Neuroscience of Consciousness, 2018, 4(1): niy001

doi: 10.1093/nc/niy001
Research article

Contextual cueing of visual search is associated with greater subjective experience of the search display configuration

Bernhard Schlagbauer[1,2,*,†], Manuel Rausch[1,2,3,†], Michael Zehetleitner[1,3], Hermann J. Müller[1,4] and Thomas Geyer[1]

[1]Department Psychologie, Ludwig-Maximilians-Universität München, Munich, Germany; [2]Graduate School of Systemic Neurosciences, Groβhaderner Str. 2, 82152 Planegg-Martinsried, Germany; [3]Fakultät für Psychologie und Pädagogik, Fachgebiet Psychologie II, Katholische Universität Eichstätt-Ingolstadt, Ostenstraβe 25, 85072 Eichstätt, Germany; [4]Department of Psychological Sciences, Birkbeck, University of London, Malet Street, London WC1E 7HX, UK

*Correspondence address. Department Psychologie, Lehrstuhl für Allgemeine und Experimentelle Psychologie, Ludwig-Maximilians-Universität München, Leopoldstraße 13, 80802 München, Germany. Tel: 0049-(0)89-2180-16587; Fax:0049-(0)89-2180-5211; E-mail:bernhard.schlagbauer@psy.lmu.de
†These authors contributed equally to this work.

Abstract

Visual search is facilitated when display configurations are repeated over time, showing that memory of spatio-configural context can cue the location of the target. The present study investigates whether memory of the search target in relation to the configuration of distractors alters subjective experience of the visual search target and/or the subjective experience of the display configuration. Observers performed a masked localization task for targets embedded in repeated vs. non-repeated (baseline) arrays of distractors items. After the localization response, observers reported their subjective experience of either the target or the display configuration. Bayesian analysis revealed that repeated displays resulted in a stronger visual experience of both targets and display configurations. However, subsequent analysis showed that repeated search displays increased the correlation between the experience of the display configuration and localization accuracy, but there was no such effect on experience of the target stimulus. We suggest that memory of visual context enhances the representation of the current visual search display. This representation improves visual search and at the same time increases observers' subjective experience of the display configuration.

Key words: contextual cueing; visual search; perception; implicit learning; subjective experience

Introduction

In everyday scenes, visual search targets do not appear in isolation but are embedded within configurations of non-target or distractor objects. When observers encounter a target consistently embedded within a stable spatial configuration of distractors, target detection becomes more efficient over time, because incidentally learned configurations expedite visual search, an effect referred to as contextual cueing (Chun and Jiang 1998; Chun 2000). A controversial issue in research on contextual cueing is whether the effect is implicit or explicit (Chun and Jiang 2003; Smyth and Shanks 2008; Schlagbauer *et al.* 2012; Vadillo *et al.* 2015; Colagiuri and Livesey 2016). The present study

Received: 22 February 2016; **Revised:** 1 December 2017; **Accepted:** 30 January 2018

investigated a new aspect of awareness in contextual cueing, namely, whether learned spatial contexts modulate subjective experience of the display configuration in addition to subjective experience of the target item.

Contextual cueing of visual search

In their pioneering study, Chun and Jiang (1998) asked their observers to search for a T-shaped target amongst L-shaped distractor items. Unknown to observers, the spatial configurations of targets and distractors were repeated in half of the trials, while in so-called non-repeated displays, shown in the other half of trials, only the locations of targets were held constant across repetitions. Thus, the effects of target location repetition (see e.g. Jiang *et al.* 2013; Schlagbauer *et al.* 2014) were equated across the two types of displays and differences in search performance could only be attributed to the effects of repeated distractor configurations. Reaction times (RTs) decreased with more practice on the experimental task, but this effect was larger for repeated compared to non-repeated displays (=contextual cueing effect). As observers' ability to recognize repeated displays was only at chance level, Chun and Jiang (1998) concluded that contextual cueing is an implicit effect. However, in recent years the question whether the cueing effect is inaccessible to awareness has become a controversial issue. For instance, a meta-analysis of performance in recognition tasks demonstrated that participants in contextual cueing experiments perform above-chance level (Vadillo *et al.* 2015), suggesting that previous non-significant results were likely due to insufficient statistical power of the individual recognition tasks. These observations are consistent with theories according to which all learning processes are associated with some degree of awareness, including repeated displays (Smyth and Shanks 2008). However, a follow-up study again challenged the view of a single memory system in contextual cueing of visual search (Colagiuri and Livesey 2016). The authors used large samples and found that contextual cueing was associated with weaker, not stronger, recognition of learned visual search displays. This led Colagiuri and Livesey (2016) to surmise that contextual cueing is supported by an implicit memory system.

Contextual cueing and subjective visual experience

In the present article, we investigate whether context memory has the capability to affect other processes than visual search, specifically, whether it influences observers' subjective experience of visual properties of the current search displays. Contextual cueing may influence subjective visual experience in at least two distinct ways. First, context memory might alter the subjective visual experience ("clarity") of the configuration of display elements. Previous studies showed that when observers are presented with repeated display configurations, they learn to associate the target with the entire configuration of distractor elements (Jiang and Wagner 2004), though target-context associations are particularly strong for distractors in close spatial proximity of the target (Brady and Chun 2007). When the repeated search displays are encountered later on, spatio-configural memory representations make visual search more efficient, for example, by guiding attention faster towards the target location (Johnson *et al.* 2007). Crucially, at the same time, these context representations could also enhance observers subjective experience of the display configuration. For instance, observers might feel that they see a clearer configuration of display elements when these configurations are stored in context memory. However, contextual cueing could also reduce subjective experience of the configuration, as previous studies could not rule out a reversed, that is, negative relationship between context learning in visual search and the conscious recollection of the repeated displays (Colagiuri and Livesey 2016).

Second, contextual cueing might also influence the subjective visual experience of the target stimulus. The reason for this hypothesis is that contextual cueing can speed up processes after visual selection, for example, the perceptual analysis of the target (which is necessary for performing a discrimination task; see Töllner *et al.* 2013) and/or response selection (Kunar *et al.* 2007; Hout and Goldinger 2012). Because context memory influences later stages of the search process, it could be expected that these processes do not alter visual experience of the display configuration, but instead increase subjective experience associated with the target stimulus. In other words, context memory may exert a specific influence on observers' experience of the target stimulus.

