figure 1.1c) were generally lower in the simultaneous condition than in the sequential condition, $F(1,22) = 27.17$, $p < .001$. Unlike in the PPA, however, the competition effect was not modulated by the configuration of the object pairs (presentation order × pair configuration interaction, $F(1,22) = 1.16$, $p = .29$, Figure 1.1e). The activation pattern observed in LO was significantly different from the pattern in PPA (three-way interaction including region, $F(1,22) = 5.59$, $p = .027$). Additional face-selective control regions showed the same pattern of results as LO (see Supplementary Material).

Together, these fMRI results indicate that competitive interactions between preferred (houses in PPA) and non-preferred (objects in PPA) stimuli are reduced when objects are positioned according to real-world regularities.

2.2.2 Visual Search Experiments

To test whether the reduced competition observed at the neural level leads to behavioral facilitation, we modified the displays of the fMRI experiment for use in a behavioral paradigm aimed at measuring accuracy of visual perception. In a series of visual search experiments (Figure 1.2a), participants located single target objects surrounded by pairs of distracter objects (Figure 1.2b), which were positioned in either their regular or irregular configuration. We used the same object arrays that were used in the fMRI experiment, but replaced the house stimuli with uniquely nameable everyday objects as targets (see Methods). The search displays were presented briefly (200 ms) and participants were instructed to indicate as accurately as possible whether the target object appeared to the left or to the right of fixation.

Accuracy in localizing the target object was higher when the pairs of distracter objects were presented in their regular configurations relative to when they were presented in irregular configurations $t(16) = 2.88$, $p = .011$, with no difference in response times, $t(16) = 0.70$, $p = .50$ (Figure 1.2c). This suggests that distracters positioned according to real-world regularities are more efficiently processed, leaving more resources for target detection in visual search. These results were replicated in a response-time-based version of this experiment, showing shallower search slopes for distracter pairs presented in regular compared with irregular configurations (see Supplementary Material).

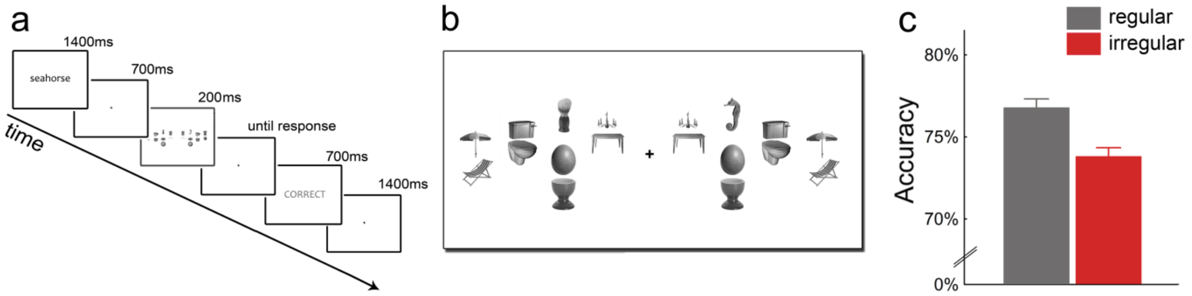

Figure 1.2. Enhanced visual search performance when distracters are positioned according to real-world regularities. (a) The visual search paradigm consisted of a word cue that corresponded to a single target object. Participants indicated whether the target was on the right or on the left side of a briefly presented cluttered visual display. (b) Search arrays were the same displays as in the fMRI experiment, but houses were replaced by the search target on one side and a single distracter on the other side. Again, all distracter pairs could be presented in a regular or irregular configuration. (c) Regular distracter pairs led to higher accuracy than irregular pairs.

To control for potential low-level differences between regular and irregular object pairs, in a second experiment we added conditions with inverted distracter pairs (Figure 1.3a). Inversion preserves all low-level differences between regular and irregular conditions but disrupts higher-level grouping. The benefit of regularly positioned distracters was again found for upright displays, $t(13) = 4.49$, $p = .006$, but, crucially, not for inverted displays, $t(13) = 0.071$, $p = .94$; interaction, $F(1,13) = 5.57$, $p = .035$ (Figure 1.3b). Response times did not differ significantly between conditions, all$F(1,13) < 0.40$, $p > .50$). This rules out the possibility that low-level visual differences between the distracter arrays accounted for the effect.

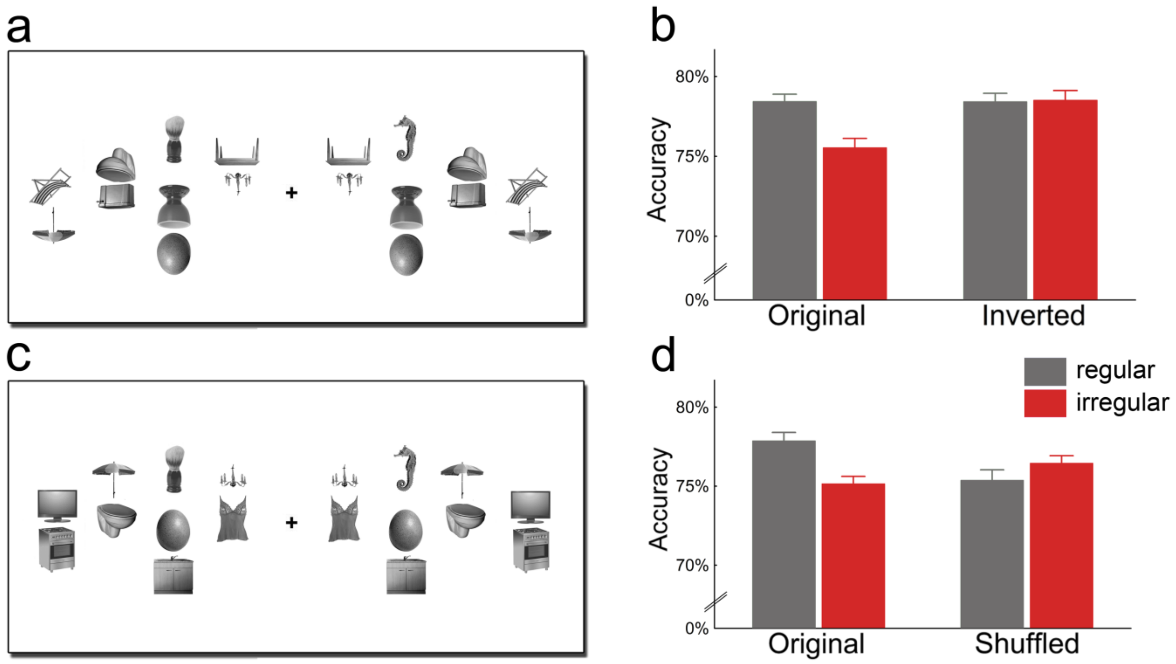

Figure 1.3. Improved detection of targets among regularly positioned distracters cannot be explained by low-level grouping: When the distracter pairs were inverted (a), regular and irregular distracters led to comparable target detection accuracy (b). Also the relative position of single objects cannot account for the effect: When the top objects were interchanged between pairs (shuffled condition, c), the accuracy benefit for regular configurations disappeared (d).Original conditions are independent replications of the first experiment.

For a second control experiment, we generated new object pairs by shuffling the top objects of the original regular and irregular pairs (Figure 1.3c). These shuffled pairs did not follow real-world regularities, but the specific locations at which single objects were presented were identical to the original pairs. Results again showed a benefit for regularly relative to irregularly positioned distracter pairs, $t(17) = 2.96$, $p = .009$, but no corresponding benefit for the shuffled pairs, $t(17) = 0.85$, $p = .41$; interaction, $F(1,17) = 5.63$, $p = .030$ (Figure 1.3d). Again, response times did not differ between conditions, all *F*(1,17) < 1.40, *p* > .20. Thus, the specific position of single objects is not sufficient to explain the effect.

Together, these visual search experiments demonstrate improved perception of target objects when distracter objects are positioned according to real-world regularities, thus providing behavioral evidence for reduced competition from regularly positioned distracters.

2.3 Discussion

Our findings demonstrate that the visual system exploits learned regularities to perceptually group objects that typically co-occur in specific configurations. Through this process, the effective number of objects that compete for representation is reduced. These findings have implications for attentional selection in real-world situations where multiple, but often regularly positioned, distracter objects compete for visual representation.

Previous studies have demonstrated that contextual factors can reduce competitive interactions among simple, artificial stimuli that were perceptually grouped based on physical similarity, geometric relationships, or gestalt principles (McMains & Kastner, 2010, 2011). Distracters that can be grouped based on such low-level cues can be rejected at once rather than on an item-by-item basis, leading to enhanced target detection (Banks & Prinzmetal, 1976; Bundesen & Pedersen, 1983; Donnelly, Humphreys, & Riddoch, 1992; Humphreys, Quinlan, & Riddoch, 1989; Rauschenberger & Yantis, 2006; Roelfsema & Houtkamp, 2011; Wolfe & Bennett, 1997). For example, when distracters can be grouped by color, search performance depends on the number of distracter groups rather than on the number of individual distracters in each group (Bundesen & Pedersen, 1983). Our results show that benefits of grouping are not limited to grouping based on low-level cues, but that these can also be observed for grouping based on knowledge about the typical spatial relations between objects in our visual environment.

The present way of measuring neural competition closely resembles the logic of monkey electrophysiology work on attentional competition (Miller et al., 1993; Moran & Desimone, 1985; Rolls & Tovee, 1995), in that we recorded neural activity to a region's preferred stimuli in the presence of competing non-preferred stimuli. Reduced neural competition from non-preferred stimuli was reflected in an increased PPA response to the region's preferred house stimuli when the PPA's non-preferred object stimuli could be grouped based on real-world regularities. The sequential presentation condition, in which houses and objects did not compete for attention, provided an important control, showing that the increased PPA response was not driven by response differences between the regular and irregular object pairs themselves.

This raises the interesting question of whether there are brain regions that differentially respond to regular and irregular object pairs. None of our regions of interest (ROIs) showed such a difference, and no regions were found in a whole-brain analysis testing for the main effect of pair configuration (see Supplementary Material). Previous work that tested for response differences as a function of action relations between objects (e.g., a hammer positioned to hit a nail) provided evidence for greater LO activity to interacting objects than to non-interacting objects (Kim & Biederman, 2011; Roberts & Humphreys, 2010). Patient and transcranial magnetic stimulation studies further showed that action relationships are processed independently of attentional influences from parietal cortex (Kim, Biederman, & Juan, 2011; Riddoch, Humphreys, Edwards, Baker, & Willson, 2003). Together with the absence of grouping effects in LO in the current study, these previous findings suggest a special status of object grouping based on action relations (Riddoch et al., 2003). Future studies are needed to test this notion, directly comparing effects of grouping based on real-world regularities, action cues (Kim & Biederman, 2011; Roberts & Humphreys, 2010), and more basic perceptual cues (Altmann, Bülthoff, & Kourtzi, 2003; Fang, Kersten, & Murray, 2008; Murray, Kersten, Olshausen, Schrater, & Woods, 2002; Kim & Biederman, 2012).

Beneficial effects of grouping are not limited to object perception and attentional competition but have also been observed in studies of visual working memory (VWM). Similar to its effects on attention, low-level grouping has been shown to enhance VWM capacity (Woodman, Vecera, & Luck, 2003; Xu, 2006). Recent studies have started to investigate VWM grouping based on statistical regularities in relative stimulus positions (Brady & Tenenbaum, 2013): Stimuli that appeared in regular combinations were better remembered (Brady, Konkle, & Alvarez, 2009; Olson, Jiang, & Moore, 2005), as if they had been compressed into a single VWM representation. An interesting avenue for future study will be to test whether VWM capacity is similarly enhanced for real-world object pairs like those used here, as suggested by accurate memory for objects in natural scenes (Hollingworth, 2006).

The reduced competition from regularly positioned objects demonstrated in the present study may constitute a powerful neural and perceptual mechanism to contend with the multitude of visual information contained in real-world scenes. The present findings could thus contribute to the understanding of perceptual efficiency in real-world scenes: Target detection in natural scenes is surprisingly efficient considering the large number of distracter objects present in real-world environments (Wolfe, Alvarez, Rosenholtz, Kuzmova, & Sherman, 2011a). As an explanation for this efficiency, it has been proposed that scene context guides attention to likely target locations (Chun, 2000; Wolfe, Võ, Evans, & Greene, 2011b). For example, we look above the sink when searching for a mirror. Such contextual guidance can stem from implicit or explicit memory for specific target locations within a specific context (Brockmole & Henderson, 2006; Chun, 2000; Wolfe et al., 2011b), global scene properties (Neider & Zelinski, 2006; Torralba, Oliva, Castelhano, & Henderson, 2006), and also from relations between target and non-target objects (Bar, 2004; Võ & Wolfe, 2013). At a general level, the current results might similarly reflect the learning of real-world correlational structure. However, our study differs from previous work in that it addressed the grouping of distracter objects independently of their role in guiding attention toward the search target, as the targets were completely unrelated to the distracters. Thus, such high-level grouping of objects forms an additional mechanism likely to support efficient target detection in cluttered real-world environments. Future studies are needed to extend our findings to attentional selection in real-world scenes. Because scenes contain a large number of objects that occur in regularly positioned groups of two or more objects, grouping of items according to real-world regularities might operate on many objects at the same time to greatly enhance the efficiency of real-world perception.

2.4 Methods

2.4.1 fMRI Experiment

Participants. Twenty-five participants (8 male, mean age 25.5 years, *SD* = 4.9) took part in the experiment. All procedures were carried out in accordance with the Declaration of Helsinki and were approved by the ethical committee of the University of Trento. Two participants were excluded from all analyses: one due to excessive head movement, and one because we were unable to define functional regions of interest at the adopted statistical threshold.

Stimuli. The stimulus set consisted of 12 object pairs of everyday objects with a typical spatial configuration in the vertical direction, such as a lamp above a dining table, a mirror above a bathroom sink, or an air vent above a stove. The pairs could be placed in their typical configuration (regular condition) or vertically interchanged (irregular condition). For each single object, two different exemplars were collected, resulting in four different exemplar combinations for each pair, and thus a total of 48 regular and 48 irregular pairs. Additionally, 36 images of houses were used. Each display contained four different object pairs and a house on each side of fixation. The pairs on the right side of fixation were always the perfect mirror image of the pairs left of fixation, whereas the house's position was mirrored but two different house exemplars were presented on each side. Single objects subtended a visual angle of ~1.5°. For each side, objects were placed in a jittered 4 × 4 grid, with the house stimulus always appearing in one of the four central locations of the grid (i.e., second or third row and second or third column). The nearest objects to fixation appeared with a horizontal offset of 2°. To control for interdisplay variability, each particular display (i.e., each particular combination of exemplars and positions) was used once in each condition. Stimuli were presented using the Psychtoolbox (Brainard, 1997) and projected on a translucent screen at the end of the scanner bore. Participants viewed the screen through a pair of tiled mirrors mounted on the head coil.

Main Experiment Procedure. Attentional competition was manipulated using an event-related variant of the sequential/simultaneous paradigm (Kastner et al., 1998). In the simultaneous condition, the whole display was presented for 500 ms, followed by a blank period of 500 ms. In the sequential condition, the house stimuli and the pair stimuli were presented in direct succession for 500 ms each (with the house appearing first in half of the trials and the surrounding pairs appearing first in the other half). Trials were separated by a 1,500-ms intertrial interval. Thus, the stimulation summed over a trial was the same in both conditions. However, whereas the simultaneous presentation of the house stimuli and the surrounding object stimuli was expected to induce competitive interactions, no such competition should be present when the stimuli were presented sequentially (Kastner et al., 1998). Importantly, we manipulated the regularity of the object pairs: In the regular condition, all pairs were presented in their typical configuration (e.g., lamp above a dining table), whereas in the irregular condition, all pairs were presented with individual object positions interchanged (e.g., lamp below a dining table). The resulting four conditions were randomly intermixed within each run. There was a total of eight runs, each lasting approximately 5 min and consisting of 120 trials, of which 20% were fixation-only trials. Participants were instructed to maintain fixation at a central cross throughout the experiment and to respond to small size changes of the fixation cross (size increases of ~15%). Participants detected the changes with high accuracy (92.3% correct, *SE* = 0.9%), and there were no significant differences between conditions, presentation order × pair configuration ANOVA, all $F(1,22) < 0.31$, $p > .55$. Similarly, response times did not differ between conditions, all $F(1,22) < 2.81$, $p > .10$.

Functional Localizer Procedure. In addition to the eight experimental runs, participants completed two functional localizer runs of 5 min each. Participants performed a one-back task while viewing images of faces, houses, everyday objects (different exemplars than in the main experiment), and scrambled objects. Each stimulus category included 36 individual exemplars. Within each run, there were four blocks of each stimulus category and four blocks of fixation baseline, with all blocks lasting 16 s. Block order was randomized for the first 10 blocks and then mirror reversed for the remaining 10 blocks. Each non-fixation block included two one-back stimulus repetitions. To find the maximally selective voxels for the house stimuli, we used the same house exemplars as in the main experiment.

fMRI Data Acquisition. Imaging was conducted on a Bruker BioSpin MedSpec 4T head scanner (Bruker BioSpin), equipped with an eight-channel head coil. T2*-weighted gradient-echo echo-planar images were collected as functional volumes for the main experimental runs and the functional localizer runs (repetition time = 2.0 s, echo time = 33 ms, 73° flip angle, 3 × 3 × 3 mm voxel size, 1-mm gap, 34 slices, 192 mm field of view, 64 × 64 matrix size). A T1-weighted image (MPRAGE; 1 × 1 × 1 mm voxel size) was obtained as a high-resolution anatomical reference.

fMRI Preprocessing. All neuroimaging data were analyzed using MATLAB and SPM8. The volumes were realigned, coregistered to the structural image, resampled to a 2 × 2 × 2 mm grid and spatially normalized to the Montreal Neurological Institute (MNI)-305 template (as included in SPM8). Functional volumes were then smoothed using a 6-mm full-width half-maximum Gaussian kernel. All analyses were performed on the smoothed data.

fMRI Data Analysis — Functional Localizer. The blood-oxygen-level–dependent (BOLD) signal of each voxel in each participant in the localizer runs was modeled using four regressors, one for each stimulus category (faces, houses, objects, and scrambled objects), and six regressors for the movement parameters obtained from the realignment procedure. Functional ROIs were defined in individual participants using tcontrasts. House-selective PPA (Epstein & Kanwisher, 1998; Aguirre et al., 1998) was localized with the houses > objects contrast. Object-selective LO (Malach et al., 1995) was localized with the objects > scrambled contrast. Bilateral LO and PPA ROIs were defined as spheres of 4-mm radius (including 33 voxels) around the peak MNI coordinates of activation [left PPA: x = −22.1 (1.4), y = −43.4 (2.1), z = −7.9 (1.2); right PPA: x = 25.0 (1.3), y = −45.1 (1.4), z = −6.5 (1.0); left LO: x = −45.9 (1.0), y = −79.0 (1.1), z = −3.0 (2.0); right LO: x = 46.6 (1.1), y = −77.0 (1.1), z = −5.0 (1.8); *SE*s in parentheses], with the threshold set at $p < .01$, uncorrected. We chose these relatively small spherical ROIs to maximize selectivity in PPA (the Supplementary Material shows results in the peak voxel of PPA).

fMRI Data Analysis — Main Experiment. For the main experiment, the BOLD signal of each voxel in each participant was estimated with 11 regressors in a general linear model: 4 regressors for the experimental conditions, 1 regressor for the fixation-only trials, and 6 regressors for the movement parameters obtained from the realignment procedure. All models included an intrinsic temporal high-pass filter of 1/128 Hz to correct for slow scanner drifts. ROI analysis was done using the MARSBAR toolbox for SPM8 (Brett, Anton, Valabregue, & Poline, 2002). For each ROI and each hemisphere, we estimated response magnitudes from the generalized linear model beta values of the conditions of interest relative to the beta values of the fixation-only trials. For each ROI, responses were then averaged across hemispheres.