Measuring subjective visual experience by verbal reports

Even though objective measures of memory had dominated cognitive psychology for many years (Boring 1953; Eriksen 1960), many researchers from different theoretical perspectives have argued for measuring conscious experience using subjective measures (Dehaene and Naccache 2001; Ramsøy and Overgaard 2004; Lau 2008; Seth *et al.* 2009). One approach based on subjective measures is to ask observers about their confidence in being correct, because it can be expected that participants determine their confidence judgement on the basis of all information they are aware of and consider as relevant for performing the task (Dienes and Seth 2010). A second approach is to ask observers directly to introspect and report their conscious experience. The most frequently used scale is the "Perceptual Awareness Scale," which requires observers to report visual clarity of the stimulus, feelings of "something being shown," and feelings of "certainty" (Ramsøy and Overgaard 2004, 12). Other scales required observers only to report the visibility of the stimulus (Sergent and Dehaene 2004; Rausch and Zehetleitner 2016). Directly asking observers about their visual experience of specific stimuli seems to be the most suitable approach for the present study because it enables us to differentiate between visual experiences related to the target and those related to the display configuration. However, verbal reports about experience are often dismissed as scientific data because they lack a verifiable ground truth: there is no way one can establish the "true" conscious experience of the observer. The problems of missing ground truth can be circumvented by quantifying the degree to which verbal reports predict performance in the task. If participants report conscious experiences relevant for solving the experimental task, their reports should differentiate between correct and incorrect responses.

Only one single previous study investigated the effect of repeated spatial configurations on observers' verbal reports about their visual experience of the target elements. Schlagbauer *et al.* (2012), using masked displays and verbal reports after each single trial, observed that repeated spatial context was associated with as clearer visual experience of the target stimulus as well as higher confidence in target localization judgments. However, the study did not require observers to report their experience of the display configuration. Further, the correlation between verbal reports and localization performance was not analyzed.

Rationale of the present study

The present study addressed the issue of whether memory of spatial context acquired during visual search in repeated display configurations influences the subjective experience of the display configuration or the target stimulus. Further, we assessed the relationship between observers' subjective reports about visual experience and their objective task performance. The experiment consisted of two parts: In the first part, participants had to localize a target letter "T" presented among distractor "L" letters. To induce variation in visibility, search displays were masked (see Fig. 1). Half of the displays were repeated displays: there were 12 different, but fixed target-distractor configurations presented in each block of the experiment. The other half were non-repeated displays, in which distractor locations were determined randomly at the beginning of each trial. Following observer's localization response, one group of participants rated the perceptual clarity of the display configuration, and another group the perceptual clarity of the target. In a third "control" group, observers performed only the target localization task to examine if contextual cueing was affected by the concurrent assessment of participants' search performance and verbal reports in each trial. In a second, consecutive part of the experiment, participants in all the three groups performed a short visual search task with unmasked displays. Participants were instructed to discriminate the orientation of the target letter "T" embedded in distractor "L" letters, the "standard" procedure in contextual cueing studies (for a review see, e.g. Goujon *et al.* 2015). In this task, no reports were made about the clarity of the display configuration or the target item. Instead, only RTs were recorded. The discrimination task served as a secondary check for the effects of the concurrent assessment of observers' verbal reports and search RTs on contextual cueing performance. The association between observers subjective reports and their search task performance was measured using Type 2 ROC analysis (Fleming *et al.* 2010; Galvin *et al.* 2003; see Supplementary Material).

If context memory was able to influence subjective visual experience, then the correlation between verbal reports and localization accuracy should increase over time for repeated over non-repeated displays in at least one of the two conditions, requiring observers either to report on the distractor configuration or the target stimulus. More specifically, if memory of search configurations affected subjective experience of the display configuration, then it might be expected that reports about the display configuration become more predictive of observers' search performance in repeated displays. Alternatively, or in addition, context memory of the display configuration may affect subjective experience of the target item. Then, repeated displays should give rise to a stronger correlation between subjective reports of the target item and search performance in repeated displays. Bayes factors were used for statistical testing, as both the presence and the absence of the effects are of theoretical interest (Rouder *et al.* 2009; Dienes 2011; Wetzels *et al.* 2011).

Experiment

Methods

Participants

A total of 45 observers took part in the experiment (11 male; 1 left-handed, mean age: 25.7 years). All participants reported normal or corrected-to-normal vision and provided written informed consent prior to the experiment. Participants received either €12 or course credit for their participation. The experiment was conducted according to the principles expressed in the Declaration of Helsinki (World Medical Association 2013).

Apparatus and stimuli

The experiment was conducted in a dimly lit room and run on a PC under the Windows XP operating system. The experiment was programed in MATLAB with the Psychtoolbox extension for stimulus presentation (Brainard 1997; Pelli 1997). Participants were seated in front of a 19″ CRT monitor [display resolution: 1024×768 pixels; refresh rate: 85 Hz (AOC, Amsterdam, The Netherlands)] at a viewing distance of approximately 60 cm. Search displays always consisted of one target T-shape among 11 distractor L-shapes. All 12 items in the search displays were dark gray (1.0 cd/m^2, $0.47^\circ \times 0.47^\circ$ in size) and presented against a light gray background (25.4 cd/m^2). The items were scattered inside an area of 9.28° visual angle in a way that item density and display extension was as comparable as possible across search displays. Items were positioned at pseudo-randomly chosen locations on four (imaginary) concentric circles around the display center (radii: 2.32°, 4.64°, 6.96°, and 9.28°). The position of items was constrained by a minimum distance between two adjacent items of 2.32°, at least one item on each circle and an equal number of items in each quadrant. These restrictions ensured that search displays were comparable in terms of item eccentricity and item density and that there was no guessing bias regarding the target quadrant. The "T"-shaped target stimulus was oriented randomly either 90° or 270° from the vertical midline and always appeared on the third circle from the display center but never on the horizontal or the vertical midline. There were 24 possible target locations on the third circle, of which 12 were used with repeating display configurations and 12 for the random configurations. The 11 "L"-shaped distractors were positioned at random locations on the four circles (with the restrictions above) and tilted either 0°, 90°, 180°, or 270°. In Part 1 of the experiment (localization task), the search displays were masked shortly after presentation by figure-8 shapes placed along eight concentric circles around the display center, covering the whole area of possible item locations.

Task and procedure

The sequence of events in each trial is illustrated in Fig. 1. Participants were randomly assigned to the configuration-, stimulus-, or control condition. The conditions differed only in the type of verbal reports. The procedure of the experiment, the behavioral tasks, and their order was the same across the three conditions (groups).