2.4.2 Visual Search Experiments

Participants. Eighteen participants (6 male, mean age 22.7 years, SD = 2.5) volunteered for behavioral Experiment 1, 13 participants (2 male; mean age 22.2 years, *SD* = 2.9) for Experiment 2, and 18 participants (1 male; mean age 22.8 years, *SD* = 2.2) for Experiment 3.

Stimuli. We used the same displays as in the fMRI experiment, but replaced the houses with a single target object on one side of the display and a single non-target object on the other side. For this purpose, an additional 100 uniquely nameable everyday objects were collected (taken from an online database; Brady, Konkle, Alvarez, & Oliva, 2008).

Procedure. In each trial, participants localized a single target object presented in the left or in the right hemifield. Each trial started with a word (e.g., “seahorse”) displayed for 1,400 ms, indicating the object participants had to localize. After 700 ms, a search array was displayed for 200 ms. Each array contained four different object pairs and one single object on each side of fixation. The pairs on the right of fixation were always the perfect mirror image of the pairs left of fixation, whereas the single objects’ positions were mirrored but the single objects (i.e., target and non-target object) differed between the two sides. One of the single objects was always the target item, with the target position (left versus right) randomly varying, whereas the overall probability for each side was fixed at 50%. Each single object appeared equally often in each condition as a target or a non-target, with no specific target–non-target pair being repeated multiple times throughout the experiment. To control for the variability between displays, each particular distracter array was shown once in each condition (i.e., each particular combination of distracter pairs and their positions). Participants used the left and right arrow keys on a keyboard to indicate as accurately as possible, without speed pressure, on which side the target object had appeared. After entering their response, participants received feedback. Trials were separated by an intertrial interval of 1,400 ms. The experiments were divided into blocks of 50 trials. The order of the first half of blocks was counterbalanced between subjects, and the order of the second half was generated by mirror reversing this order. In each block, the object pairs appeared in either the regular or the irregular configuration. In Experiment 1, participants completed eight blocks of the task. In Experiment 2, we exactly replicated Experiment 1, and additionally included blocks with inverted object pairs, in which all distracter pairs were presented upside down (i.e., rotated by 180°), whereas the single objects appeared in normal orientation. This inverted condition was included to control for the potential influence of low-level grouping effects, as inversion disrupts the object pairs' configuration, although all low-level properties are identical to the original upright pairs. In Experiment 3, blocks with "shuffled" pairs were included, in which the top and bottom items of the pairs were recombined into new pairs. These shuffled pairs (e.g., computer screen above stove) did not form typical spatial configurations, whereas the actual position of individual objects was identical to the original upright pairs. Thus, the inclusion of this shuffled condition allowed us to control for the potential influence of the actual position of single objects within pairs.

3 Study 2: Real-world regularities affect visual working memory for objects[2]

3.1 Introduction

Only a fraction of the massive visual input that humans face in virtually every real-life situation can be selected and used for further cognitive operations (Broadbent, 1958). As the visual input itself is constantly changing (e.g., due to eye movements), keeping past percepts in visual working memory (VWM; Baddeley, 1986) is essential for the use of this information in subsequent cognitive operations. VWM capacity is known to be extremely limited: for example, memory performance for 3-4 simple colored shapes is relatively high, but performance drops when more of these items are added to the display (Alvarez & Cavanagh, 2004; Cowan, 2001; Luck & Vogel, 1997).

However, these capacity limitations can be partly overcome by exploiting certain regularities in the sensory input (often referred to as *chunking* in the literature on verbal working memory; Chase & Simon, 1973; Cowan, 2001; Miller, 1956; Simon, 1974): When visual stimuli form regular and predictable ensembles, they can be grouped into larger unitary representations, leading to increased memory performance (Brady & Tenenbaum, 2013). For example, VWM is enhanced when individual stimuli can be grouped by Gestalt principles or by forming illusory contours (Anderson, Vogel, & Awh, 2013; Peterson & Berryhill, 2013; Woodman, Vecera, & Luck, 2003; Xu, 2006; Xu & Chun, 2007). Similarly, VWM is enhanced when participants learn to associate stimuli through arbitrary spatial contingencies (Brady, Konkle, & Alvarez, 2009; Olson, Jiang, & Moore, 2005). For example, Brady et al. (2009) found enhanced VWM performance for displays of two-colored disks when the displays contained disks with learned, predictable color

---

[2] This work has been published elsewhere: Kaiser, D., Stein, T., & Peelen, M. V. (2015). Real-world regularities affect visual working memory for objects. *Psychonomic Bulletin & Review*.

combinations. Thus, as the color conjunctions got highly experienced, VWM performance increased, as if the two colors had been grouped into a single VWM representation. Together, these findings show that stimulus regularities can be exploited to form more efficient VWM representations.

Most studies investigating the influence of stimulus regularities on VWM capacity have used simple stimuli such as colored disks. Importantly, our everyday environments contain many statistical regularities that might similarly be exploited for efficient cognitive processing. For example, the multitude of complex objects in real-world scenes commonly appear in regular and predictable locations relative to other objects. Taking such regularities into account may allow for more effective information processing (Bar, 2004; Chun, 2000). However, only few studies have investigated visual memory for naturalistic objects in real-world context. Early work suggested that memory for spatial relations among objects is better when these objects are embedded in meaningful scenes (Mandler & Johnson, 1976). More recent evidence indicates that object representations are bound to spatial locations within scenes (or object arrays; Hollingworth 2006, 2007), thereby helping to generate elaborate episodic representations of visual scenes (Hollingworth & Henderson, 2002). These findings suggest that rather than being stored independently of each other, objects are stored in memory in relation to their environment.

Whereas little is known about the effect of real-world regularities among naturalistic objects on VWM, a larger body of research has studied how object regularities influence visual perception. Some studies have explicitly focused on the relations among pairs of objects, comparing regular spatial arrangements (e.g., a hammer ready to hit a nail) to irregular configurations (e.g., a nail being positioned at the wrong end of a hammer). These studies have demonstrated more efficient visual perception for objects that are regularly positioned (Green & Hummel, 2006; Gronau & Shachar, 2014; Riddoch, Humphreys, Edwards, Baker, & Willson, 2003). These perceptual consequences of real-world regularities among objects may be related to differences in the encoding of regular and irregular object configurations in visual cortex (Gronau, Neta, & Bar, 2008; Kim & Biederman, 2011; Roberts & Humphreys, 2010). Indeed, we have recently found that when pairs of objects are positioned according to typically experienced, regular spatial configurations attentional competition among individual objects is reduced both on a behavioral and neural level (Kaiser, Stein, & Peelen, 2014a). This indicates that objects appearing in such regular spatial configurations can be grouped to reduce the number of competing stimuli and to allow for more efficient processing.

Object grouping based on real-world regularities may not only facilitate visual perception, but could also represent a powerful mechanism to overcome the capacity limitations of VWM. To test this hypothesis, we measured VWM performance in a delayed change-detection task, where participants had to memorize multiple objects that were presented in pairs. The objects of each pair were either placed in their typical, regular configurations (e.g., a lamp over a dining table or a mirror above a bathroom sink), or in irregular configurations, with their positions interchanged. We hypothesized that the visual system groups objects based on their typical real-world configurations, leading to enhanced VWM performance for regularly positioned objects.

3.2 Experiment 1

Experiment 1 investigated whether real-world regularities enhance VWM performance. To do so, we compared performance in a delayed change-detection task between object pairs presented in their typical, regular configuration with a condition in which this regularity was disrupted by interchanging object positions (Figure 2.1).

3.2.1 Methods

Participants. Thirty-eight healthy adults (7 male, mean age 23.6 years, *SD* = 4.6) participated. All of them had normal or corrected to normal vision and received money or course credits for their participation.

Stimuli and Apparatus. We used a set of 12 object pairs of everyday-objects, which have a typical spatial configuration in the vertical direction[3]. To manipulate regularity, pairs could be placed in their typical configuration ("regular" condition) or vertically reversed ("irregular" condition). For each single object category, we collected two different exemplars, leading to four possible pairs per category. All images were matched for luminance and contrast using the SHINE toolbox (Willenbockel et al., 2010). Single objects subtended a visual angle of approximately 3°. Images were displayed on a 17'' CRT monitor (1024×768 resolution, 75Hz refresh rate). Stimulus presentation was controlled using the Psychtoolbox (Brainard, 1997).

Procedure. Participants performed a VWM task with concurrent verbal suppression (e.g., Luck & Vogel, 1997; Jackson & Raymond, 2008; Figure 2.1a). At the beginning of each trial, a string of five digits was presented for 1400 ms. Subjects had to rehearse these digits aloud throughout the whole trial. After the digit presentation,

---

[3] Effects were not significantly modulated by the presence of specific individual object pairs (see Supplementary Material).

followed by a 1000 ms blank interval, a display of two (left and right of fixation) or three (triangular around fixation) object pairs was shown for 2000 ms (Figure 2.1b).

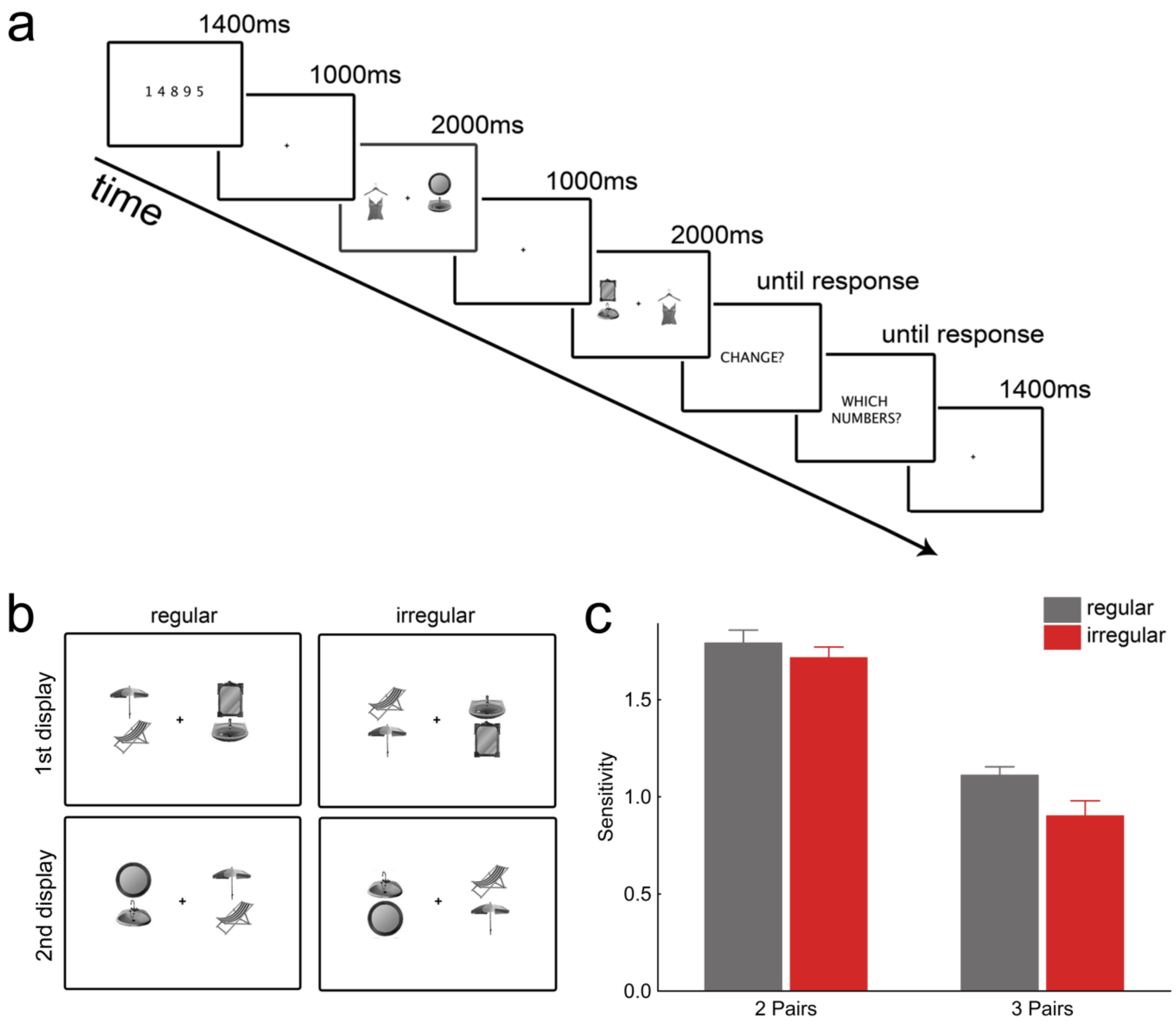

Figure 2.1. Enhanced visual working memory performance for objects that are positioned according to real-world regularities, relative to irregularly positioned objects. a) A single trial sequence. Participants had to rehearse five digits aloud while performing the change detection task. b) Example displays of change trials from the two-pairs condition in Experiment 1. Pair positions were exchanged on every trial and changes were always exemplar-level changes (e.g., one mirror changes to another mirror) that included both objects of the changing pair. c) Regular object pairs led to higher change-detection sensitivity than irregular pairs. Standard errors reflect within-subject SEM (Cousineau, 2005).

All of the groups in a display were either configured in a regular or irregular way. After a retention interval of 1000 ms, the display appeared again for 2000 ms. In the second display, all pairs appeared at different locations than on the first presentation. This was done to prevent same-different decisions from being made on the basis of shape-outline differences only. On 50% of the trials, the second display contained the same objects as the first one. On the other 50% of trials, one of the object pairs was changed: Changes were exemplar-level changes (e.g., a lamp changing into another lamp and a table changing into another table), and always both objects of the pair changed (see Figure 2.1b). After the second display, subjects were first required to report if there was a change in the objects and then they had to type in the digits of the verbal suppression task. Participants were informed that both responses should be made as accurately as possible, without speed pressure. The experiment consisted of a total of 192 trials, with the two set size conditions (two versus three object pairs) and the two configuration conditions (regular vs. irregular object pairs) randomly intermixed, leading to 48 trials (24 change trials, 24 no-change trials) per condition.

3.2.2 Results and Discussion

To test the influence of pair configuration on VWM performance, we computed *d*-prime scores [$d' = Z(\text{hit rate}) - Z(\text{false alarm rate})$] as a measure of change detection sensitivity. Trials with incorrect responses in the verbal suppression task (*M* = 6.1%, *SD* = 7.2) were excluded. A two-way repeated measures ANOVA on mean *d*-prime scores, with the factors pair configuration (regular versus irregular object pairs) and set size (two versus three object pairs) revealed a main effect of pair configuration, with higher change-detection sensitivity for regularly configured pairs than for irregularly configured pairs, $F(1,37) = 4.70$, $p = .037$ (Figure 2.1c), and a main effect of set size, reflecting better performance in the smaller set size condition than the larger set size condition, $F(1,37) = 82.12$, $p < .001$. The interaction was not significant, $F(1,37) = 1.90$, $p = .185$. Importantly, these differences in change detection sensitivity were not related to performance differences in the verbal suppression task: A two-way ANOVA on mean accuracies in the verbal suppression task did not reveal any significant main effects or interaction, all $F(1,37) < 1.04$, all $p > .316$. Thus, verbal memory strategies cannot account for the effect of pair configuration on VWM. These results demonstrate that VWM is enhanced when pairs of objects are positioned according to commonly experienced, regular configurations, as compared to pairs in which this configuration is disrupted.

3.3 Experiment 2

The results from Experiment 1 indicate that VWM is influenced by real-world regularities. However, this effect could partly reflect low-level Gestalt grouping, which is known to enhance VWM (Anderson et al., 2013; Peterson & Berryhill, 2013; Woodman et al., 2003; Xu, 2006; Xu & Chun, 2007). Although we carefully selected our stimuli to avoid that regular and irregular object pairs differed along low-level dimensions, in Experiment 2 we included a control condition to directly rule out any influence of such putative low-level differences. For this control condition all object pairs were inverted, that is, rotated by 180 degrees. Inversion disrupts the typical object configuration, while preserving all low-level stimulus properties. Furthermore, although the abstract spatial relationship among objects is unaffected by inversion, the pairs no longer follow typical real-world viewing conditions. Thus, if the VWM effect in Experiment 1 reflected the impact of real-world regularities rather than low-level grouping, inversion should abolish the effect.

3.3.1 Methods

Participants. Thirty-eight healthy adults (10 male, mean age 24.4 years, *SD* = 4.5) participated, of which eleven had also participated in Experiment 1.

Stimuli and Apparatus. The apparatus, stimuli, and setup were identical to Experiment 1. In addition to varying the pair configuration, we also manipulated orientation by presenting the original pairs or inverted versions, where the same pairs were presented upside-down (Figure 2.2a).

Procedure. We used the same design as in Experiment 1, except that we now used only displays with three object pairs and added the inverted condition. This again led to a total of 192 trials (48 per condition), with the two configuration (regular versus irregular) and the two orientation (original versus inverted) conditions being randomly intermixed.