In the initial localization task, each trial started with the presentation of a fixation cross in the center of the screen for 1500 ms, followed by a blank interval of 200 ms. Next, the search display was presented for an individually adjusted stimulus onset asynchrony (SOA; see Supplementary Material for the staircase procedure) until it was masked by the figure-8 shapes. Participants were asked to indicate in which quadrant of the screen the target was localized using the keys on the numeric key pad of a standard computer keyboard with their right hand ("1" for the lower left, "3" for the lower right, "7" for the upper left, and "9" for the upper right quadrant). Following participants' localization response and a blank interval of 200 ms, a question appeared on screen. In the configuration condition, participants were asked: "How clearly did you see the configuration"; in the stimulus condition participants were asked: "How clearly did you see the T?" The questions of both experimental conditions were presented together with a scale from 1

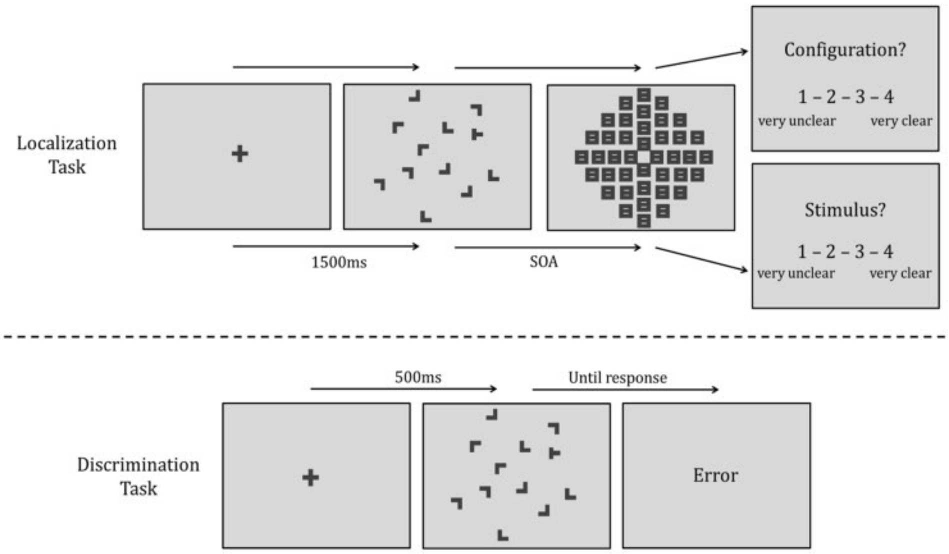


Figure 1. Sequence of events in the localization task (top panel) and discrimination task (bottom panel). In the localization task, participants were required to respond to the screen quadrant of the target stimulus as accurate as possible. The display was presented for an individually adjusted SOA (aiming at 75% correct performance), before it was masked by figure-8-shapes. After their localization response, participants in the configuration condition were asked to rate how clearly they saw the configuration and participants in the stimulus condition were asked to rate the clarity of the target stimulus. No questions were administered in the control condition. After the localization task, participants continued with a discrimination task, which was identical for the three rating conditions. In this task, the displays were visible until observers responded to the orientation of the target as fast and as accurate as possible. No masking occurred and no reports were collected in the discrimination task.

("very unclear") to 4 ("very clear"). Reports were given by participants pressing the corresponding key ("1", "2", "3", or "4" key) on the keyboard using their left hand. After the verbal report, the next trial started with a blank interval of 200 ms. No question was asked in the control condition, while the inter-trial interval was prolonged to 1000 ms. No feedback was given.

In the later discrimination task, trials started with a fixation cross for 500 ms, followed by a blank interval of 200 ms. When the search display appeared, observers were asked to respond to the orientation (to the left or the right) of the target stimulus as fast and as accurate as possible by pressing the left or right arrow key using the corresponding index finger. After a correct response, the next trial started after a blank interval of 500 ms. An erroneous response resulted in the display of the word "Error" in the center of the screen for 1000 ms.

The localization task consisted of 480 trials, divided into 20 blocks of 24 trials each. In each block, 12 of the displays were repeated (repeated displays); the other 12 displays were generated randomly, with only the target position remaining constant across all trial blocks (non-repeated displays). The same repeated displays were used in the discrimination task, which consisted of 96 trials divided into 4 blocks of 24 trials each. Consequently, each repeated display was shown 20 times in Part 1 and 4 times in Part 2. Overall, the experiment took about 90 min.

The design of the experiment was a between-subject design, so that each participant only had to give ratings about the display configuration or the target stimulus (or no ratings at all). The between-subject design was chosen to ensure that verbal reports with different contents could not interfere with each other as well as holding task difficulty at a tolerable level. In the configuration condition, participants were asked to report the clarity of the configuration of the display after their localization response. They were instructed that configuration refers to the general outline of the search array, its form or shape and that they should report how well they perceived the display as a whole entity. In the stimulus condition, participants were asked to report the clarity of the target stimulus. They were instructed that this refers to the letter T only and that they should report how vividly they saw this item. In both conditions, participants were asked whether they understood the instruction, and this was also double checked at the end of the staircase procedure. The control condition was identical to the other two conditions except that no ratings were administered. The discrimination task was identical in all three conditions/groups.

Data analysis

The data from the localization task were collapsed into two epochs, with each epoch representing an average of 10 consecutive blocks, to obtain reasonably stable estimates of contextual cueing and the association between verbal reports and localization accuracy, the latter assessed by the area under Type 2 ROC curves (Fleming *et al.* 2010, see Supplementary Material for details). Type 2 ROC analysis quantifies the degree to which verbal reports predict trial accuracy independent of participants' propensity to report high visual experience. Type 2 ROC curves control for rating criteria unlike gamma correlation coefficients (Masson and Rotello 2009) and logistic regression (Rausch and Zehetleitner 2017) and can be calculated even when there are more than two response options, unlike meta-d' (Maniscalco and Lau 2012). Moreover, in Type 2 ROC analysis, no assumptions about the distributions of evidence in correct and incorrect trials has to be made (Fleming and Lau 2014).The data were analyzed using R (R Core Team 2014) Bayes Factors were calculated with the package "BayesFactor" (Morey and Rouder 2015).