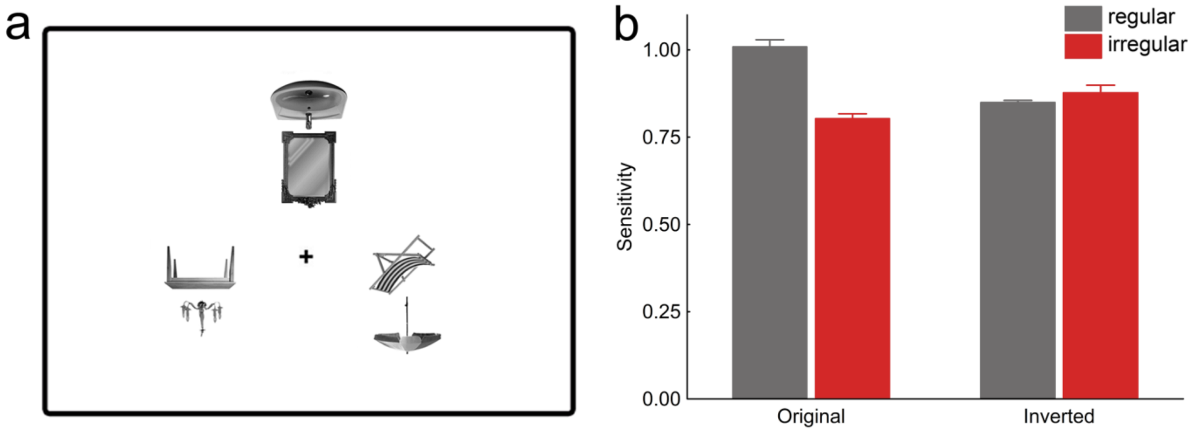

Figure 2.2. The VWM regularity effect is abolished for inverted object pairs. a) An example display from the regular inverted condition. b) While in Experiment 2 the results from Experiment 1 were replicated ("original" condition), there was no effect of pair configuration in the inverted condition. Standard errors reflect within-subject SEM (Cousineau, 2005).

3.3.2 Results and Discussion

Trials with incorrect responses in the verbal suppression task (6.1%, *SD*=5.2) were excluded from the analysis. A two-way repeated-measures ANOVA on mean *d*-prime scores, with the factors pair configuration (regular versus irregular) and orientation (original versus inverted), yielded a significant interaction, $F(1,37) = 4.16$, $p = .049$, but no significant main effects, both $F(1,37) < 1.59$, both $p > .216$ (Figure 2b). For the original pairs, sensitivity was significantly higher for the regular configuration than for the irregular configuration, $t(37) = 2.40$, $p = .021$. This VWM benefit for regular pairs was abolished by inversion: For the inverted pairs, there was no significant difference in sensitivity between the two pair configurations, $t < 1$. Performance in the verbal suppression task did not differ between conditions, all $F(1,37) < 1.39$, all $p > .246$. These results show that the VWM effect obtained in Experiment 1 cannot be explained by low-level differences between the regular and irregular pairs.

3.4 Experiment 3

The first two experiments demonstrated a VWM benefit for regularly positioned objects in comparison to irregularly positioned objects. It is possible, however, that this benefit reflected more efficient perceptual processing of regular configurations: Perhaps an encoding time of 2 seconds (Figure 2.1a) was sufficient to perceptually encode the regular but not the irregular displays. We conducted Experiment 3 to experimentally rule out the possibility that the VWM benefit observed here was fully due to perceptual limitations. Experiment 3 therefore included a condition where participants were given 4 seconds to encode the displays. If the effect of object configuration was primarily due to better perceptual encoding of regular displays we would expect a decreased effect for the four-second encoding condition, because limitations in perceptual encoding are reduced. Alternatively, if the benefit for regular displays reflected more efficient VWM storage or retrieval, the effect should be independent of encoding time.

3.4.1 Methods

Participants. Thirty-eight healthy adults (13 male; mean age 23.3 years, *SD* = 4.9) participated, of which four had also participated in Experiment 1 and five had participated in Experiment 2.

Stimuli and Apparatus. The apparatus, stimuli, and setup were identical to Experiment 1, but here we additionally manipulated the encoding time (i.e., the time for which the first display was presented).

Procedure. We used the same design as in Experiment 1, except that we now used only displays with three object pairs and added a condition in which the first display was presented for 4 seconds. This again led to a total of 192 trials (48 per condition), with the two configuration (regular versus irregular) and the two encoding time (2 seconds versus 4 seconds) conditions being randomly intermixed.

3.4.2 Results and Discussion

Trials with incorrect responses in the verbal suppression task (6.6%, *SD* = 6.1) were excluded from the analysis. A two-way repeated-measures ANOVA on mean *d*-prime scores, with the factors pair configuration (regular versus irregular) and encoding time (2 seconds versus 4 seconds), yielded a significant main effect of configuration, with higher sensitivity for the regular configuration than for the irregular configuration, $F(1,37) = 4.73$, $p = .036$. While longer encoding time led to better overall performance, $F(1,37) = 10.95$, $p = .002$, the effect of configuration was independent of encoding time, as indicated by a non-significant interaction[4], $F(1,37) < 1$, $p = .832$ (Figure 2.3). Performance in the verbal suppression task did not differ significantly between conditions, all $F(1, 37) < 1.36$, all $p > .250$. As the VWM benefit for regularly positioned objects was independent of the available encoding time, it seems unlikely that differential perceptual processing of regular and irregular configurations could fully account for the effect.

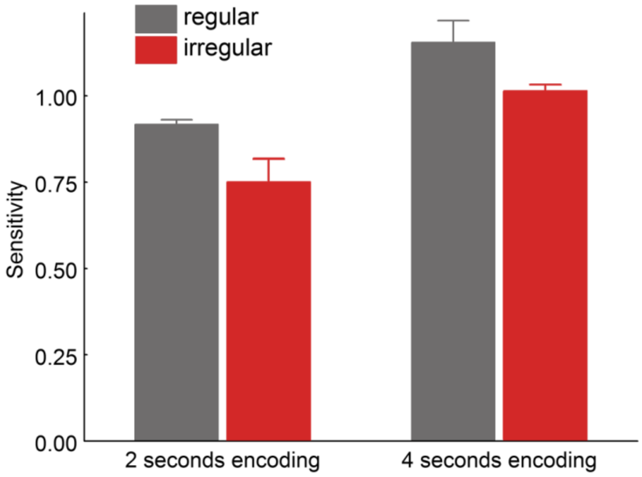

Figure 2.3. Longer encoding time does not reduce the VWM regularity effect. In the two- and four-second encoding durations we found an equally pronounced effect of object configuration, indicating that the effect does not depend on perceptual limitations. Standard errors reflect within-subject SEM (Cousineau, 2005).

3.5 General Discussion

Our study investigated the influence of real-world object regularities on VWM performance using a delayed change-detection paradigm with concurrent verbal suppression. We found that VWM performance was enhanced when pairs of objects

---

[4] A power analysis revealed that with a sample size of $N = 38$ and statistical power of 80%, interaction effects up to a difference in sensitivity differences of $\Delta d' = 0.34$ could have been reliably detected by this analysis.

were positioned according to such regularities, in comparison to an irregular positioning of the same objects. Crucially, this effect of regularity was significantly reduced when object pairs were inverted. Another control experiment (see Supplementary Material) ruled out that these findings were due only to the typical position of individual objects relative to any other object, independent of this other object's identity: Rather, both congruent object identities and regular relative positioning within a pair are required to give rise to the VWM effect. Thus, neither verbal memory strategies nor low-level grouping processes or the position of individual objects alone can account for the results. These results complement previous findings of better performance in perceptual tasks for regularly positioned objects (e.g., Gronau & Shachar, 2014; Kaiser et al., 2014a). Because the VWM effect was statistically independent of encoding time, our results are unlikely to solely reflect improved perception of regularly positioned objects. Rather, they indicate that real-world regularities are additionally associated with more efficient storage of objects in VWM. While previous work revealed the influence of associations among simple artificial stimuli on VWM (Brady et al., 2009; Olson et al., 2005), our findings show that lifelong experience with specific spatial configurations of real-world objects similarly facilitates VWM performance.

While our change-detection paradigm directly tested VWM, it is important to stress that the VWM benefit for regularly positioned objects can only emerge by additional recruitment of long-term memory processes. Stored knowledge can provide a conceptual link that allows for higher quality VWM representations by offering elaborated and structured coding frames. Such effects can be seen in enhanced VWM capacity for objects of expertise (Curby, Glazek, & Gauthier, 2009; Scolari, Vogel, & Awh, 2008), in contrast to dramatically reduced capacity for artificial stimuli that do not belong to separate categories (Olsson & Poom, 2005). As VWM can be enhanced when stimuli match pre-defined templates, but can also be harmed when stimuli violate these templates, it is worth noting that our results could in principle reflect either a VWM benefit for regularly positioned objects or a VWM cost for irregularly positioned objects (see Supplementary Material for further elaboration on this point).

Assuming that relational knowledge stored in long-term memory provides efficient schemata for organizing information in VWM, this process could be operating on all different stages of the memory process (i.e., encoding, maintenance, retrieval). Although we provide evidence that the effect observed here is not merely perceptual in nature, our findings do not directly address the question at which stage of the memory process the benefit for object regularities arises. We would expect that each of these stages can benefit from more effective information representation to some extent, but future work is necessary to pinpoint their exact contributions to the overall effect.

Interestingly, our study provides evidence that grouping influences VWM even if the tested stimulus dimension is different from the stimulus dimension underlying the grouping of items. Previous VWM studies investigating the effect of grouping typically used the same dimension for inducing the grouping of items and for testing memory performance. For example, Anderson et al. (2013) induced perceptual grouping by illusory contours that depended on the rotation of pacman-like inducers and found that VWM for the rotation of single inducers improved when an illusory contour was formed. Similarly, in studies that investigated the grouping of items based on color, VWM performance was measured by color judgments (Brady et al., 2009; Peterson & Berryhill, 2013). By contrast, in the present study memory performance was assessed by exemplar-level object discrimination, while the grouping was based on spatial relations and, importantly, on the object-category level. Thus, such high-level grouping of objects based on spatial-relational knowledge impacts VWM even when unrelated to the specific task.

The more effective VWM representation of objects that follow real-world regularities can be highly useful in natural perception. Indeed, memory for objects within natural scenes has been shown to be more effective than predicted by classical VWM models (Hollingworth, 2006; Hollingworth & Henderson, 2002). Because real-world spatial regularities appear on multiple levels and include a multitude of objects, the amount of grouping in natural scenes can be very high. This grouping of complex objects according to spatial-relational knowledge might thus represent a powerful mechanism of enhancing VWM in natural visual environments.

4 Study 3: Inter-object grouping facilitates visual awareness[5]

4.1 Introduction

Although visual scenes generate a complex, ambiguous mosaic of light on the retina, we have a stable, coherent conscious perception of our visual environment composed of objects, parts of objects, and groups of objects (Palmer, 1999). In organizing visual perception the visual system takes advantage of regularities in the visual input to group related image elements into higher-order perceptual units. Principles of such perceptual grouping determine the part-whole hierarchy among objects in a visual scene, thereby shaping conscious perception and contributing to the efficiency of visual processing (Wagemans et al., 2012). Most work on perceptual grouping has been carried out in the tradition of Gestalt psychology, investigating how physical regularities among simple stimuli such as dots, lines, or simple shapes influence visual perception.

However, our visual environment is not only structured by such physical regularities among simple image elements but also contains regularities among more complex, meaningful stimuli at more conceptual levels. For example, objects in real-world scenes do not appear at random locations, but are typically experienced at regular, predictable positions relative to each other (Bar, 2004): Lamps usually appear above not below tables. Recent evidence indicates that the visual system does extract such real-world spatial regularities among meaningful stimuli to perceptually group complex, natural objects we typically encounter in our everyday environments (Gronau & Shachar, 2014; Kaiser, Stein, & Peelen, 2014a; Riddoch, Humphreys, Edwards, Baker, & Wilson, 2003). Grouping based on this prior knowledge of the typical spatial configurations of

---

[5] This work has been published elsewhere: Stein, T., Kaiser, D., & Peelen, M. V. (2015). Inter-object grouping facilitates visual awareness. *Journal of Vision*, *15*, 10.

objects can improve object identification, short-term memory, and long-term memory retrieval (Kaiser, Stein, & Peelen, 2015; Roberts & Humphreys, 2011; Tibon, Gronau, Scheuplein, Mecklinger, & Levy, 2014). These initial findings raise the intriguing possibility that grouping of complex, meaningful objects enhances the efficiency of visual processing, in a way analogous to the well-established effects of Gestalt-like grouping among simple stimuli. Indeed, object grouping according to real-world spatial regularities is reflected in reduced attentional competition (Kaiser et al., 2014a), similar to reduced attentional competition for Gestalt-like grouping based on cues such as illusory contours (McMains & Kastner, 2010). Interestingly, physical regularities in the visual input can also determine whether we consciously perceive a stimulus in the first place. Recently, it has been found that simple stimuli that can be grouped by forming an illusory contour, are prioritized for access to conscious awareness (Wang, Weng, & He, 2012).

In the present study, we asked whether the grouping of natural, meaningful objects according to real-world regularities has a similar impact on the contents of conscious perception. To address this question, we tested whether grouping among complex objects can occur before these objects become available for conscious access and hence determine which objects gain access to conscious awareness. Visual awareness is thought to reflect the transient dominance of neural assemblies representing the conscious percept over competing assemblies representing other aspects of the visual input (Koch, 2004). These competitive dynamics can be tracked using continuous flash suppression (CFS), in which high-contrast patterns flashed into one eye can suppress conscious perception of stimuli presented to the other eye for several seconds (Tsuchiya & Koch, 2005). By tracking the duration of perceptual suppression under CFS for different stimuli the *breaking CFS* paradigm (b-CFS, Stein, Hebart, & Sterzer, 2011a) allows a direct comparison of the potency of different stimuli to gain access to awareness (e.g., Gayet, Van der Stigchel, & Paffen, 2014; Jiang, Costello, & He, 2007; Wang et al., 2012).

Adopting a b-CFS paradigm we compared suppression durations for object pairs presented in their typical, regular configuration with an irregular condition where the position of the individual objects was interchanged, thus disrupting regularity (see Figure 3.1a). If objects that can be grouped based on real-world regularities had a competitive advantage in gaining access to awareness, regularly positioned object pairs should break suppression more quickly than irregularly positioned pairs.

4.2 Experiment 1

Experiment 1 tested whether suppression durations would be shorter for regular than for irregular object pairs. Although regular and irregular pairs consisted of identical single objects with identical pixel values that only differed in their spatial configuration within the pairs, differences in breaking CFS could in principle be related to differences in the configuration of simple image elements between regular and irregular pairs (e.g., differences in Gestalt-like properties such as parallelism or symmetry). To control for such potential differences, we included a condition in which all pairs were inverted, that is, rotated by 180 degrees. Inversion disrupts the typical configuration of the pairs, while preserving all potential differences related to the grouping of simple image elements. Thus, inversion should abolish any genuine effect of real-world regularities.

4.2.1 Methods

Participants. In Experiment 1 we explored the possibility that inter-object positional regularities influence suppression durations for upright pairs but not for inverted pairs. For this first exploratory experiment we decided to test a relatively small sample size of *N* = 14. All 14 volunteers (all female, age range 18–36 years, mean 23.6 years) recruited through the University of Trento subject pool participated for course credit or payment. All participants gave informed written consent, reported normal or corrected-to-normal vision and were naïve as to the purpose of the experiment. The study protocol was approved by the ethical committee of the University of Trento and was carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki.

Apparatus and stimuli. Observers viewed a 19-in CRT monitor (1280 × 1024 pixels resolution, 100 Hz refresh rate) dichoptically through a custom-built mirror stereoscope. The observer's head was stabilized by a chin-and-head rest at a viewing distance of approximately 50 cm. The mirrors of the stereoscope were adjusted for each observer to promote stable binocular fusion. The screen was black except for the uniform light-gray area in which the stimuli were presented. Two red frames (10.4° × 10.4°) were displayed side-by-side on the screen such that one frame was shown to each eye (distance between the centers of the two frames 22.0°). To further support binocular fusion, noise contours (width 0.5°) consisting of random pixels were presented within the red frames. In the center of each frame a red fixation dot (0.5° × 0.5°) with a black dot (0.2° × 0.2°) in its center was displayed. Participants were asked to maintain stable fixation throughout the experiment. Visual stimuli were presented with Matlab (The MathWorks, Natick, MA) using the Cogent 2000 toolbox functions (www.vislab.ucl.ac.uk/cogent.php).

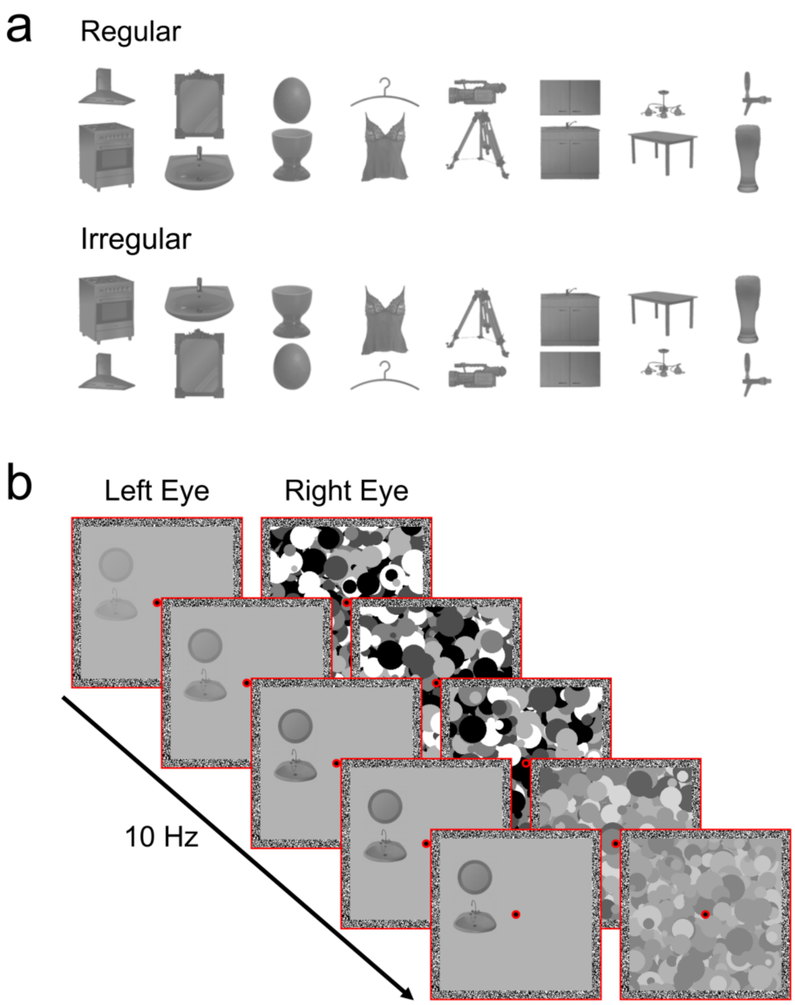

Figure 3.1. Stimuli and procedure. (a) Examples of upright regular (top row) and upright irregular (bottom row) object pairs. Regular and irregular pairs consisted of the same individual everyday objects. For the regular pairs, these objects were arranged according to their typical real-world configuration. For the irregular pairs the positions of the individual objects were interchanged. (b) Schematic of an example trial. To induce interocular suppression, CFS masks flashing at 10 Hz were presented to one eye, while a target stimulus was gradually introduced to the other eye. Participants indicated on which side of fixation the target stimulus or any part of the target stimulus became visible. The contrast of the target stimulus increased over the first second of a trial, while the contrast of the CFS masks was slowly ramped down over the course of a trial.

Target stimuli were 12 pairs of everyday-objects with a typical spatial configuration in the vertical direction, for example a lamp above a dining table, a bathroom mirror above a bathroom sink, or a TV screen above a DVD player. In the "regular" condition, these pairs were presented in their typical configuration (e.g., lamp above a dining table), whereas in the "irregular" condition all pairs were presented with individual object positions interchanged (e.g., lamp below a dining table; Figure 1a). For each single object pair, there were two different exemplars, resulting in a set of 24 object pairs (size 1.6°–2.8° × 2.8°–5.0°) per condition. Inverted versions of these regular and irregular pairs were created by presenting them upside-down (i.e., rotated by 180 degrees). To induce interocular suppression, we generated high-contrast, contour-rich CFS masks (9.2° × 9.2°) consisting of randomly arranged white, black, and gray circles (diameter 0.4°–1.8°; see Figure 3.1b).

An independent group of 16 observers answered two questions about 11 of the 12 regularly positioned object pairs (the two exemplars being counterbalanced across observers) in order to test if the two objects constituting a pair (a) were judged as commonly experienced together in this specific configuration and (b) were nevertheless perceived as two distinct objects. Participants answered on an ordinal scale from 1 ("fully disagree") to 7 ("fully agree"). For the first question ("I see these two objects often in this particular spatial arrangement") the mean of the average ratings for the different object pairs was high ($M = 5.78$), with little variability across pairs ($SD = 0.64$), demonstrating that the manipulation of regularity was successful. Also for the second question ("These are two distinct objects") ratings for the different object pairs were high with little variability ($M = 4.86$, $SD = 0.98$), meaning that despite the regularity manipulation the two individual objects constituting the pairs were still perceived as two separate objects.

Procedure. Each trial started with a 1-s fixation period. Subsequently, CFS masks changing at 10 Hz were presented to one eye and a target stimulus was introduced to the other eye. To avoid abrupt gradients, target stimuli were gradually faded in over the first second of each trial (by linearly increasing the contrast and simultaneously decreasing the luminance from light- to mid-gray) and then remained constant until the end of the trial (Figure 3.1b). Beginning 2 seconds after trial onset, the contrast of the CFS masks was linearly decreased to zero over 7 seconds. This contrast ramp was implemented to reduce the number of trials in which the target stimulus was not perceived at all. Target stimuli were presented until response or for a maximum trial duration of 10 seconds either to the left or to the right of the fixation dot (horizontal center-to-center distance 2.8°) at a random vertical position above or below the fixation dot (maximum vertical center-to-center distance 1.5°). Participants were required to press the left or the right arrow key on the keyboard to indicate whether the target stimulus appeared left or right to fixation. They were instructed to respond as soon as any part of the target stimulus became visible and to be as fast and accurate as possible. At the beginning of the experiment participants were informed about the presentation of two vertically arranged objects on every trial, but no information regarding the regularity manipulation was provided.

There were 192 trials (separated by breaks after 64 and 128 trials) in which each combination of two pair configurations (regular, irregular), two target orientations (upright, inverted), 24 target exemplars, and two eyes for target stimulus presentation occurred once. Trial order was randomized and the location of the target was selected at random for each trial.

Analysis. Only trials with correct responses and response times longer than 300 ms (*M* = 98.0%, *SD* = 1.4%) were included in the analyses. For our main analysis and for intuitive eyeballing of the results, we calculated means from the raw suppression durations. In addition, we conducted the same statistical analyses on log-transformed suppression durations to account for their positive skew (Heyman & Moors, 2014; Stein, End, & Sterzer, 2014a; Stein, Thoma, & Sterzer, 2015). Throughout this paper, we report Cohen's *d* as an effect size estimate for the paired *t*-tests, computed as the mean of the difference scores divided by the standard deviation of the difference scores.

4.2.2 Results

A repeated measures ANOVA with the factors pair configuration (regular, irregular) and target orientation (upright, inverted) on the means calculated from the raw suppression durations revealed a significant main effect of pair configuration, $F(1, 13) = 9.62$, $p = .008$, $\eta_p^2 = .43$, a marginally significant main effect of target orientation, $F(1, 13) = 4.00$, $p = .067$, $\eta_p^2 = .24$, and, most importantly, a significant interaction, $F(1, 13) = 8.97$, $p = .010$, $\eta_p^2 = .41$. When targets were presented in their normal upright orientation, suppression durations for regular pairs were significantly shorter than for irregular pairs, $t(13) = -3.54$, $p = .004$, $d = 0.95$ ($M = -448$ ms, $SD = 473$ ms, 95% CI [−720 ms, −175 ms], see Figure 3.2). Thus, regular object pairs overcame CFS and broke into awareness more quickly than irregular object pairs. Crucially, for inverted targets there was no significant difference in suppression durations between regular and irregular pairs, $t(13) = -0.39$, $p = .706$, $d = 0.10$ ($M = -29$ ms, $SD = 280$ ms, 95% CI [−191 ms, 133 ms]). Thus, differences in the grouping of simple image elements (which are preserved in inverted targets) are unlikely to account for the difference in suppression durations between upright regular and irregular pairs.

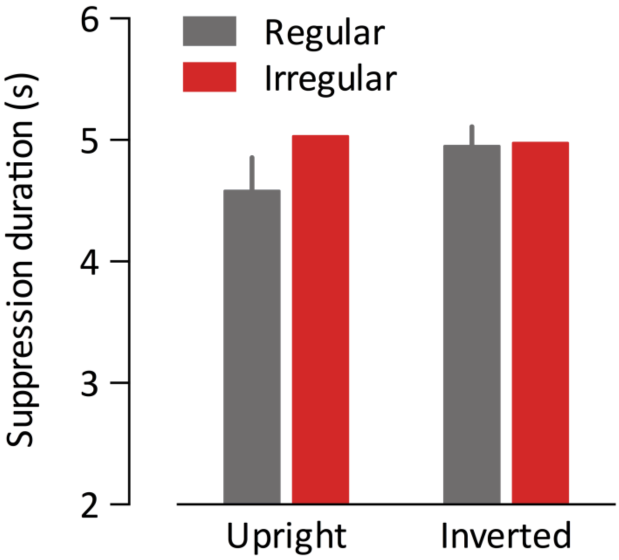

Figure 3.2. Results from Experiment 1. Bar plots show mean suppression durations for regular and irregular pairs, separately for target stimuli presented in their normal upright orientation and in inverted orientation (i.e., rotated by 180 degrees). Error bars denote 95% CIs for the mean difference between regular and irregular pairs, separately for upright and inverted targets.

An additional analysis on the log-transformed suppression durations revealed a similar pattern of results: There was a significant main effect of pair configuration, $F(1, 13) = 9.73$, $p = .008$, $\eta_p^2 = .43$, and a significant interaction between pair configuration and target orientation, $F(1, 13) = 9.84$, $p = .008$, $\eta_p^2 = .43$, while the main effect of target orientation did not reach significance, $F(1, 13) = 2.32$, $p = .152$, $\eta_p^2 = .15$. For upright object pairs log-transformed suppression durations were significantly shorter for regular pairs than for irregular pairs, $t(13) = -3.80$, $p = .002$, $d = 1.01$, whereas no such difference was found for inverted pairs, $t(13) = -0.43$, $p = .677$, $d = 0.11$. These results show that object pairs that are positioned according to real-world regularities have an advantage in gaining access to awareness.

Linear mixed effects analysis. To account for variability in suppression durations between individual stimulus items, we also performed linear mixed effects analyses using the lme4 package (Bates, Maechler, & Bolker, 2012) for R (R Core Team) on the raw suppression durations and on the log-transformed suppression durations (for similar b-CFS analyses see Heyman & Moors, 2014; Stein et al., 2014a). These analyses had random intercepts for participants and for individual exemplars of the object pairs. Reduced models containing only these random effects of participants and pair exemplars were tested against models including fixed effects of pair configuration (regular, irregular) or target orientation (upright, inverted) using likelihood ratio tests. To test for the interaction effect, models with the pair configuration-by-target orientation interaction were compared to models with the two fixed factors only.

For the analysis of raw suppression durations the comparison of the reduced model with the model containing the additional fixed factor of pair configuration was significant, $\chi^2(1) = 7.52$, $p = .006$, while the comparison with the model containing the additional fixed factor of target orientation was only marginally significant, $\chi^2(1) = 3.34$, $p = .068$. Most importantly, the interaction was significant, $\chi^2(1) = 5.65$, $p = .017$. Follow-up analyses for upright and inverted object pairs separately revealed that the main effect of pair configuration was significant only for upright pairs, $\chi^2(1) = 13.36$, $p < .001$, but not for inverted object pairs, $\chi^2(1) = 0.06$, $p = .810$. The results of the analysis of log-transformed suppression durations were similar, for pair configuration, $\chi^2(1) = 8.96$, $p = .003$, for target orientation, $\chi^2(1) = 2.18$, $p = .139$, and for the interaction, $\chi^2(1) = 6.31$, $p = .012$. Also for log-transformed suppression durations the main effect of pair configuration was significant only for upright pairs, $\chi^2(1) = 15.26$, $p < .001$, but not for inverted object pairs, $\chi^2(1) = 0.10$, $p = .751$. Thus, these results show that the influence of real-world regularities on access to awareness under CFS persisted after accounting for variability across individual object pair exemplars.

4.3 Experiment 2

A confirmatory second experiment was conducted to provide an internal replication (Experiment 2a) and to test the possibility that differences in the vertical position of individual objects could have accounted for the advantage of regular over irregular pairs in breaking CFS (Experiment 2b). Although regular and irregular pairs consisted of identical single objects that only differed in their configuration, these individual objects occupied slightly different spatial locations depending on whether they belonged to a regular or irregular pair: Objects that were presented on top of other objects in regular pairs (e.g., bathroom mirror) appeared on average further up in the CFS frames when they were part of a regular pair than when they were part of an irregular pair. Conversely, objects that were presented below other objects in regular pairs (e.g., bathroom sink) appeared on average further down in the CFS frames when they were part of a regular pair than when they were part of an irregular pair. In Experiment 2b, we presented single objects at the same positions as in the regular and irregular pairs. If the positioning of individual objects was driving the effect, we would expect to obtain shorter suppression durations for single objects that appeared at the same positions as in the regular pairs. If, however, faster awareness of regular pairs was related specifically to the relative positioning of the two objects forming a pair, that is, to their real-world configuration, no effect would be expected for single objects.

4.3.1 Methods

Participants. Experiment 2a was an identical replication of Experiment 1. For this confirmatory study we decided to run a larger sample size than in Experiment 1, in order to have sufficient power for detecting the effect of interest. We therefore decided to add another 10 participants to the sample size of Experiment 1, resulting in a total *N* of 24. Based on the effect size estimation from Experiment 1, this sample size yielded a power of 0.96 for obtaining the critical interaction between pair configuration and object orientation. This new set of 24 volunteers (21 female, age range 18–33 years, mean 22.7 years) participated for course credits or payment.

Apparatus, stimuli, and procedure. Experiment 2a was identical to Experiment 1. Experiment 2b was designed to control for differences between regular and irregular pairs regarding the vertical position of individual objects on the screen. For Experiment 2b, we created single object stimuli by replacing either the top or the bottom object from the regular and irregular pair images with the light-gray background. This resulted in a set of 48 "regular" and 48 "irregular" single-object target stimuli (24 "top objects", e.g., bathroom mirror, and 24 "bottom objects", e.g., bathroom sink, respectively, in each of their two possible positions within a pair). Only upright versions of these single objects were included in Experiment 2b.

The general procedure was identical to Experiment 1. The positions at which "regular" and "irregular" single-object targets could appear were the exact same positions at which the individual objects in regular and irregular pairs could appear. Experiment 2b contained 192 trials (separated by breaks after 64 and 128 trials) in which each combination of two target conditions ("regular", "irregular"), 48 target exemplars (24 "top objects", 24 "bottom objects"), and two eyes for target stimulus presentation occurred once. Trial order was randomized and the location of the target was selected at random for each trial. Half of the participants began with Experiment 2a, and the other half with Experiment 2b. The two experiments were separated by a short break.

Analysis. Again, only trials with correct responses and response times longer than 300 ms (Experiment 2a: *M* = 97.8%, *SD* = 1.8%; Experiment 2b: *M* = 97.8%, *SD* = 1.7%) were included in the computation of raw suppression durations and in the additional analysis of log-transformed suppression durations.

4.3.2 Results

Experiment 2a – Replication. The results of Experiment 2a replicated the findings of Experiment 1: A repeated measures ANOVA with the factors pair configuration (regular, irregular) and object orientation (upright, inverted) revealed a marginally significant main effect of pair configuration, $F(1, 23) = 4.00$, $p = .058$, $\eta_p^2 = .15$, no significant main effect of object orientation, $F(1, 23) = 1.99$, $p = .172$, $\eta_p^2 = .08$, but a significant interaction, $F(1, 23) = 14.30$, $p = .001$, $\eta_p^2 = .38$. For upright object pairs, suppression durations were again significantly shorter for regular pairs than for irregular pairs, $t(23) = -4.43$, $p < .001$, $d = 0.90$ (*M* = −276 ms, *SD* = 306 ms, 95% CI [−405 ms, −147 ms], see Figure 3.3a). As in Experiment 1, there was no significant difference in suppression durations between regular and irregular pairs when they were shown in inverted orientation, $t(23) = 0.81$, $p = .428$, $d = 0.16$ (*M* = 62 ms, *SD* = 375 ms, 95% CI [−97 ms, 220 ms]).

Again, an additional analysis of the log-transformed suppression durations confirmed these findings: There was a significant main effect of pair configuration, $F(1, 23) = 4.33$, $p = .049$, $\eta_p^2 = .16$, no significant main effect of object orientation, $F(1, 23) = 1.93$, $p = .178$, $\eta_p^2 = .08$, but a significant interaction, $F(1, 23) = 11.30$, $p = .003$, $\eta_p^2 = .33$. Log-transformed suppression durations for regular pairs were significantly shorter than for irregular pairs when presented in upright orientation, $t(23) = -4.39$, $p < .001$, $d = 0.90$, but not when presented in inverted orientation, $t(23) = 0.77$, $p = .452$, *d*

= 0.16. These results confirm the findings of Experiment 1, again demonstrating that objects positioned according to real-world regularities gain privileged access to awareness.

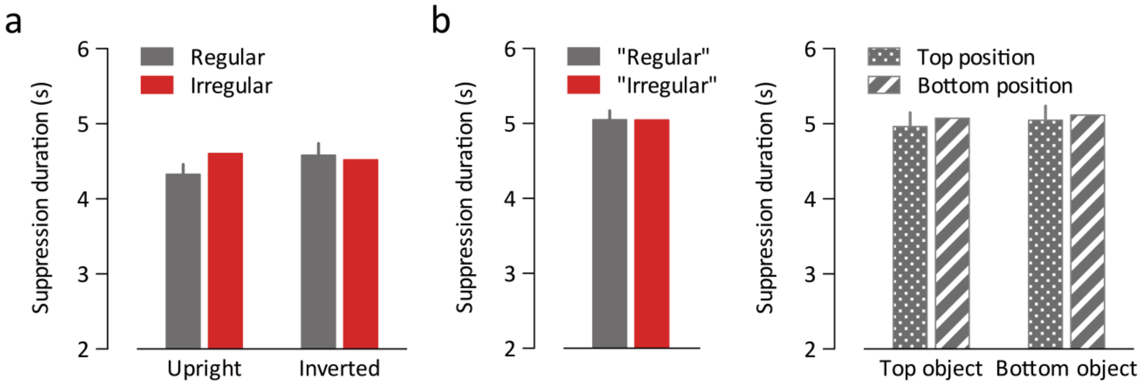

Figure 3.3. Results from Experiment 2. (a) Results from Experiment 2a, which was an exact replication of Experiment 1. Bar plots show mean suppression durations for regular and irregular pairs, separately for target stimuli presented in their normal upright orientation and in inverted orientation. Error bars denote 95% CIs for the mean difference between regular and irregular pairs, separately for upright and inverted targets. (b) Results from Experiment 2b, which served to control for differences in the vertical position of individual objects in the previous experiments. The bar plots on the left denote mean suppression durations for individual objects derived from regular pairs and irregular pairs. The bar plots on the right show mean suppression durations for single objects as a function of target position (top: above fixation, bottom: below fixation) and object type, that is, depending on whether the single object was presented on top of another object in the original pairs (top object, e.g., bathroom mirror) or below another object in the original pairs (bottom object, e.g., bathroom sink). Error bars denote 95% CIs for the mean difference between single objects from regular and irregular pairs and between top and bottom positions, respectively.

Linear mixed effects analysis. In addition, as for Experiment 1 we carried out a linear mixed effects analysis to account for variability between individual object pair exemplars. The analysis of raw suppression durations yielded no significant main effects of pair configuration, $\chi^2(1) = 2.70$, $p = .010$, or target orientation, $\chi^2(1) = 1.81$, $p = .179$, but, importantly, a significant interaction, $\chi^2(1) = 7.11$, $p = .008$. The main effect of pair configuration was significant only for upright pairs, $\chi^2(1) = 9.40$, $p = .002$, but not for inverted object pairs, $\chi^2(1) = 0.49$, $p = .486$. Similarly, the analysis of log-transformed suppression durations yielded no significant main effects of pair configuration, $\chi^2(1) = 2.52$, $p = .112$, or target orientation, $\chi^2(1) = 1.69$, $p = .193$, but a significant interaction, $\chi^2(1) = 6.55$, $p = .011$. Again, the main effect of pair configuration was significant only for upright pairs, $\chi^2(1) = 8.67$, $p = .003$, but not for inverted object pairs, $\chi^2(1) = 0.43$, $p = .513$. Thus, these results show that also in Experiment 2a the beneficial influence of real-world regularities on access to awareness persisted after accounting for variability across individual object pair exemplars.