Localization accuracy, verbal reports, and the relationship between verbal reports and localization accuracy were analyzed with ANOVA-equivalent Bayes factors using Bayesian linear models with report condition as between-subject factor (three levels for accuracy: configuration, vs. stimulus vs. control; only two levels for verbal reports and the association between verbal reports and localization accuracy: configuration vs. stimulus), display type (two levels: repeated, non-repeated), and epoch (two levels: epoch 1 vs. epoch 2) as within-subject factors. The Bayes Factor of each main effect or interaction is obtained by comparing a linear model including the effect of interest to a model where the effect is omitted. This procedure allowed us to include covariates in the linear models (as implemented in the R package "BayesFactor" by Morey and Rouder 2015). As priors, we used previously suggested default variance priors for linear models with a scale parameter of $\sqrt{2}/4$ (Rouder and Morey 2012). The evidence for or against an effect was considered as substantial if its Bayes Factor was larger than 3 or lower than 1/3 (Wetzels *et al.* 2011). As *post hoc* tests, we computed the Bayesian equivalent of a one-sided paired t-test comparing the association between verbal reports and localization accuracy between repeated and non-repeated displays. We assumed a Cauchy distribution of the standardized effect sizes with the scale parameter $r = \sqrt{2}/2$ over the interval 0 to ∞, which was suggested as a default prior in psychology (Morey and Rouder 2015).

In the discrimination task, incorrect responses were discarded from the analysis (overall error rate: 3.3%). RTs were analyzed with Bayes factors calculated analogously to the analysis of the localization task performance with the factors report condition (configuration, stimulus, control; between-subject variable) and display type (repeated, non-repeated; within-subject variable).

Results

Localization task

Localization accuracy

Figure 2 shows that observers' accuracy in the localization task increased over time (epochs). This improvement in performance was comparable across the three groups; however, it was larger for repeated over non-repeated displays, revealing a beneficial effect of learned spatial context on search performance. The Bayes factors for accuracy as dependent variable indicated main effects of display type [$BF_{10} = 73.17$] and epoch [$BF_{10} = 38.04$], but only anecdotal evidence for their two-way interaction [$BF_{10} = 2.56$]. A relatively early onset of the contextual cueing effect is not unexpected, given that the current localization task was split into only two epochs of 10 blocks each, while contextual cueing usually emerges already after approximately 4–6 repetitions/blocks (e.g. Chun and Jiang 1998). Importantly, there were no interactions of report condition and display type [$BF_{10} = 0.19$], as well as report condition, display type, and epoch [$BF_{10} = .17$]. There was no conclusive evidence regarding the main effect of report condition [$BF_{10} = 0.45$], but substantial evidence for the interaction of report condition and epoch [$BF_{10} = 4.91$]. Direct tests showed that localization accuracy was comparable across report conditions in both the first and second epoch (see Supplementary Material for details).

In sum, we observed a general improvement in localization accuracy with training, which was higher in the configuration and control relative to the stimulus condition (see Fig. 2). Further, there was context-dependent learning, reflected by higher localization accuracy in repeated over non-repeated displays. Crucially, context-dependent learning was comparable across groups, suggesting that the acquisition of context memory was not selectively influenced by the concurrent assessment of verbal reports (see Fig. 2; exact descriptive statistics are provided in the Supplementary Material).

Verbal reports

An analogous 2 (report condition) × 2 (display type) × 2 (epoch) analysis with verbal reports as dependent variable revealed substantial main effects of display type [$BF_{10} = 5.53$] and epoch [$BF_{10} = 216.03$]. There was anecdotal evidence for a main effect of report condition [$BF_{10} = 2.08$], and an interaction between display type and epoch [$BF_{10} = 1.28$]. The analysis also revealed the absence of interactions between report condition and display type [$BF_{10} = 0.11$], as well as between report condition and epoch [$BF_{10} = 0.13$]. The three-way interaction was inconclusive, trending towards evidence for its absence [$BF_{10} = 0.41$]. These results suggest the operation of context-dependent and context-independent (i.e. procedural) learning in the present task and mirror those from the analysis of localization accuracy. The clarity of the display configuration and the target identity was greater in repeated relative to non-repeated displays. This is the effect of context-dependent learning. Further, clarity ratings increased in general through extended practice on the task (effect of context-independent learning), and there was a trend for clarity reports being higher when observers had to report on target identity compared to display configuration (see Fig. 3).

Association between verbal reports and localization accuracy

Type 2 ROC analysis, used to measure the relation between verbal reports and localization accuracy, revealed a substantial three-way interaction of report condition, display type and epoch [$BF_{10} = 15.14$]. All main effects were inconclusive [report condition: $BF_{10} = 1.55$; display type: $BF_{10} = 0.83$; epoch: $BF_{10} = 0.56$]. The interactions of report condition and display type [$BF_{10} = 0.35$] and of epoch and display type [$BF_{10} = 0.40$] were inconclusive, although leaning towards the null hypothesis. There was substantial evidence against an interaction of report condition and epoch [$BF_{10} = 0.27$].

As depicted in Fig. 4, the average area under ROC curves was greater in repeated compared to non-repeated displays, but only in the configuration condition. A *post hoc* analysis performed for the target condition revealed anecdotal evidence for a null effect of display type in epoch 1 [$BF_{10} = 0.71$] and substantial evidence for a null effect of display type in epoch 2 [$BF_{10} = 0.27$]. For the configuration condition, in contrast, there was substantial evidence for a null effect of display type in epoch 1 [$BF_{10} = 0.27$] and substantial evidence for an effect of display type in epoch 2 [$BF_{10} = 3.70$]. This suggests that repeated contexts are associated with a greater area under the ROC curve in the "late" epoch 2 in case of reports about the display configuration. But context memory did not exert an influence on the area under the ROC curve for reports about the target item in epoch 2.—Two control analyses investigated whether the effects of contextual cueing on the area under the ROC curve for reports about the display configuration were modulated by performance improvements in the localization task (for details, see Supplementary Material). In the first control analysis, localization accuracy and RTs were included as covariates in the analysis of Type 2 ROC curves. In the second control analysis, we examined only participants who displayed comparable accuracy scores in their localization performance in the three report conditions. Both control analyses confirmed the results pattern depicted in Fig. 4: the area under ROC curve for ratings of the display configuration increased over time. In contrast, no such effect was observed for ratings about the target item.