Experiment 2b – Single-objects control. To test whether this effect could have been due to the slightly different positioning of individual objects in regular and irregular pairs, we compared suppression durations for single objects that appeared at the same spatial locations as in the pairs. Crucially, there was no significant difference in suppression durations between single objects from regular and irregular pairs, $t(23) = 0.07$, $p = .941$, $d = 0.02$ ($M = 4$ ms, $SD = 268$, 95% CI [−109 ms, 117 ms], see Figure 3.3b). Moreover, when directly comparing Experiment 2a and 2b, the difference in suppression durations between (upright) regular and irregular pairs was larger than between “regular” and “irregular” single objects, as reflected in a significant interaction between experiment and configuration, $F(1, 23) = 7.67$, $p = .011$, $\eta_p^2 = .25$.

The analysis of log-transformed suppression durations from Experiment 2b revealed similar results: There was no significant difference between single objects from regular and irregular pairs, $t(23) = -0.57$, $p = .576$, $d = 0.12$, and the advantage of (upright) regular over irregular pairs in Experiment 2a was larger than the difference between "regular" and "irregular" single objects, $F(1, 23) = 6.70$, $p = .016$, $\eta_p^2 = .23$. Thus, the relative position of individual objects cannot explain the advantage of regular over irregular pairs in access to awareness.

Finally, we further explored whether spatial locations influence access to awareness of objects as a function of whether an object is typically seen on top or below other objects. Objects that are typically seen above other objects generally more often fall in the upper part of the visual field and could thus be detected better when appearing in the upper as compared to the lower visual field, whereas objects that are typically seen below other objects more often fall in the lower visual field and might be detected better there. To address this possibility, we computed mean suppression durations depending on the position of the target (above vs. below fixation) and the type of object ("top object", e.g., bathroom mirror, vs. "bottom object", e.g., bathroom sink). An ANOVA yielded no significant main effects of target position, $F(1, 23) = 2.64$, $p = .118$, $\eta_p^2 = .10$, or object type, $F(1, 23) = 1.57$, $p = .223$, $\eta_p^2 = .06$, and, most importantly, no significant interaction, $F(1, 23) = 0.10$, $p = .759$, $\eta_p^2 < .01$, meaning that the spatial location in the CFS frames did not influence breakthrough into awareness differently for different types of objects (see Figure 3b). Similarly, the analysis of log-transformed suppression durations yielded no significant main effects of target position, $F(1, 23) = 1.37$, $p = .254$, $\eta_p^2 = .06$, or object type, $F(1, 23) = 2.32$, $p = .141$, $\eta_p^2 = .09$, and no significant interaction, $F(1, 23) = 0.09$, $p = .765$, $\eta_p^2 < .01$. These results further support the notion that the difference in suppression duration between regular and irregular pairs is due to the configuration of the pairs rather than to their positions on the screen.

4.4 Discussion

The present results demonstrate that objects that can be grouped based on real-world spatial regularities are prioritized for access to conscious awareness. Two experiments revealed faster access to awareness for object pairs that were positioned in the configuration in which they typically co-occur in the real world. This advantage of regularly positioned object pairs was abolished by stimulus inversion, meaning that the effect cannot reflect physical stimulus differences or grouping of simple image elements. Rather, our findings indicate that experience-based grouping of complex, meaningful objects can occur before these objects become available for conscious access, thereby determining which objects are consciously perceived in the first place.

This advantage for grouped objects is similar to the advantage in breaking CFS for simple shapes that can be grouped to a Kanizsa figure through illusory contours (Wang et al., 2012). Thus, both grouping of simple stimuli as well as grouping of meaningful, complex stimuli can transpire before conscious access. The underlying mechanisms, however, are most likely markedly different. The representation of physical (e.g.,geometrical) relationships among simple stimuli, such as those leading to the formation of illusory contours, seems to rely on both early visual cortical areas and higher-level ventral stream areas (e.g., Abu Bakar, Liu, Conci, Elliott, & Ioannides, 2009; Stanley & Rubin, 2003; von der Heydt, Peterhans, & Baumgartner, 1984), whereas the representation of object-object relations likely involves only higher occipitotemporal object processing areas (Kim & Biederman, 2010; Roberts & Humphreys, 2010). Distributed patterns of activity in these areas evoked by two objects can be modeled as a linear combination of the response patterns to the individual objects (MacEvoy & Epstein, 2009; Reddy, Kanwisher, & VanRullen, 2009) and the relative weighting of the two patterns seems to be altered when the two objects form meaningful spatial relationships (Baeck, Wagemans, & Op de Beeck, 2013; but see Kaiser, Strnad, Seidl, Kastner, & Peelen, 2014b), indicating that these object configurations are represented in visual cortex activity patterns. Furthermore, Kanizsa-type figures do not only induce the perception of illusory contours but also of an illusory surface, which constitutes a salient region that “pops out” in visual search (Davis & Driver, 1994; Gurnsey, Poirier, & Gascon, 1996). Thus, differences in suppression durations for these stimuli may reflect differences in preconsciously extracted bottom-up saliency (cf. Gayet et al., 2014). By contrast, the present findings cannot be due to differences in bottom-up saliency between regular and irregular pairs, but likely reflect a mechanism of perceptual grouping based on knowledge about the relative positions of objects that often co-occur in the real world.

This central role of real-world perceptual experience in modulating access to visual awareness is consistent with findings from other studies showing that the dynamics of interocular competition are influenced by experience with our environment (Gayet et al., 2014). For example, stimuli whose low-level properties follow natural image statistics tend to dominate perception in binocular rivalry (Baker & Graf, 2009). Also natural objects such as human faces and bodies overcome CFS more quickly when they are presented in their familiar upright orientation than when their typical spatial configuration is disrupted by inversion (e.g., Jiang et al., 2007; Stein, Peelen, & Sterzer, 2011b; Stein, Sterzer, & Peelen, 2012; Stein et al., 2014a; Yang, Zald, & Blake, 2007; Zhou, Zhang, Liu, Yang, & Qu, 2010). The present results go beyond these previous studies by showing for the first time that the relative spatial position of two upright, locally identical objects can determine access to awareness. Thus, while previous findings can be explained by a general advantage for more recognizable or meaningful stimuli (e.g., upright faces), differences in recognizability of individual objects cannot explain our results, because individual stimuli were identical across conditions. Only the relative positioning can render regular object pairs more meaningful and facilitate their recognition (Gronau & Shachar, 2014; Roberts & Humphreys, 2011; Tibon et al., 2014). The present findings may thus reflect the increased meaningfulness of coherent object pairs, indicating that inter-object grouping can precede conscious access.

How, then, could object grouping influence the duration of perceptual suppression? According to the unconscious binding hypothesis spatiotemporarily distributed visual stimuli can be bound into coherent objects even when rendered invisible (Lin & He, 2009). Indeed, the advantage of radial over random motion in b-CFS (Kaunitz, Fracasso, Lingnau, & Melcher, 2013) indicates that the visual system can extract physical regularities from suppressed stimuli to form coherent patterns, which are then prioritized for conscious access. Inter-object grouping that emerges from such preconscious binding of individual objects may similarly entail the formation of coherent, integrated multi-object representations, either through neural assemblies in object-sensitive cortex or through context-facilitated reentrant circuitry between frontal and occipitotemporal areas (Fenske, Aminoff, Gronau, & Bar, 2006). This unified representation of regularly positioned objects seems to be a more potent competitor for access to the capacity-limited stage of conscious awareness than the representations of single objects alone. This conclusion is consistent with the general notion that b-CFS is sensitive to complex stimulus properties such as familiarity, ecological relevance, or meaningfulness, whereas the extraction of even more complex stimulus attributes, such as word semantics, may require conscious access (Gayet et al., 2014).

In the present study we used b-CFS as a paradigm to probe potency of regular and irregular object pairs to gain access to awareness. Several previous studies have used this approach to study unconscious processing transpiring specifically under interocular suppression (e.g., Jiang et al., 2007; Mudrik, Breska, Lamy, & Deouell, 2011; Stein, Senju, Peelen, & Sterzer, 2011c; Wang et al., 2012; Zhou et al., 2010). These studies included a binocular control condition not involving interocular suppression and inferred CFS-specific unconscious processing when the effect obtained with b-CFS was larger than the effect obtained with this control condition. However, because the logic of relying on such a control condition for inferring CFS-specific unconscious processing has recently been questioned on theoretical and empirical grounds (Stein et al., 2011a; Stein & Sterzer, 2014), here we did not include such a binocular control condition (also see e.g., Gray, Adams, Hedger, Newton, & Garner, 2013; Stein et al., 2014a, 2014b, 2015; Tsuchiya, Moradi, Felsen, Yamazaki, & Adolphs, 2009; Yang et al., 2007). The current findings could thus reflect more general differences in detectability between regularly and irregularly positioned object pairs rather than differences in CFS-specific unconscious processing. Still, even such non-specific differences in stimulus detectability most likely reflect differences in the processes that precede and lead to conscious access (e.g., Kaunitz et al., 2013).

To provide unequivocal evidence for unconscious processing differences between regular and irregular object pairs, future studies will need to show that real-world spatial regularities continue to be extracted from objects that are rendered permanently invisible. However, this approach may be less sensitive to the visual processes that precede conscious access than the b-CFS paradigm in which initially invisible stimuli eventually cross the threshold to consciousness. For example, neuroimaging studies have repeatedly shown that the processing of stimuli rendered permanently invisible through CFS is strongly reduced in those higher-level visual areas along the ventral stream that are likely candidates for representing spatial regularities among complex objects (for a review, see Sterzer, Stein, Ludwig, Rothkirch, & Hesselmann, 2014). Because to date no study has investigated the spatiotemporal dynamics of the neural processes associated with competition for awareness during b-CFS, it remains possible that the advantage of regular over irregular pairs in b-CFS involves occipitotemporal and even frontal cortices.

Another important challenge for future work will be to investigate to what extent these findings obtained with the laboratory paradigm of b-CFS extend to other paradigms for measuring access to awareness and, most importantly, to more naturalistic situations and to real-world perception. Although b-CFS seems to be a particularly sensitive device for probing differences in stimulus detectability, recent studies that used both b-CFS and other psychophysical paradigms for studying access to awareness have shown similar effects with b-CFS and standard sandwich masking (Stein et al., 2014b) and rapid serial visual presentation (Gobbini et al., 2013). It is thus likely that findings obtained with b-CFS can similarly be found with other, sufficiently sensitive psychophysical laboratory techniques. One promising avenue for determining the extent to which perceptual mechanisms uncovered with such laboratory experiments influence behavior in real-world situations consists in using more naturalistic stimulus material, such as photographs of real-world scenes (for a review, see Peelen & Kastner, 2014). For example, the current stimuli could be embedded in naturalistic scenes to test whether inter-object grouping facilitates perceptual performance in a more ecological setting.

While our results show that objects in regular configurations are prioritized for conscious access, another recent b-CFS study found shorter suppression durations for photographs of complex scenes that contained semantically incongruent objects, for example a checkerboard in an oven (Mudrik et al., 2011). However, in contrast to the present approach in which we only changed the configuration of identical objects, this study compared suppression durations to physically different stimuli and could therefore not rule out that these results reflected visual rather than semantic factors. Nevertheless, their findings suggest that gross violations of semantic context are rapidly detected, bringing an unexpected stimulus more quickly into awareness, perhaps through a preconscious novelty or surprise response. This advantage of incongruent scenes is not necessarily inconsistent with the present findings, as the two objects in our irregular condition were always semantically congruent. Thus, in the absence of gross semantic violations the visual system is tuned to those stimuli that are typically encountered in real-world environments. The present findings now demonstrate that this principle applies even to the complex spatial-relational regularities among natural objects: Objects that follow these real-world regularities are prioritized for conscious access.

5 Concluding Discussion

Here, we provide evidence for a novel inter-object grouping process that allows the brain to efficiently reduce visual complexity by making use of ubiquitously encountered real-world regularities. Using fMRI we demonstrate that grouping based on such regularities has a profound impact on neural responses by reducing competition between objects. We furthermore show that this grouping process enhances behavioral performance in a range of tasks: (1) Grouped objects can be rejected more efficiently when they are distracters, boosting visual search performance, (2) grouped objects can be more successfully stored in visual working memory, and (3) grouped objects are prioritized when competing for awareness. Importantly, in a series of control experiments, we ensured that none of these effects were attributable to low-level grouping processes or relative object positions. By contrast, the benefits of real-world regularities were only observed when two specific objects were both presented in their typical spatial locations, and in an upright orientation. These findings indicate that inter-object grouping requires the integration of both object category and object position to facilitate the processing of typical object arrangements.

But what is the neural mechanism underlying the grouping process we observed here? One possibility is that the visual system houses compound representations that are established based on frequent co-occurrence of objects in real-world environments. If we encounter lamps above tables very often, the visual system could adapt and make use of this specific, recurring pattern by forming a representation for the two objects as a group. Such compound representations for object groups would ultimately have to integrate information about object category and position to reflect the information inherent to real-world properties, that is, such a representation shouldn’t be recruited for lamps appearing above other things in general (although such representations might also exist because of the facts that lamps preferentially appear in the upper hemifield), and also not for lamps below tables or next to tables (as this wouldn't reflect the spatial dependencies in the world). Evidence for the existence of such position-sensitive compound representations has been found in fMRI studies investigating grouping based on action cues (Baeck et al., 2013), and the cortical representation of the human body and face (Bernstein, Oron, Sadeh, & Yovel, 2014; Schmalzl, Zopf, & Williams, 2012; but see Kaiser et al., 2014b). Similar representations for groups defined by real-world regularities seem capable of explaining the pattern of results described here. On one hand, the activation of additional compound representations would facilitate detection of stimuli under inter-ocular suppression and enhance memory performance, as in both cases regular stimuli would be more thoroughly processed by the additional recruitment of the group representation. On the other hand, compound representations can also reduce the impact of unwanted information on the processing of relevant information. As demonstrated in our fMRI and visual search experiments, distracting objects groups interfere less when they follow real-world regularities. One explanation for this is a reduction of effective set size, when group representations become the units of processing. Going beyond that, the effect observed in our visual search experiments can also be interpreted as a more efficient suppression (see Seidl, Peelen, & Kastner, 2012) of the representation of entire distracter groups instead of a more inefficient suppression of individual objects.

Alternatively, inter-object grouping could originate from close interactions between distinct and independent object representations. When certain objects are frequently encountered together in specific combinations, the wiring between these object representations can be strengthened by associative learning. Neurophysiological studies suggest that learned relations between pairs of stimuli can influence neural tuning in primate inferior temporal cortex: Neurons previously selective for only one of the two stimuli show equal firing behaviour for both stimuli after the two items have been behaviourally associated (Messinger, Squire, Zola, & Albright, 2001; Sakai & Miyashita, 1991). Such co-activations after associative learning can be explained by entrained neural connections between representations that result from repeatedly pairing the stimuli. These entrained connections would in turn allow for activation to spread over to the associated representational nodes (see Collins & Loftus, 1975). However, the grouping observed here goes beyond a pure semantic association between items, as it is also tied to the relative spatial locations of the objects. Hence, if inter-object grouping is explicable by enhanced connectivity between independent object representations, connections have to be differentially recruited as a function of spatial regularities. It is worth emphasizing that these two possible mechanisms underlying inter-object grouping are not mutually exclusive. It is clear that objects that can be grouped based on the regularities described here are very highly related in the semantic domain, and thus could be represented relatively similarly to begin with, allowing for effective connections between individual object representations. Typical spatial regularities can then influence this connectivity, or can activate distinct compound representation, or, crucially, can do both at the same time. Further studies are needed to disentangle these possible mechanisms. A promising avenue for resolving the issue of integrated versus separate neural representations is the reconstruction of response patterns for regular and irregular object arrangements from response patterns for their constituent objects: If a linear combination of single object response patterns accurately approximates the response pattern to an object group, additional compound representations are unlikely to be recruited (see Baeck et al., 2013; Kaiser et al., 2014b).

The inter-object grouping mechanism described here is one of many different integration (and segregation) mechanisms that are going on in parallel and on different hierarchical levels during natural vision. Constrained by the inherent receptive field organization of the visual cortex (Rolls, 2000; Smith, Singh, Williams, & Greenlee, 2001), visual input is carved up from locally defined, simple attributes in the beginning of the processing stream to more global and abstract properties towards the end of the visual hierarchy. On the lower levels of this hierarchy, input is organized by simple features within locally restricted spatial locations. During these early stages, integration and segregation of display elements is thus determined by low-level attributes: Elements of similar luminosity, color, orientation, and spatial location can be grouped together by rules referred to as Gestalt principles (Wagemans et al., 2012; Wertheimer, 1923). These processes support the formation of objects by supporting feature binding, texture segmentation, and figure-ground organization (Wagemans et al., 2012). On the level of objects, first complex feature conjunctions are represented, which may constitute parts of other objects (Hayworth & Biederman, 2006; Tanaka, 1996). Subsequently, these parts have to be integrated to form representations of meaningful real-world categories such as faces, bodies, and tools (Downing, Chan, Peelen, Dodds, & Kanwisher, 2006). Even later in processing, whole visual scenes are represented, integrating information about individual objects and spatial layout to implement tasks such as scene categorization (Walther, Caddigan, Fei-Fei, & Beck, 2009) and navigation (Epstein & Vass, 2014). Our work provides evidence for an additional step of information integration within this processing hierarchy: At the level of meaningful objects, processing can be facilitated by grouping objects based on real-world regularities, which can in turn contribute to scene representations.

Previous research has shown that scene-selective brain regions like the PPA respond more strongly to objects that fulfil certain criteria like large real-world size (Cate, Goodale, & Köhler, 2011), landmark suitability (Troiani, Stigliani, Smith, & Epstein, 2012), space-defining properties (Mullally & Maguire, 2011), and strong associations with a particular scene context (Bar, Aminoff, & Schacter, 2008). Some of these scene-defining object characteristics, such as the degree to which they convey information about scene layout and category, are potentially easier to extract from frequently encountered groups of objects that are arranged in a spatially prototypical way (e.g., seeing a sink, a mirror, and a bathtub together in a typical spatial configuration certainly creates a more vivid impression of a bathroom scene than just one of the three objects alone). Although our fMRI study doesn't provide evidence for PPA preferring grouped over non-grouped stimuli per se, the representation of object groups might act as an interface between object and scene processing. Clearly, further investigations are needed to explore how object grouping can contribute to the formation of scene representations, for example by testing whether the discrimination of different scene categories in visual cortex is enhanced as a function of object regularities.

Interesting parallels to the inter-object grouping process described here can be found in the domain of reading. For the visual processing of words a similar hierarchy can be traced in the brain, where on different levels of the visual processing stream more and more complex arrangements (from single letters to letter combinations to words) are integrated (Dehaene, Cohen, Sigman, & Vinckier, 2005; Vinckier et al., 2007). While posterior regions are most selective for the visual features of a displayed font, independently of the arrangements of the letters, more anterior regions show a preference for frequent groups of letters (e.g., “QU” or “WH”) over infrequent ones (e.g., “QH” or “WU”; Binder, Medler, Westbury, Liebenthal, & Buchanan, 2006; Vinckier et al., 2007) and even further up the hierarchy, neurons in the visual word form area (VWFA; Cohen et al., 2000) show a preference of words over pseudowords (Bruno, Zumberge, Manis, Lu, & Goldman, 2008; Schurz et al., 2005; but see Vinckier et al., 2007). According to the local combination detector model of word processing (Dehaene et al., 2005), word processing is implemented by a hierarchical combination of receptive fields that are tuned to increasingly complex input: (1) specific letter shapes, (2) abstract letter information, (3) frequent local letter combinations, and (4) recurring substrings and words. The intermediate step of letter grouping based on typically encountered local configurations (3) is nicely congruent with our results from the object domain: On this level, both identity and relative position of the letters need to be taken into account to activate a group representation. The combination of individual items into group representation based on highly experienced arrangements thus seems to be applied for different hierarchically decomposable stimuli – at least for both visual scenes and written text. Future studies could explore to which extent the local to global tuning for increasingly complex subsets of words along the posterior-to-anterior axis in visual processing (Vinckier et al., 2007) conceptually resembles a grouping mechanism that integrates over an increasingly large number of real-world objects (from single objects to pairs of objects to more complex ensembles to whole scenes) along the same axis.

To summarize, our studies provide compelling evidence for an inter-object grouping mechanism that is based on real-world regularities. Grouping scene elements on the level of objects helps to reduce the complexity of real-world environments and allows the brain to carve up scenes on the level of object groups, thereby reducing competition between objects. We believe that these results reflect a selective tuning of the visual system for frequently experienced object arrangements. Altogether, our findings help to bridge the gap between laboratory experiments and real-world tasks: Inter-object grouping provides a novel explanation for highly efficient behaviour in complex real-world environments, which often seems to be in apparent contradiction to the capacity limitations revealed in laboratory experiments using simple stimuli.

6 References

Abu Bakar, A., Liu, L., Conci, M., Elliott, M. A., & Ioannides, A. A. (2008). Visual field and task influence illusory figure responses. *Human Brain Mapping, 29*, 1313-1326.

Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). An area within human ventral cortex sensitive to “building” stimuli: Evidence and implications. *Neuron, 21*, 373-383.

Altmann, C. F., Bülthoff, H. H., & Kourtzi, Z. (2003). Perceptual organization of local elements into global shapes in the human visual cortex. *Current Biology, 13*, 342-349.

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science, 15*, 106-111.

Anderson, D. E., Vogel, E. K., & Awh, E. (2013). Selection and storage of perceptual groups is constrained by a discrete resource in working memory. *Journal of Experimental Psychology: Human Perception and Performance, 39*, 824-835.

Baddeley, A. D. (1986). *Working Memory*. Oxford, UK: Clarendon Press.

Baeck, A., Wagemans, J., & Op de Beeck, H. P. (2013). The distributed representation of random and meaningful object pairs in human occipitotemporal cortex: The weighted average as a general rule. *Neuroimage, 70*, 37-47.

Baker, D. H., & Graf, E. W. (2009). Natural images dominate in binocular rivalry. *Proceedings of the National Academy of Sciences USA, 106*, 5436-5441.

Banks, W. P., & Prinzmetal, W. (1976). Configurational effects in visual information processing. *Perception & Psychophysics, 19*, 361-367.

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

Bar, M., Aminoff, E., & Schacter, D. L. (2008). Scenes unseen: The parahippocampal cortex intrinsically subserves contextual associations, not scenes or places per se. *Journal of Neuroscience, 28*, 8539-8544.

Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using Eigen and S4. Version 1.1-6. Available online at: http://cran.r-project.org/web/packages/lme4/.

Beck, D. M., & Kastner, S. (2005). Stimulus context modulates competition in human extrastriate cortex. *Nature Neurocscience, 8*, 1110-1116.

Beck, D. M., & Kastner, S. (2007). Stimulus similarity modulates competitive interactions in human visual cortex. *Journal of Vision, 7*, 1-12.

Bernstein, M., Oron, J., Sadeh, B., & Yovel, G. (2014). An integrated face-body representation in the fusiform gyrus but not the lateral occipital cortex. *Journal of Cognitive Neuroscience, 26*, 2469-2478.

Binder, J., Medler, D. A., Westbury, C. F., Liebenthal, E., & Buchanan, L. (2006). Tuning of the human left fusiform gyrus to sublexical orthographic structure. *Neuroimage, 33*, 739-748.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: Using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General, 138*, 487-502.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences USA, 105*, 14325-14329.

Brady, T. F., & Tenenbaum, J. B. (2013). A probabilistic model of visual working memory: Incorporating higher order regularities into working memory capacity estimates. *Psychological Review, 120*, 85-109.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433-436.

Brett, M., Anton, J.-L., Valabregue, R., & Poline, J.-B. (2002). Region of interest analysis using an SPM toolbox. *Neuroimage, 16*, 1140–1141.

Broadbent, D. E. (1958). *Perception and communication.* London: Pergamon Press.

Brockmole, J. R., & Henderson, J. M. (2006). Recognition and attention guidance during contextual cueing in real-world scenes: Evidence from eye movements. *Quarterly Journal of Experimental Psychology, 59*, 1177-1187.

Bruno, J. L., Zumberge, A., Manis, F. R., Lu, Z. L., & Goldman, J. G. (2008). Sensitivity to orthographic familiarity in the occipito-temporal region. *Neuroimage, 39*, 1988-2001.

Bundesen, C., & Pedersen, L. F. (1983). Color segregation and visual search. *Perception & Psychophysics*, *33*, 487-493.

Cate, A. D., Goodale, M. A., & Köhler, S. (2011). The role of apparent size in building- and object-specific regions of ventral visual cortex. *Brain Research, 1388*, 109-122.

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215–281). New York: Academic Press.

Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences, 4*, 170-178.

Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S, Dehaene-Lambertz, G., Hénaff, M. A., & Michel, F. (2000). The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain, 123*, 291-307.

Cohen, M. A., Konkle, T., Rhee, J. Y., Nakayama, K., & Alvarez, G. A. (2014). Processing multiple visual objects is limited by overlap in neural channels. *Proceedings of the National Academy of Sciences USA, 111*, 8955-8960.

Cohen, M. A., Nakayama, K., Konkle, T., Stantic, M., & Alvarez, G. A. (2015). Visual awareness is limited by the representational architecture of the visual system. *Journal of Cognitive Neuroscience, 27*, 2240-2252.

Collins, A. M., & Loftus, E. F. (1975). A spreading activation theory of semantic processing. *Psychological Review, 82*, 407-428.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Massons's method. *Tutorials in Quantitative Methods for Psychology, 1*, 42-45.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87-114.

Curby, K. M., Glazek, K., & Gauthier, I. (2009). A visual short-term memory advantage for objects of expertise. *Journal of Experimental Psychology: Human Perception and Performance, 35*, 94-107.

Davis, G., & Driver, J. (1994). Parallel detection of Kanizsa subjective figures in the human visual system. *Nature, 371*, 791-793.

Dehaene, S., Cohen, L., Sigman, M., & Vinckier, F. (2005). The neural code for written words: A proposal. *Trends in Cognitive Sciences, 9*, 335-341.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18*, 193-222.

Donnelly, N., Humphreys, G. W., & Riddoch, M. J. (1991). Parallel computation of primitive shape descriptions. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 561-570.

Downing, P. E., Chan A. W., Peelen, M. V., Dodds, C. M., & Kanwisher, N. (2006). Domain specificity in visual cortex. *Cerebral Cortex, 16*, 1453-1461.

Driver, J. (1995). Object segmentation and visual neglect. *Behavioral and Brain Sciences, 71*, 135-146.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General, 113*, 501-517.

Epstein, R. A., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392*, 598-601.

Epstein, R., & Vass, L. K. (2013). Neural systems for landmark-based wayfinding in humans. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*, 20120533.

Fang, F., Kersten, D., & Murray, S. O. (2008). Perceptual grouping and inverse fMRI activity patterns in human visual cortex. *Journal of Vision, 8*, 1-9.

Feng, M. J., Aminoff, E., Gronau, N., & Bar, M. (2006). Top-down facilitation of visual object recognition: Object-based and context-based contributions. *Progress in Brain Research, 155*, 3-21.

Franconeri, S. L., Alvarez, G. A., & Cavanagh, P. (2013). Flexible cognitive resources: Competitive content maps for attention and memory. *Trends in Cognitive Sciences, 17*, 134-141.

Gayet, S., Van der Stigchel, S., & Paffen, C. L. E. (2014). Breaking continuous flash suppression: Competing for consciousness on the pre-semantic battlefield. *Frontiers in Human Neuroscience, 5*, 460.

Gobbini, I., Gors, J. D., Halchenko, Y. O., Rogers, C., Guntupalli, S., Hughes, H., & Cipolli, C. (2013). Prioritized detection of personally familiar faces. *PLoS ONE, 8*, e66620.

Gray, K. L. H., Adams, W. J., Hedger, N., Newton, K. E., & Garner, M. (2013). Faces and awareness: Low-level, not emotional factors determine perceptual dominance. *Emotion, 13*, 537-544.

Green, C. B., & Hummel, J. E. (2006). Familiar interacting object pairs are perceptually grouped. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 1107-1119.

Gronau, N., Neta, M., & Bar, M. (2008). Integrated contextual representation for objects' identities and their locations. *Journal of Cognitive Neuroscience, 20*, 371-388.

Gronau, N., & Schachar, M. (2014). Contextual integration of visual objects necessitates attention. *Attention, Perception, & Psychophysics, 76*, 695-714.

Gurnsey, R., Poirier, F. J., & Gascon, E. (1996). There is no evidence that Kanisza-type subjective contours can be detected in parallel. *Perception, 25*, 861-874.

Hayworth, K. J., & Biederman, I. (2006). Neural evidence for intermediate representations in object recognition. *Vision Research, 46*, 4024-4031.

Heyman, T., & Moors, P. (2014). Frequent words do not break continuous flash suppression differently from infrequent or nonexistent words: Implications for semantic processing of words in the absence of awareness. *PLoS ONE, 9*, e104719.

Hollingworth, A. (2006). Scene and position specificity in visual memory for objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 58-69.

Hollingworth, A. (2007). Object-position binding in visual memory for natural scenes and object arrays. *Journal of Experimental Psychology: Human Perception and Performance, 33*, 31-47.

Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 113-136.

Humphreys, G. W., Quinlan, P. T., & Riddoch, M. J. (1989). Grouping processes in visual search: Effects with single- and combined-feature targets. *Journal of Experimental Psychology: General, 118*, 258-279.

Jackson, M. C., & Raymond, J. E. (2008). Familiarity enhances visual working memory for faces. *Journal of Experimental Psychology: Human Perception and Performance, 34*(3), 556-568.

Jiang, Y., Costello, P., & He, S. (2007). Processing of invisible stimuli: Advantage of upright faces and recognizable words in overcoming interocular suppression. *Psychological Science*, 18, 349-355.

Kaiser, D., Stein, T., & Peelen, M. V. (2014a). Object grouping based on real-world regularities facilitates perception by reducing competitive interactions in visual cortex. *Proceedings of the National Academy of Sciences USA, 111*, 11217-11222.

Kaiser, D., Stein, T., & Peelen, M. V. (2015). Real-world spatial regularities affect visual working memory for objects. *Psychonomic Bulletin & Review*.

Kaiser, D., Strnad, L., Seidl, K. N., Kastner, S., & Peelen, M. V. (2014b). Whole person-evoked fMRI activity patterns in human fusiform gyrus are accurately modeled by a linear combination of face- and body-evoked activity patterns. *Journal of Neurophysiology, 111*, 82-90.

Kapadia, M. K., Ito, M., Gilbert, C. D., & Westheimer, G. (1995). Improvement in visual sensitivity by changes in local context: Parallel studies in human observers and in V1 of alert monkeys. *Neuron, 15*, 843-856.

Kastner, S., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1998). Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. *Science, 282*, 108-111.

Kastner, S., & Ungerleider, L. G. (2001). The neural basis of biased competition in the human visual cortex. *Neuropsychologia, 39*, 1263-1276.

Kaunitz, L., Fracasso, A., Lingnau, A., & Melcher, D. (2013). Non-conscious processing of motion coherence can boost conscious access. *PLoS ONE, 8*, e60787.

Kim, J. G., & Biederman, I. (2011). Where do objects become scenes? *Cerebral Cortex, 21*, 1738-1746.

Kim, J. G., & Biederman, I. (2012). Greater sensitivity to nonaccidental than metric changes in the relations between simple shapes in the lateral occipital cortex. *Neuroimage, 63*, 1818-1826.

Kim, J. G., Biederman, I., & Juan, C. H. (2011). The benefit of object interactions arises in the lateral occipital cortex independent of attentional modulation from the intraparietal sulcus: A TMS study. *Journal of Neuroscience,* 31, 8320-8324.

Knierim, J. J., & van Essen, D. C. (1992). Neural responses to static texture patterns in area V1 of the alert macaque monkey. *Journal of Neurophysiology, 67*, 961-980.

Koch, C. (2004). *The quest for consciousness: A neurobiological approach.* Colorado: Roberts & Company Publishers.

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences USA, 99*, 9596-9601.

Lin, Z., & He, S. (2009). Seeing the invisible: The scope and limits of unconscious processing in binocular rivalry. *Progress in Neurobiology, 87*, 195–211.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*, 279-281.

MacEvoy, S. P., & Epstein, R. A. (2009). Decoding the representation of multiple simultaneous objects in human occipitotemporal cortex. *Current Biology, 19*, 943–947.

Malach, R., et al. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences USA, 92*, 8135-8139.

Mandler, J. M., & Johnson, N. S. (1976). Some of the thousand words a picture is worth. *Journal of Experimental Psychology: Human Learning & Memory, 2*, 529-540.

McMains, S. A., & Kastner, S. (2010). Defining the units of competition: Influences of perceptual organization on competitive interactions in human visual cortex. *Journal of Cognitive Neuroscience, 22*, 2417-2426.

McMains, S. A., & Kastner, S. (2011). Interactions of top-down and bottom-up mechanisms in human visual cortex. *Journal of Neuroscience, 31*, 587-597.

Messinger, A., Squire, L. R., Zola, S. M., & Albright, T. D. (2001). Neuronal representations of stimulus associations develop in the temporal lobe during learning. *Proceedings of the National Academy of Sciences USA, 98*, 12239-12244.

Miller, E. K., Gochin, P. M., & Gross, C. G. (1993). Suppression of visual responses of neurons in inferior temporal cortex of the awake macaque by addition of a second stimulus. *Brain Research, 616*, 25-29.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

Montoro, P. R., Luna, D., & Ortells, J. J. (2014). Subliminal Gestalt grouping: Evidence of perceptual grouping by proximity and similarity in absence of conscious perception. *Consciousness and Cognition, 25*, 1-8.

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science, 229*, 782-784.

Mudrik, L., Breska, A., Lamy, D., & Deouell, L. Y. (2011). Integration without awareness: Expanding the limits of unconscious processing. *Psychological Science, 22*, 764-770.

Mullally, S. L., & Maguire, E. A. (2011). A new role for the parahippocampal cortex in representing space. *Journal of Neuroscience, 31*, 7441-7449.

Murray, S. O., Kersten, D., Olshausen, B. A., Schrater, P., & Woods, D. L. (2002). Shape perception reduces activity in human primary visual cortex. *Proceedings of the National Academy of Sciences USA, 99*, 15164-15169.

Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research, 46*, 614-621.

Olson, I. R., Jiang, Y., & Moore, K. S. (2005). Associative learning improves visual working memory performance. *Journal of Experimental Psychology: Human Perception and Performance, 31*, 889-900.

Olsson, H., & Poom, L. (2005). Visual memory needs categories. *Proceedings of the National Academy of Sciences USA, 102*, 8776-8780.

Palmer, S. E. (1999). *Vision science: Photons to phenomenology.* Cambridge, MA: MIT Press.

Peelen, M. V., & Kastner, S. (2014). Attention in the real world: Toward understanding its neural basis. *Trends in Cognitive Sciences, 18*, 242-250.

Peterson, D. J., & Berryhill, M. E. (2013). The Gestalt principle of similarity benefits visual working memory. *Psychonomic Bulletin & Review, 20*, 1282-1289.

Purves, D., Wojtach, W. T., & Lotto, R. B. (2011). Understanding vision in wholly empirical terms. *Proceedings of the National Academy of Sciences USA, 108*, 15588-15595.

Rauschenberger, R., & Yantis, S. (2006). Perceptual encoding efficiency in visual search. *Journal of Experimental Psychology: General, 135*, 116-131.

Reddy, L., Kanwisher, N. G., & VanRullen, R. (2009). Attention and biased competition in multi-voxel object representations. *Proceedings of the National Academy of Sciences USA, 106*, 21447-21452.

Riddoch, M., 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, 82-80.

Roberts, K. L., & Humphreys, G. W. (2010). Action relationships concatenate representations of separate objects in the ventral visual system. *Neuroimage, 52*, 1541-1548.

Roberts, K. L., & Humphreys, G. W. (2011). Action relations facilitate the identification of briefly-presented objects. *Attention, Perception, & Psychophysics, 73*, 597-612.

Roelfsema, P. R., & Houtkamp, R. (2011). Incremental grouping of image elements in perception. *Attention, Perception, & Psychophysics 73*, 2542-2572.

Rolls, E. T. (2000). Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition. *Neuron, 27,* 205-218.

Rolls, E. T., & Tovee, M. J. (1995). The responses of single neurons in the temporal visual cortical areas of the macaque when more than one stimulus is present in the receptive field. *Experimental Brain Research, 103,* 409-420.

Sakai, K., & Miyashita, Y. (1991). Neural organization for the long-term memory of paired associations. *Nature, 354,* 152-155.

Schmalzl, L., Zopf, R., & Williams, M. A. (2012). From head to toe: Evidence for selective brain activation reflecting visual perception of whole individuals. *Frontiers in Human Neuroscience, 6*, 108.

Schurz, M., Sturm, D., Richlan, F., Kronbichler, M., Ladurner, G., & Wimmer, H. (2010). A dual-route perspective on brain activation in response to visual words: Evidence for a length by lexicality interaction in the visual word form area (VWFA). *Neuroimage, 49,* 2649-2661.

Scolari, M., Vogel, E. K., & Awh, E. (2008). Perceptual expertise enhances the resolution but not the number of representations in working memory. *Psychonomic Bulletin & Review, 15*, 215-222.

Seidl, K. N., Peelen, M. V., & Kastner, S. (2012). Neural evidence for distracter suppression during visual search in real-world scenes. *Journal of Neuroscience, 32*, 11812-11819.