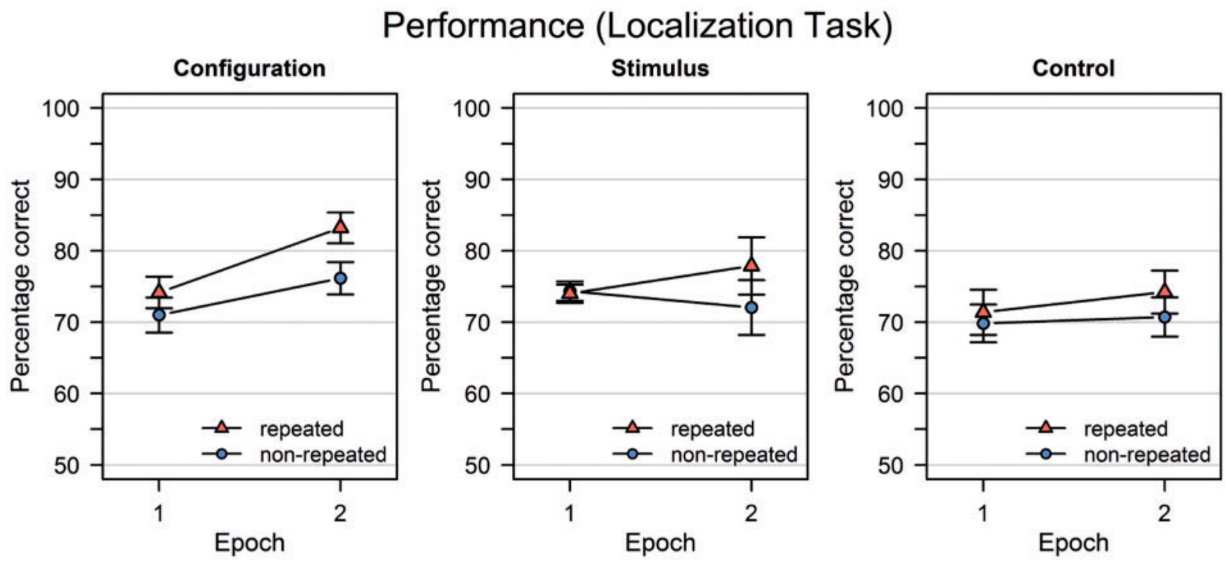


Figure 2. Mean localization performance as a function of epoch in the configuration, stimulus, and control condition (left, middle, and right panel, respectively).

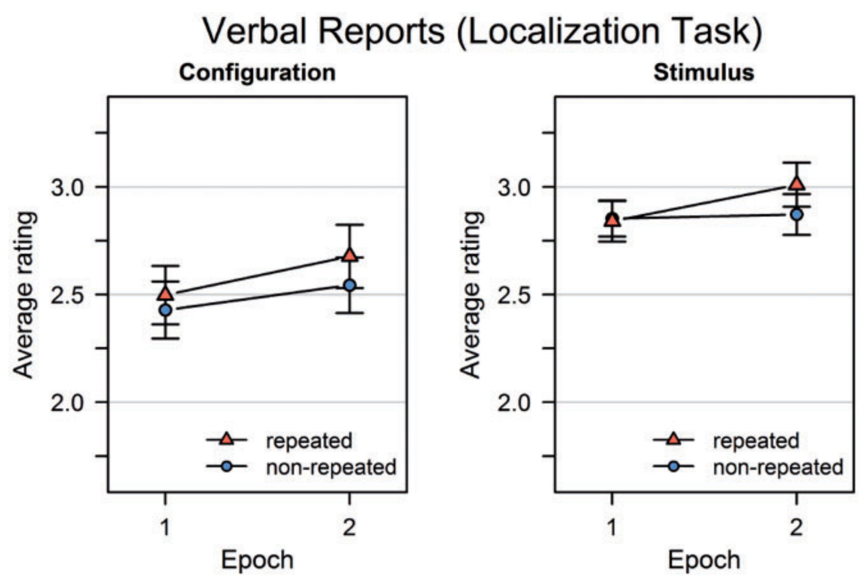


Figure 3. Mean verbal reports in the localization task as a function of epoch in the configuration and stimulus condition (left and right panel, respectively).

Discrimination task

Reaction times

The analysis of RTs revealed a substantial main effect of display type [$BF_{10}=441.62$], as well as evidence against the interaction between report condition and display type [$BF_{10}=0.09$]. The main effect of report condition was inconclusive [$BF_{10}=0.73$]. These results show that contextual learning acquired initially in a localization task is able to transfer to a subsequent discrimination task. Crucially for the present investigation, the transfer effects were comparable across the three report conditions (see Fig. 5). This would also mean that the processes of learning the repeated displays were equally efficient across the three report conditions.

Discussion

The present experiment investigated the effect of learned spatial configurations on subjective visual experience of the display configuration and the target item in a visual search task. We observed that the learning of repeated target-distractor configurations was associated with an increase of visual experience of the display configurations as well as a greater association of subjective experience of the display configurations and localization accuracy. Concerning the visual experience of the target, context memory did not modulate the correlation between target ratings and localization performance, though learned contexts were associated with greater visual experience of the target item.

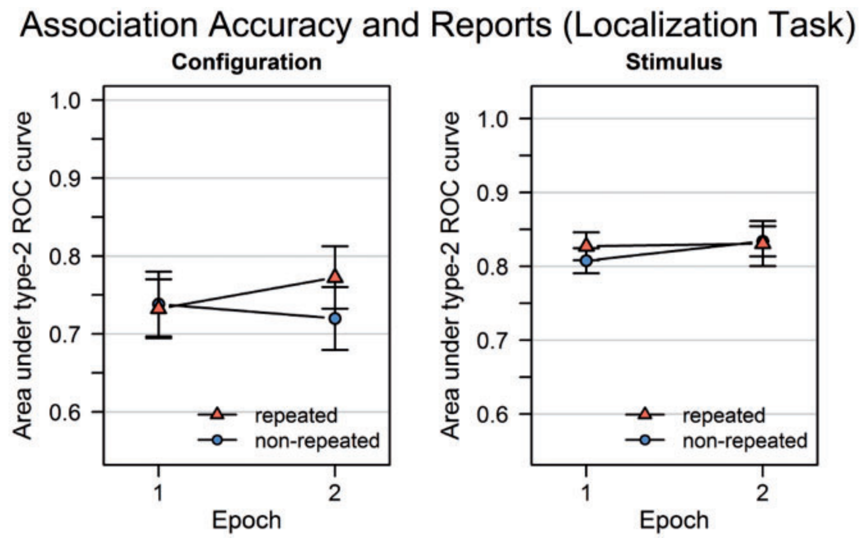


Figure 4. Area under the Type 2 ROC curve quantifying the association between accuracy in the localization task and configuration reports (left panel) and target reports (right panel) as a function of epoch and context (repeated vs. non-repeated).

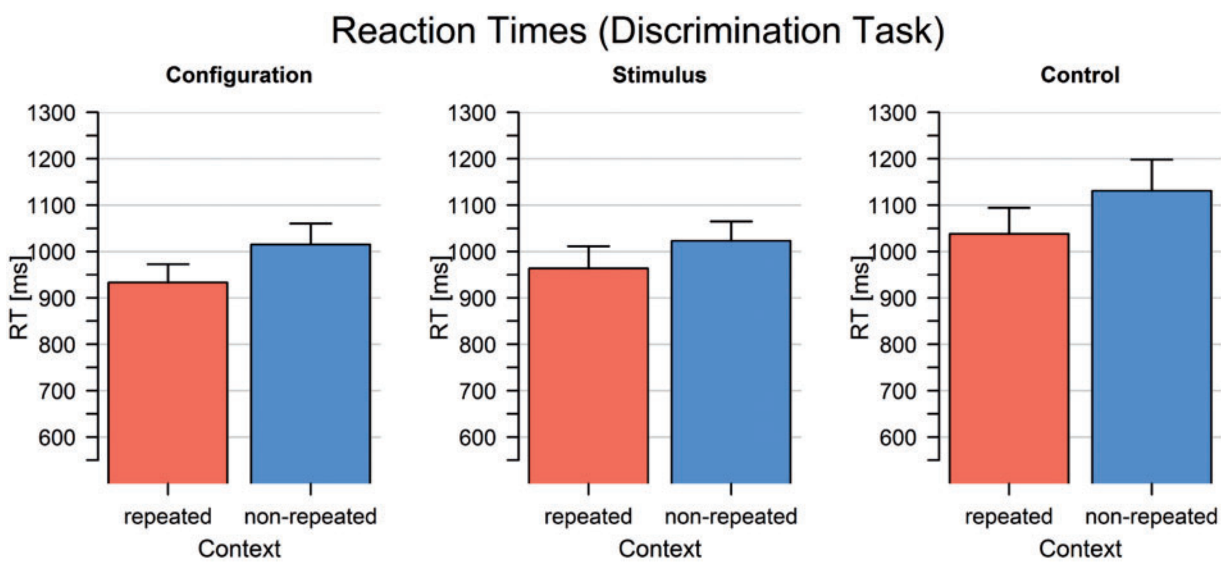


Figure 5. Mean RTs in the discrimination task as a function of display type (repeated vs. non-repeated) in the configuration, stimulus, and control condition (left, middle, and right panel, respectively).

The most parsimonious explanation for the present pattern of results is that memory of visual context enhances the representation of the current display configuration. This improved representation then guides visual attention to the target of visual search, thus speeding up visual search. Previous studies suggested that guidance of attention is one of the mechanisms underlying visual search benefits of repeated displays (Chun and Jiang 1998; Johnson *et al.* 2007; Geyer *et al.* 2010). At the same time, the improved representation might form the basis for reports about the subjective visual experience of the display configuration. Learning of the display configuration will also increase the correlation between verbal reports and localization accuracy: Assuming that the quality of the representation of the current display configuration varies from trial to trial, then, whenever there is a strong representation, there will be both a vivid experience of the display configuration and a high probability of correct target localization. Whenever the representation of the display is rather poor, there is neither a clear experience of display configuration nor a high chance of detecting (localizing) the target item. When displays are non-repeated, the quality of the representation of the configuration varies as well, but representation of the configuration is no longer predictive of the location of the target stimulus and thus the quality of the representation is less predictive of accurate target localization. At a consequence, the associations between subjective reports about the experience of the configuration and localization accuracy is reduced.

The idea that repeated displays enhance the representation of the configuration, which then guides attention to the target, can also provide an explanation why repeated displays increase the visual experience of the target, but leave the correlation with localization accuracy unchanged: When attention is allocated to the target more efficiently, there will also be more time and resources to process the target, eventually giving rise to a more vivid experience of the target item. However, repeated displays seem not to have changed the experience of the target over and above the effect caused by improved localization of the target: If this were the case, repeated displays should have increased subjective experience of the target specifically in correct trials but not in incorrect trials, thus increasing the correlation between subjective experience of the target and localization performance. Instead, repeated displays increased subjective experience of the target item indiscriminately for both correct and incorrect search trials, rather than for correct trials alone. Consequently, it is likely that the effect of repeated displays on experience of the target are only a by-product of the facilitation of visual search by repeated search contexts instead of enhancing conscious processing of the target item. This observation relates to previous investigations of context memory, showing that repeated displays do not only guide attention, but may also speed up processes after attention has already been allocated at the target, including perceptual analysis of the target (Töllner *et al.* 2013) and response selection (Kunar *et al.* 2007; Hout and Goldinger 2012). One possibility is that context memory indeed speeds up processing of the target, but the effect does not affect conscious experience of the target item. A second possibility is that specific features of the present task diminished the influence of learned displays on post-selective processes: Because observers were asked to localize, not identify the target, it is possible that observers only processed the target to a minimal extent.

In summary, we interpret the present findings from a localization task as evidence for the beneficial effects of context memory on conscious experience of the display configuration. We find no evidence for the effects of context memory on subjective experience of the target stimulus.

Is metacognition influenced by learned context?

Verbal reports about subjective experience always rely on some degree of metacognition: participants need to know about conscious experience in order to report it (Dienes 2004; Seth 2008; Zehetleitner and Rausch 2013). Specifically, verbal reports about visual experience of the display configuration seem to require metacognition about perceptual processing of the display configuration. Consequently, the present results may point out an effect of contextual memory on metacognition and not only on conscious experience. In line with this interpretation, context was observed to increase the correlation between subjective reports experience of the display configuration and localization accuracy. The correlation between subjective reports and objective performance is often considered one important aspect of metacognition (Fleming and Lau 2014). Noteworthy, metacognitive access to visual context is consistent with a recent eye-tracking study (Kaunitz *et al.* 2016). In their study, Kaunitz *et al.* asked their observers to search for face targets that were embedded in a crowded natural scene (of other persons/faces). Following search, they were asked about their incidental memory for visually inspected distractor faces and the confidence about their decisions. Interestingly, memory accuracy was correlated with confidence ratings, suggesting that the memory of distractor faces was consciously accessible. The results of the present study may add to their findings in at least two aspects: First, while Kaunitz *et al.* used faces as stimuli, the present results may indicate metacognition for relatively untrained and seemingly meaningless laboratory stimuli of letter arrangements. Second, according to the present study, metacognition may be influenced by more permanent (durable) visual memory instead of only the memory related to the immediate search trial (as in Kaunitz *et al.* 2016).