Simon, H. A. (1974). How big is a chunk? *Science, 183,* 482-488.

Smith, A. T., Singh, K. D., Williams, A. L., & Greenlee, M. W. (2001). Estimating receptive field size from fMRI data in human striate and extrastriate visual cortex. *Cerebral Cortex, 11*, 1182-1190.

Stanley, D. A., & Rubin, N. (2003). fMRI activation in response to illusory contours and salient regions in the human lateral occipital complex. *Neuron, 37*, 323-331.

Stein, T., End, A., & Sterzer, P. (2014a). Own-race and own-age biases facilitate visual awareness of faces under interocular suppression. *Frontiers in Human Neuroscience, 8*, 582.

Stein, T., Hebart, M. N., & Sterzer, P. (2011a). Breaking continuous flash suppression: A new measure of unconscious processing during interocular suppression? *Frontiers in Human Neuroscience, 5*, 167.

Stein, T., Kaiser, D., & Peelen, M. V. (2015). Inter-object grouping facilitates visual awareness. *Journal of Vision*, 15, 10.

Stein, T., Peelen, M. V., & Sterzer, P. (2011b). Adults' awareness of faces follows newborns' looking preferences. *PLoS ONE,* 6, e29361.

Stein, T., Senju, A., Peelen, M. V., & Sterzer, P. (2011c). Eye contact facilitates awareness of faces during interocular suppression. *Cognition, 119*, 307-311.

Stein, T., Seymour, K., Hebart, M. N., & Sterzer, P. (2014b). Rapid fear detection relies on high spatial frequencies. *Psychological Science, 25*, 566-574.

Stein, T., &, Sterzer, P. (2014). Unconscious processing under interocular suppression: Getting the right measure. *Frontiers in Psychology, 5*, 387.

Stein, T., Sterzer, P., & Peelen, M. V. (2012). Privileged detection of conspecifics: Evidence from inversion effects during continuous flash suppression. *Cognition, 125*, 64-79.

Stein, T., Thoma, V., & Sterzer, P. (2015). Priming of object detection under continuous flash suppression depends on attention but not on part-whole configuration. *Journal of Vision, 15*, 15.

Sterzer, P., Stein, T., Ludwig, K., Rothkirch, M., & Hesselmann, G. (2014). Neural processing of visual information under interocular suppression: A critical review. *Frontiers in Psychology, 5*, 453.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience, 19*, 109-139.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*, 520-522.

Tobon, R., Gronau, N., Scheuplein, A. L., Mecklinger, A., & Levy, D. A. (2014). Associative recognition processes are modulated by the semantic unitizability of memoranda. *Brain and Cognition, 92*, 19-31.

Torralba, A., Oliva, A., Castelhano, 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*, 766-786.

Treisman, A. M. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 194-214.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97-136.

Troiani, V., Stigliani, A., Smith, M. E., & Epstein, M. A. (2014). Multiple object properties drive scene-selective regions. *Cerebral Cortex, 24*, 883-397.

Tsuchiya, N., & Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nature Neuroscience, 8*, 1096-1101.

Tsuchiya, N., Moradi, F., Felsen, C., Yamazaki, M., & Adolphs, R. (2009). Intact rapid detection of fearful faces in the absence of the amygdala. *Nature Neuroscience, 12*, 1224-1225.

Vinckier, F., Dehaene, S., Jobert, A., Dubus, J. P., Sigman, M., & Cohen, L. (2007). Hierarchical coding of letter strings in the ventral stream: Dissecting the inner organization of the visual word-form system. *Neuron, 55*, 143-156.

Võ, M. L.-H., & Wolfe, J. M. (2013). The interplay of episodic and semantic memory in guiding repeated search in scenes. *Cognition, 126*, 198–212.

von der Heydt, R., Peterhans, E., & Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science, 224*, 1260-1262.

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological Bulletin, 138*, 1172-1217.

Walther, D. B., Caddigan, E., Fei-Fei, L., & Beck, D. M. (2009). Natural scene categories revealed in distributed patterns of activity in the human brain. *Journal of Neuroscience, 29*, 10573-10581.

Wang, L., Weng, X., & He, S. (2012). Perceptual grouping without awareness: Superiority of Kanizsa triangle in breaking interocular suppression. *PLoS ONE, 7*, e40106.

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt II. *Psychologische Forschung, 4*, 301-350.

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods, 42*, 671-684.

Wolfe, J. M., Alvarez, G. A., Rosenholtz, R. E., Kuzmova, Y. I., & Sherman A. M. (2011a). Visual search for arbitrary objects in real scenes. *Attention, Perception, & Psychophysics, 73*, 1650-1671.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception & Performance, 16*, 879-892.

Wolfe, J. M., Võ, M. L.-H., Evans, K. K., & Greene, M. R. (2011b). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Science, 15*, 77-84.

Wolfe, J. M., & Bennett, S. C. (1997). Preattentive object files: Shapeless bundles of basic features. *Vision Research, 37*, 25-43.

Woodman, G. F., Vecera, S. P., & Luck, S. J. (2003). Perceptual organization influences visual working memory. *Psychonomic Bulletin & Review, 10*, 80-87.

Yang E., Zald D. H., & Blake R. (2007). Fearful facial expressions gain preferential access to awareness during continuous flash suppression. *Emotion, 7*, 882-886.

Xu, Y. (2006). Understanding the object benefit in visual short-term memory: The roles of feature proximity and connectedness. *Perception & Psychophysics, 68*, 815-828.

Xu, Y., & Chun, M. M. (2007). Visual grouping in human parietal cortex. *Proceedings of the National Academy of Sciences USA, 104*, 18766-18771.

7 Supplementary Material

7.1 Supplementary Material - Study 1

7.1.1 Supplementary fMRI Results

Face-selective control regions. To further highlight the specificity of the results observed in PPA, we also analyzed activation profiles in fusiform face area (FFA; Kanwisher, McDermott, & Chun, 1997) and occipital face area (OFA; Gauthier et al., 2000). We defined the peaks of activation for FFA [average MNI coordinates; left: x = -40.3 (0.7), y = -50.0 (1.1), z = -20.8 (0.6); right: x = 41.9 (0.9), y = -49.7 (1.1), z = -21.9 (0.6); standard errors in brackets] and OFA [left: x = -42.1 (0.9), y = -78.0 (1.1), z = -13.1 (0.8); right: x = 43.3 (0.9), y = -76.9 (1.1), z = -13.8 (0.8)] using the face > object contrast from the functional localizer runs, thresholded at $p < .01$ (uncorrected). As in the main experiment, all analyses were carried out on 4mm spheres around the peak voxel (see Methods for details). In a repeated measures ANOVA with the factors of presentation order (simultaneous vs. sequential) and configuration (regular vs. irregular), we observed a main effect of presentation order, with higher responses in the sequential condition, in both FFA , $F(1,22) = 22.67$, $p < .001$ (Figure S1.1a) and OFA, $F(1,22) = 22.67$, $p <.001$ (Figure S1.1b). However, neither a main effect of configuration nor an interaction of the two factors was observed in both regions, all $F < 1.68$, $p >.20$. These results match the activation profile of object-selective LO, further indicating that the different PPA response profile is due to this region's selectivity for the house stimuli.

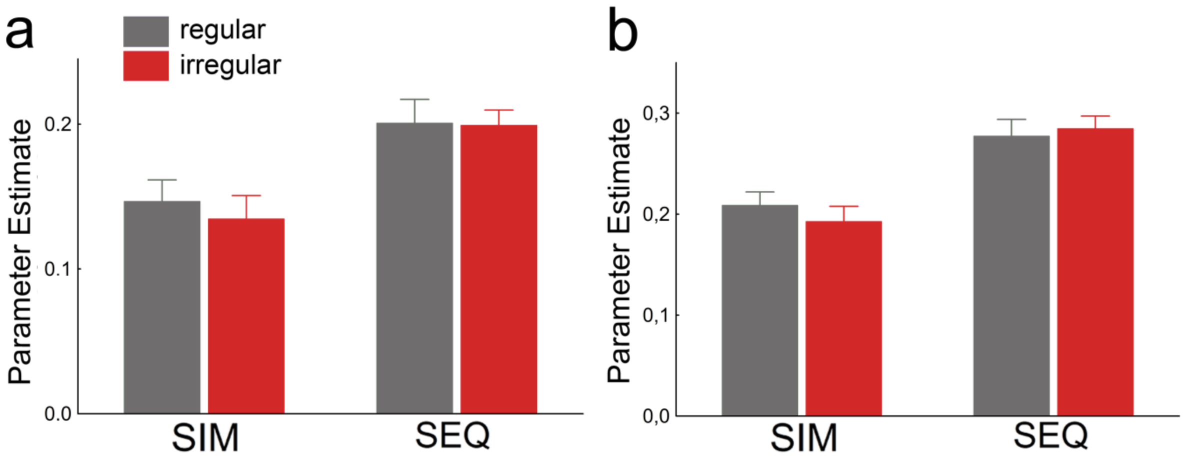

Figure S1.1. Data from face-selective FFA (a) and OFA (b). Both face-selective regions show reduced responses in the simultaneous condition, but no modulation connected to the regular versus irregular configuration.

Are the results in PPA explicable by anti-selectivity to objects? We defined PPA on the basis of the house>object contrast in our functional localizer runs. This way of defining the region could make it possible that the effects observed in PPA are not linked to its house-selectivity in the first place, but alternatively to its anti-selectivity for objects. To exclude this possibility, we repeated the analysis, we defined a new region (PPA*) on the basis of the house>face localizer contrast. Using again a liberal threshold of $p < .01$ uncorrected, we were able to define bilateral PPA* in 21 participants [average MNI coordinates; left: x = -25.1 (1.3), y = -43.8 (1.4), z = -8.2 (1.1); right: x = 26.7 (1.3), y = -45.1 (1.8), z = -8.6 (0.9); standard errors in brackets]. Again, a repeated measures ANOVA with the factors of presentation order (simultaneous vs. sequential) and configuration (regular vs. irregular) was carried out on the data from 4 mm spheres around the peak voxel. Importantly, this analysis revealed a significant interaction, $F(1,20) = 4.94$, $p = .038$ (Figure S1.2a), with no difference between regular and irregular object pairs in the sequential condition, $t(20) = 0.39$, $p = .70$, but lower responses for the irregular than the regular pairs in the simultaneous condition, *t*(20) = 1.86, p = .078. This pattern of results closely resembles the pattern of results obtained from conventionally defined PPA, and thus indicates that house-selectivity – rather than object anti-selectivity – is the key property that can explain the results in PPA. Additional evidence for this stems from the data obtained from the face-selective ROIs: If object anti-selectivity were the cause for the interaction observed in PPA, we should also observe such an interaction in FFA and/or OFA (which are also defined against objects as a control condition).

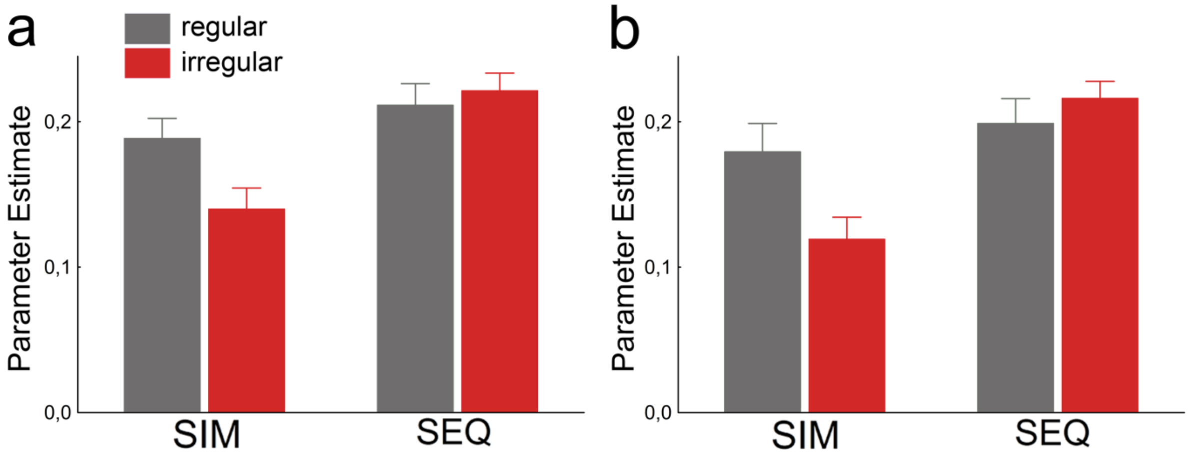

Figure S1.2. The results from PPA are not due to object anti-selectivity: The PPA response profile is preserved, even if the region is defined on the basis of a house>face contrast (a). The response profile is also maintained if – instead of a spherical ROI – only the peak voxel activation is used for the analysis (b).

PPA response profile based on peak voxel activation. We used relatively small spherical ROIs (containing 33 voxels) in the main experiment to ensure maximum PPA selectivity to the house stimuli used. Following this logic, similar modulations of PPA activity should be seen if we only analyze the response profile of the peak voxel. With this data, a repeated measures ANOVA with the factors of presentation order (simultaneous vs. sequential) and configuration (regular vs. irregular) showed a significant interaction, *F*(1,20) = 5.94, *p* = .023 (Figure S1.2b). As in the main experiment, we observed lower responses for the irregular than the regular condition in the simultaneous, *t*(20) = 2.26, *p* = .034, but not in the sequential condition, *t*(20) = 0.86, *p* = .40.

Event-related time courses. We also replicated the results of the ROI analysis using event-related time courses of the BOLD signal (Figure S1.3). To do so, we extracted the mean intensity values from the smoothed functional images for every condition, for both PPA and LO (using the same ROIs as in the main experiment). We computed these values for the eight TRs following the onset of the trial (rounded down to the nearest TR). Then, separately for each run, we subtracted the time course obtained for the fixation trials. To assess differences between conditions, we performed a repeated measures ANOVA with the factors of presentation order (simultaneous vs. sequential) and configuration (regular vs. irregular) on the mean value of the third and fourth TR after trial onset (representing the peak of the time course). In PPA, we found a significant interaction, *F*(1,22) = 22.89, *p* < .001, with lower responses for the irregular than the regular condition in the simultaneous, *t*(22) = 4.31, *p* < .001, but not in the sequential condition, *t*(22) = 1.53, *p* = .14. By contrast, we could not find a significant interaction in LO, *F*(1,22) = 1.04, *p* = .32; interaction including ROI, *F*(1,22) = 6.60, *p* = .018. These results nicely confirm the pattern of results obtained in the main experiment.

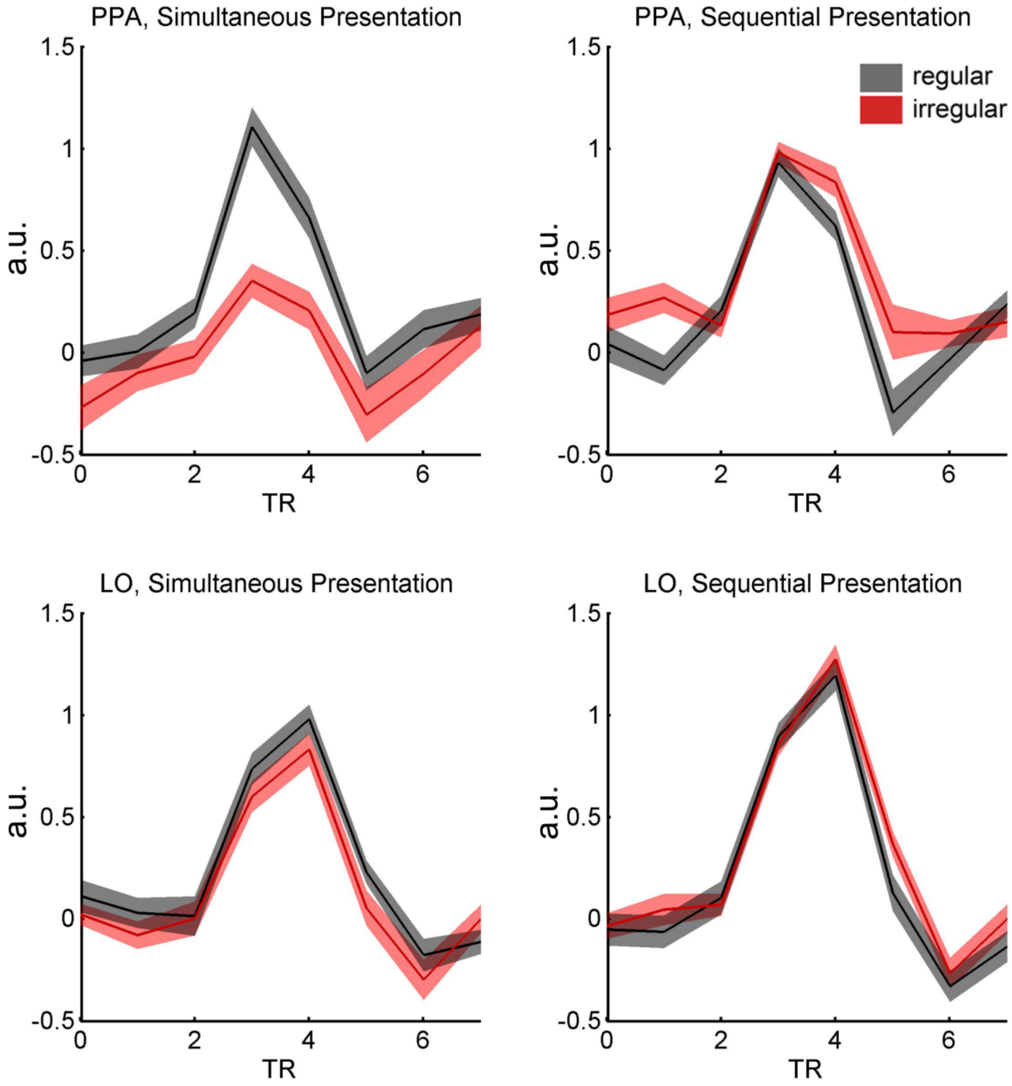

Figure S1.3. Event-related times courses of activation in PPA and LO. The pattern of results nicely confirms the interaction of presentation order and pair configuration in PPA, with a highly pronounced difference between the regular and irregular configurations in the simultaneous condition. By contrast, LO activity is not modulated by pair configuration.

Whole-brain analysis. To investigate whether there are regions outside the visual cortex ROIs reported before that show activity modulations, we conducted a whole-brain analysis. Similarly to the ROI analyses, we looked at the effects of presentation order (simultaneous vs. sequential) and configuration (regular vs. irregular). When we contrasted the simultaneous and sequential conditions ($p_{FDR} < .05$), we found that large clusters in visual cortex showed reduced responses in the simultaneous as compared to the sequential conditions. These clusters were spanning bilateral ventral and lateral occipital cortex, fusiform gyrus and parts of the parahippocampal gyrus (Figure S1.4). However, we were neither able to find a main effect of configuration nor an interaction, even at a liberal threshold of $p < .001$ (uncorrected), indicating that there are no other regions showing response modulations caused by the regularity manipulation, at least not at the whole-brain level.