We do of course acknowledge that the present study does not involve conventional measures of metacognition. The standard approach is to ask participants about their confidence of being correct in the experimental task, here: target localization accuracy, instead of asking them about the clarity of the distractor configuration vs. the target stimulus (Nelson and Narens 1990; Kepecs and Mainen 2012; Fleming and Lau 2014). Confidence judgments are not always equivalent to subjective reports about visual experience (Sandberg *et al.* 2010; Zehetleitner and Rausch 2013; Wierzchoń *et al.* 2014; Jachs *et al.* 2015; Rausch *et al.* 2015). For example, subjective reports about visual experience may underestimate metacognition if participants have knowledge about their performance that is not experienced visually, for example, an intuition of being correct in the task (Rausch and Zehetleitner 2016). Notably, the assessment of metacognition may come to different results when experimental procedures designed to minimize the impact of bias are used: By asking observers which of two preceding task decisions they are willing to bet on, metacognition in masked orientation judgments was shown to be even Bayes-optimal (Peters and Lau 2015). At a consequence, future studies should investigate the relation between contextual memory and metacognition using more direct measures of metacognition.

Implication for the neurocognitive mechanisms of contextual cueing

The present results are well in line with neuroscientific investigations suggesting that contextual cueing of visual search is supported by the medial temporal lobe (MTL) and specifically the hippocampus (HC, Chun and Phelps 1999; Geyer *et al.* 2012; Greene *et al.* 2007). The traditional view is that these structures are essential for declarative memory, which typically includes awareness of learned materials (Manns and Squire 2001). If contextual cueing is based on MTL/HC structures, and if these structures are accessible to consciousness, the question arises why contextual cueing is not associated with awareness. A previously suggested solution to this controversy is that MTL and HC are not exclusively dedicated to explicit memory, but serve other forms of relational memory, which can be implicit, too (Chun and Phelps 1999; Henke 2010). This idea is consistent with a recent functional magnetic resonance imaging study (Geyer *et al.* 2012), suggesting that individual repeated search displays that yield above-chance knowledge in an explicit recognition test (performed concurrently with the search task) are associated with increased MTL/HC activations relative to non-repeated displays. Interestingly, these areas also showed decreased activations in the absence of awareness for other individual repeated displays. Thus, the very same (MTL/HC) areas would process explicit and implicit search displays, though these areas would exert their effects in functionally different ways (repetition enhancement vs. suppression, respectively). The present findings may point out another possibility, namely that MTL/HC activity during contextual cuing is associated with consciousness, although not with awareness of learning, but with a changed subjective experience of search arrays. Under this account, MTL/HC activations during contextual cuing would no longer be special in the sense that they do not give rise to conscious experience. However, neuroscientific studies are required to put this proposal into test.

Conclusions

The present study suggests that the effects of spatial memory acquired during repeated encounters of identical search arrays go beyond the effects on visual search behavior and modulate, that is, enhance the subjective experience of the display configuration.

Supplementary data

Supplementary data is available at *NCONSC Journal* online.

Funding

This research was supported by the German Research Foundation (Grants GE 1889/3-1 to T.G. and Z.E. 887/3-1 to M.Z.). The funding institution had no role in study design, data collection, analysis, decision to publish, or preparation of the manuscript.

Statement of Data Availability

The raw data, the analysis codes, and the reported results are publicly available at the Open Science Framework (https://osf.io/c25wb/?view_only=bae27e28fe5542fc8c2e695f266408e4), to facilitate reproduction of the present study and replication of its results (Ince *et al.* 2012; Morin *et al.* 2012; Simonsohn 2013; Wicherts 2013).

References

Boring EG. A history of introspection. *Psychol Bull* 1953;**50**:169–89.

Brady TF, Chun MM. Spatial constraints on learning in visual search: modeling contextual cuing. *J Exp Psychol* 2007;**33**: 798–815.

Brainard DH. The psychophysics toolbox. *Spat Vis* 1997;**10**:433–6.

Chun MM. Contextual cueing of visual attention. *Trends Cogn Sci* 2000;**4**:170–8.

Chun MM, Jiang Y. Contextual cueing: implicit learning and memory of visual context guides spatial attention. *Cogn Psychol* 1998;**71**:28–71.

Chun MM, Jiang Y. Implicit, long-term spatial contextual memory. *J Exp Psychol* 2003;**29**:224–34.

Chun MM, Phelps EA. Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. *Nat Neurosci* 1999;**2**:844–7.

Colagiuri B, Livesey EJ. Contextual cuing as a form of nonconscious learning: Theoretical and empirical analysis in large and very large samples. *Psychon Bull Rev* 2016;**23**:1996–2009.

Dehaene S, Naccache L. Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition* 2001;**79**:1–37.

Dienes Z. Assumptions of subjective measures of unconscious mental states. *J Conscious Stud* 2004;**11**:25–45.

Dienes Z. Bayesian versus orthodox statistics: which side are you on? *Perspect Psychol Sci* 2011;**6**:274–90.

Dienes Z, Seth AK. Measuring any conscious content versus measuring the relevant conscious content: comment on Sandberg et al. *Conscious Cogn* 2010;**19**:1079–80.

Eriksen CW. Discrimination and learning without awareness: a methodological survey and evaluation. *Psychol Rev* 1960;**67**: 279–99.

Fleming SM, Lau HC. How to measure metacognition. *Front Hum Neurosci* 2014;**8**:1–9.

Fleming SM, Weil RS, Nagy Z *et al*. Relating introspective accuracy to individual differences in brain structure. *Science* 2010; **329**:1541–3.

Galvin SJ, Podd JV, Drga V *et al*. Type 2 tasks in the theory of signal detectability: discrimination between correct and incorrect decisions. *Psychon Bull Rev* 2003;**10**:843–76.

Geyer T, Baumgartner F, Müller HJ *et al*. Medial temporal lobe-dependent repetition suppression and enhancement due to implicit vs. explicit processing of individual repeated search displays. *Front Hum Neurosci* 2012;**6**:1–13.

Geyer T, Zehetleitner M, Müller HJ. Contextual cueing of pop-out visual search: when context guides the deployment of attention. *J Vis* 2010;**10**:20.

Goujon A, Didierjean A, Thorpe S. Investigating implicit statistical learning mechanisms through contextual cueing. *Trends Cogn Sci* 2015;**19**:524–33.