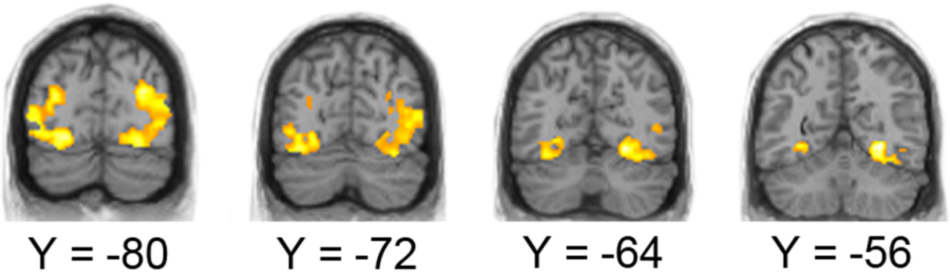

Figure S1.4. Bilateral visual areas show competitive interactions, indicated by reduced activity to simultaneously as compared to sequentially presented displays (whole brain-analysis, $p_{FDR} < .05$).

7.1.2 Supplementary Visual Search Experiment

In our visual search experiments, we used an accuracy measure to quantify the efficiency of search among regular and irregular distracters. Similar approaches have been used in numerous previous studies (e.g., Estes, 1964; Schoonveld, Shizomaki, & Eckstein, 2007; Shiffrin & Gardner, 1972). These studies have revealed important insights about visual search in the absence of overt attention (Eckstein, 2011) and have informed models of parallel attentional allocation (Bundesen, 1990; Bundesen, Habekost, & Kyllingsbaek, 2005). However, another major branch of the visual search literature refers to search efficiency specifically as the search time to set size relation (Neisser, 1964; Treisman & Gelade, 1980; Wolfe, 1998). Here, the higher the additional search time when adding an item to a display (i.e., the steeper the search slope), the less efficient the search process is considered. In an additional visual search experiment, we aimed to demonstrate that the benefit of real-world object regularities is not only visible in higher search accuracy, but can also be observed in shallower search slopes in a reaction-time-based variant of the task.

Participants. Eleven participants (1 male, mean age 22.3 years, *SD* = 2.0) volunteered for the experiment.

Stimuli. All stimuli were identical to the ones used in the accuracy-based search experiments.

Procedure. We used a similar design as in Visual Search Experiment 1 (see Figure 1.2a), but made a number of changes that allowed us to estimate search efficiency based on reaction times. Again, a word cue was indicating the target on every trial. However, the task was now changed to a present/absent forced-choice judgment: On 50% of trials, the cued target was present, while on the 50% of trials it was absent. The search display stayed on the screen until a response was given and participants were instructed to respond as fast as possible on every trial (while responses were non-speeded in the accuracy-based experiments). Importantly, to be able to estimate search slopes we also manipulated set size: We either presented two pairs and one single object on one side of fixation (i.e., a total of five objects; see Figure S1.5a), or we doubled the number of objects and presented two pairs and a single object on either side of fixation (i.e., 10 objects; see Figure S1.5b). To make the potential target locations less predictable, in contrast to the accuracy-based experiments the location of the target and distracter objects was no longer restricted to the central locations, and no mirroring of the positions was applied between hemifields. The experiment was split into blocks of 36 trials, in which the distracter pairs always appeared either in regular or irregular configurations, while the two set sizes and target absent/present trials were randomly intermixed.

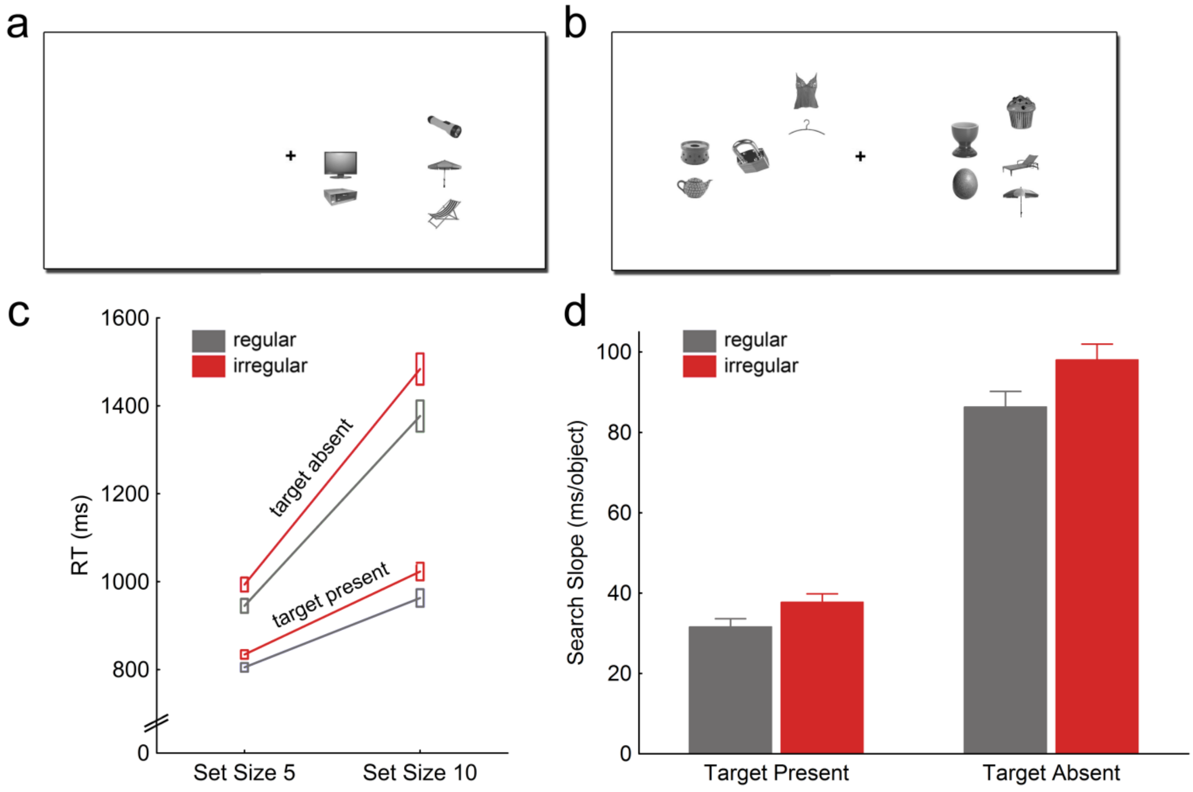

Figure S1.5. Real-world regularities increase efficiency also in speeded search. a) Example display for the set size 5 condition with regular pair configurations. b) Example display for the set size 10 condition with irregular pair configurations. c) Reaction times were generally lower in the regular condition than in the irregular condition, both in target absent and target present trials, with a more pronounced effect in the set size 10 condition. d) Search slopes were increased when the distracters were presented in regular configurations, both in target present and target absent trials.

Results and Discussion. We analyzed reaction times in a three-factorial ANOVA with the factors of pair configuration (regular vs. irregular), set size (5 vs. 10 objects), and target presence (present vs. absent). We found significant main effects of target presence, $F(1,10) = 62.7$, $p < .001$, and set size, $F(1,10) = 42.7$, $p < .001$, indicating slower responses for the larger set size and in target absent trials, with a higher difference between target absent and target present trials in the set size 10 condition, interaction: *F*(1,10) = 38.98, *p* < .001 (Figure S1.5c). We also found a significant main effect of pair configuration, *F*(1,10) = 21.08, *p* < .001, showing faster responses for regular than for irregular pairs. Importantly, if regular distracter pairs led to higher search efficiency, we would expect an interaction of pair configuration and set size, indicating differing search slopes between the two conditions. This interaction was significant, *F*(1,10) = 5.81, *p* = .037, with an effect of pair configuration in both set sizes, both *t*(10) > 6.67, *p* < .001, but a larger benefit for regular pairs in the set size 10 condition. To confirm this, we calculated the set size to reaction time slopes (as the reaction time difference between the set size 10 and set size 5 conditions divided by 5), and found greater slopes for irregular than for regular distracters (Figure S1.5d). When we analyzed target detection accuracy, we found no main effect of, or interaction with, pair configuration, all *F*(1,10) < 0.51, *p* > .49. Altogether, we were able to show that real-world distracter regularities also have a beneficial impact, which also shows up when investigating the reaction time to set size slopes in a speeded search task. These data confirm the results of our accuracy-based experiments, showing that real-world objects regularities among distracters can increase visual search efficiency.

7.1.3 Supplementary References

Bundesen, C. (1990). A theory of visual attention. *Psychological Review, 97*, 523-547.

Bundesen, C., Habekost, T., & Kyllingsbaek, S. (2005). A neural theory of visual attention: Bridging cognition and neurophysiology. *Psychological Review, 112*, 291-328.

Eckstein, M. P. (2011). Visual search: A retrospective. *Journal of Vision, 11*, 14.

Estes, W. K. (1964). A detection method and probabilistic models for assessing information processing from brief visual displays. *Science, 144*, 562.

Gauthier, I., Tarr, M. J., Moylan, J., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). The fusiform 'face area' is part of a network that processes faces at the individual level. *Journal of Cognitive Neuroscience, 12*, 195-504.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17*, 4302-4311.

Neisser, U. (1964). Visual Search. *Scientific American, 210*, 94-102.

Schoonveld, W., Shimozaki, S. S., & Eckstein, M. P. (2007). Optimal observer model of single-fixation oddity search predicts a shallow set-size function. *Journal of Vision, 7*, 1.

Shiffrin, R. M., & Gardner, G. T. (1972). Visual processing capacity and attentional control. *Journal of Experimental Psychology, 93*, 72-82.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97-136.

Wolfe, J. M. (1998). What can 1,000,000 trials tells us about visual search? *Psychological Science, 9*, 33-39.

7.2 Supplementary Material - Study 2

7.2.1 Pair-Related Effects

To explore the contribution of single pairs to the VWM effect, we looked at performance in the VWM task as a function of the presence of each of the single pairs. We computed sensitivity scores for all relevant regular and irregular conditions in Experiments 1 - 3 (all trials from Experiment 1, upright trials from Experiment 2, and all trials from Experiment 3), for all displays that contained a particular object pair. For every pair, we then computed the regularity effect, as the sensitivity difference between the regular and irregular conditions. A one-way ANOVA on these values across the three experiments revealed no significant difference between the different object pairs, *F*(11,1243) = 1.14, *p* = .328, indicating that there were no differences among object pairs in generating the effect of regularity in these experiments (Figure S2.1).

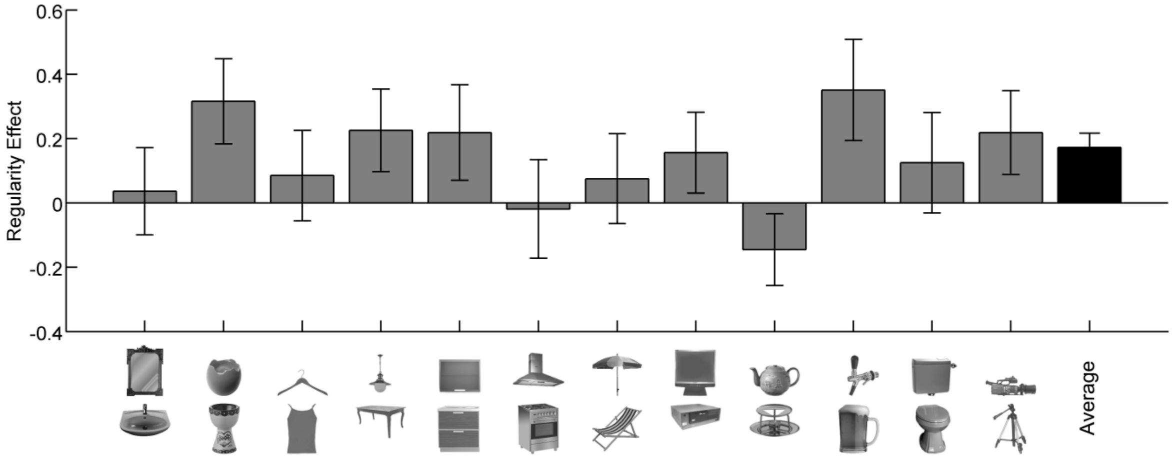

Figure S2.1. The VWM regularity effect is consistent across pairs: No significant difference in regularity effects (sensitivity for regular – sensitivity for irregular configurations) was observed across pairs. Error bars reflect SEM.

7.2.2 “Shuffled” Control Experiment

Another possible explanation for the results obtained in Experiments 1 - 3 is that single objects are memorized more efficiently if they are presented in their typical real-world locations. For example lamps are generally hanging from the ceiling and hence are encountered most often in upper parts of the visual field. As a consequence, they might be better memorized when presented in the upper position within a pair, regardless of the category of the second object. To test for the contribution of such single-object location effects, we recombined the top and bottom objects of different pairs such that the positions of single objects were preserved but the pair configuration was disrupted. If the regularity effect observed in Experiments 1 - 3 was caused by the positions of single objects, rather than object *pairs* being positioned according to their real-world configuration, the same effect of regularity would be expected for these shuffled pairs.

Participants. Twenty-two healthy adults (3 male; mean age 24.4 years, *SD* = 7.6) participated, of which 3 also participated in Experiment 1, 2 participated in Experiment 2, and 2 participated in Experiment 3.

Stimuli and Apparatus. The apparatus, stimuli, and setup were identical to Experiments 1 - 3. In addition to varying the pair configuration, we included a shuffled condition, in which we recombined the pairs. For this shuffled condition the top object of each pair was recombined with the bottom object of another arbitrarily selected pair (e.g., a parasol above a toilet; see Figure S2.2a). Thus, in the shuffled condition the configuration of the object pairs was disrupted while the single objects making up the pairs were again presented either in regular or irregular positions.

Procedure. We used the same design as in Experiment 2, but we replaced the inverted condition with the shuffled condition, where all top and bottom objects of the pairs were recombined. This again led to a total of 192 trials (48 per condition), with the two configuration and two pair type (original vs. shuffled) conditions being randomly intermixed.

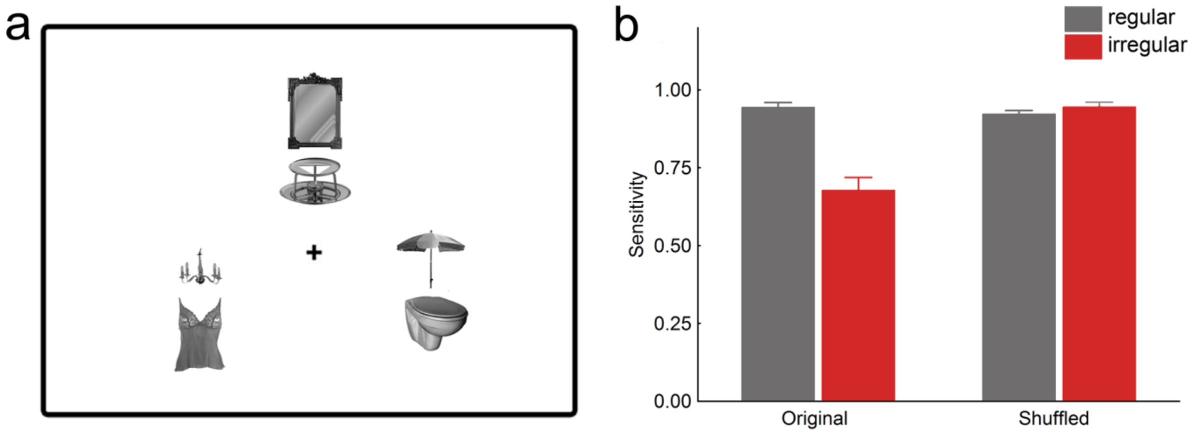

Figure S2.2. The VWM regularity effect is abolished for shuffled pairs, where individual items are recombined into pairs. a) An example display from the regular shuffled condition. b) While again the effect of regularity was replicated ("original" condition), there was no effect of pair configuration in the shuffled condition. Note that absolute performance is not directly comparable across the original and shuffled conditions due to the different pairs presented in these conditions. Standard errors reflect within-subject SEM (Cousineau, 2005).

Results and Discussion. Trials with incorrect responses in the verbal suppression task (4.5%, *SD* = 4.1) were excluded from the analysis. A 2-factorial ANOVA on the d-prime scores with the factors of pair type (original vs. scrambled) and pair configuration (regular vs. irregular) revealed no significant main effects, both $F(1, 21) < 2.96$, $p > .10$, but a significant interaction, $F(1,21) = 5.26$, $p = .032$: For the original pairs sensitivity was significantly higher for the regular than for the irregular configuration, $t(21) = 2.60$, $p = .017$. By contrast, for the shuffled pairs there was no difference in sensitivity between the two pair configurations, $t < 1$ (Figure S2.2b). Again, accuracy in the verbal suppression task was comparable between conditions, all $F(1,21) < 2.43$, $p > .135$. These results offer evidence that single object positions alone cannot account for the VWM enhancement observed for regular as compared to irregular pairs.

Although the most important result from this experiment is the difference between the regular and irregular pairs in the original but not the shuffled conditions, we also observed lower change-detection sensitivity in the original irregular condition as compared to all other conditions, all $t(21) > 2.19$, $p < .039$. At first glance, this result suggests an irregularity cost rather than a regularity benefit. While we cannot exclude this interpretation of the observed regularity effect, we favor an interpretation related to a VWM enhancement due to real-world regularities, in line with research on regularity effects with simple stimuli. Importantly, the data of the shuffled control experiment is not inconsistent with this “regularity benefit” interpretation. This is because performance levels across the original and shuffled conditions are not directly comparable – the stimuli in both conditions differed physically due to the recombination of pairs. More specifically, the shuffling of objects might have led to more easily detectable changes for the pairs in the shuffled conditions. This could happen, for example, if changes to the two objects making up an original pair were both relatively easy or both relatively hard to detect (e.g., the upper and lower kitchen cabinets); recombining these objects would result in more pairs with at least one “easy” object. Alternatively, it is of course also possible that there is both a cost related to the explicit disruption of real-world regularities, and a benefit for regularly positioned objects. As the present data cannot conclusively decide between these possibilities, further research is needed to disentangle the contribution of these two potential mechanisms.

Altogether, we conclude that the VWM regularity effect is best described as an interactive effect between objects, taking into account both object identity and position.

We suggest that this effect stems from a benefit for regularly positioned objects, possibly combined with decreased VWM performance for irregularly positioned objects.