Greene AJ, Gross WL, Elsinger CL *et al*. Hippocampal differentiation without recognition: an fMRI analysis of the contextual cueing task. *Learn Mem* 2007;**14**:548–53.

Henke K. A model for memory systems based on processing modes rather than consciousness. *Nat Rev Neurosci* 2010;**11**: 523–32.

Hout MC, Goldinger SD. Incidental learning speeds visual search by lowering response thresholds, not by improving efficiency: evidence from eye movements. *J Exp Psychol* 2012;**38**:90–112.

Ince DC, Hatton L, Graham-Cumming J. The case for open computer programs. *Nature* 2012;**482**:485–8.

Jachs B, Blanco MJ, Grantham-Gill S *et al*. On the independence of visual awareness and metacognition: a signal detection theoretic analysis. *J Exp Psychol* 2015;**41**:269–76.

Jiang Y, Wagner LC. What is learned in spatial contextual cuing–configuration or individual locations? *Percept Psychophys* 2004; **66**:454–63.

Jiang YV, Swallow KM, Capistrano CG. Visual search and location probability learning from variable perspectives. *J Vis* 2013;**13**: 1–13.

Johnson JS, Woodman GF, Braun E *et al*. Implicit memory influences the allocation of attention in visual cortex. *Psychol Bull Rev* 2007;**14**:834–9.

Kaunitz LN, Rowe EG, Tsuchiya N. Large Capacity of Conscious Access for Incidental Memories in Natural Scenes. *Psychol Sci* 2016;**27**:1266–77.

Kepecs A, Mainen ZF. A computational framework for the study of confidence in humans and animals. *Philos Trans R Soc Lond B* 2012;**367**:1322–37.

Kunar MA, Flusberg S, Horowitz TS *et al*. Does contextual cuing guide the deployment of attention? *J Exp Psychol* 2007;**33**: 816–28.

Lau HC. Are we studying consciousness yet? In: Weiskrantz L, Davies M (eds.), *Frontiers of Consciousness: Chichele Lectures*. 1st ed. Oxford, UK: Oxford University Press, 2008, 245–60.

Maniscalco B, Lau H. A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Conscious Cogn* 2012;**21**:422–30.

Manns JR, Squire LR. Perceptual learning, awareness, and the hippocampus. *Hippocampus* 2001;**11**:776–82.

Masson MEJ, Rotello CM. Sources of bias in the Goodman–Kruskal gamma coefficient measure of association: implications for studies of metacognitive processes. *J Exp Psychol Learn Mem Cogn* 2009;**35**:509–27.

Morey RD, Rouder JN. *BayesFactor: Computation of Bayes Factors for Common Designs*. R package version 0.9.10-1, 2015. https://cran.r-project.org/package=BayesFactor.

Morin A, Urban J, Adams PD *et al*. Shining light into black boxes. *Science* 2012;**336**:159–60.
Nelson T, Narens L. Metamemory: a theoretical framework and new findings. *Psychol Learn Motiv* 1990;**26**:125–73.
Pelli DG. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis* 1997;**10**:437–42.
Peters MAK, Lau H. Human observers have optimal introspective access to perceptual processes even for visually masked stimuli. *eLife* 2015;**4**:e09651.
R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2014.
Ramsøy TZ, Overgaard M. Introspection and subliminal perception. *Phenomenol Cogn Sci* 2004;**3**:1–23.
Rausch M, Müller HJ, Zehetleitner M. Metacognitive sensitivity of subjective reports of decisional confidence and visual experience. *Conscious Cogn* 2015;**35**:192–205.
Rausch M, Zehetleitner M. Visibility is not equivalent to confidence in a low contrast orientation discrimination task. *Front Psychol* 2016;**7**:591.
Rausch M, Zehetleitner M. Should metacognition be measured by logistic regression? *Conscious Cogn* 2017;**49**:291–312.
Rouder JN, Morey RD. Default Bayes factors for model selection in regression. *Multivariate Behav Res* 2012;**47**:877–903.
Rouder JN, Speckman PL, Sun D *et al*. Bayesian t tests for accepting and rejecting the null hypothesis. *Psychon Bull Rev* 2009;**16**: 225–37.
Sandberg K, Timmermans B, Overgaard M *et al*. Measuring consciousness: is one measure better than the other? *Conscious Cogn* 2010;**19**:1069–78.
Schlagbauer B, Geyer T, Müller HJ *et al*. Rewarding distractor context versus rewarding target location: a commentary on Tseng and Lleras (2013). *Atten Percept Psychophys* 2014;**76**:669–74.
Schlagbauer B, Müller HJ, Zehetleitner M *et al*. Awareness in contextual cueing of visual search as measured with concurrent access- and phenomenal-consciousness tasks. *J Vis* 2012;**12**: 1–31.
Sergent C, Dehaene S. Is consciousness a gradual phenomenon? Evidence for an all-or-none bifurcation during the attentional blink. *Psychol Sci* 2004;**15**:720–8.
Seth AK. Post-decision wagering measures metacognitive content, not sensory consciousness. 2008;**17**:981–83.
Seth AK, Dienes Z, Cleeremans A *et al*. Measuring consciousness: relating behavioural and neurophysiological approaches. *Trends Cogn Sci* 2009;**12**:314–21.
Simonsohn U. Just post it: the lesson from two cases of fabricated data detected by statistics alone. *Psychol Sci* 2013;**24**: 1875–88.
Smyth AC, Shanks DR. Awareness in contextual cuing with extended and concurrent explicit tests. *Memory* 2008;**36**: 403–15.
Töllner T, Conci M, Rusch T *et al*. Selective manipulation of target identification demands in visual search: the role of stimulus contrast in CDA activations. *J Vis* 2013;**13**:1–13.
Vadillo MA, Konstantinidis E, Shanks DR. Underpowered samples, false negatives, and unconscious learning. *Psychon Bull Rev* 2015;**23**:87–102.
Wetzels R, Matzke D, Lee MD *et al*. Statistical evidence in experimental psychology: an empirical comparison using 855 t tests. *Perspect Psychol Sci* 2011;**6**:291–8.
Wicherts JM. Science revolves around the data. *J Open Psychol Data* 2013;**1**:1–4.
Wierzchoń M, Paulewicz B, Asanowicz D *et al*. Different subjective awareness measures demonstrate the influence of visual identification on perceptual awareness ratings. *Conscious Cogn* 2014;**27**:109–20.
World Medical Association. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. *J Am Med Assoc* 2013;**310**:2191–94.
Zehetleitner M, Rausch M. Being confident without seeing: what subjective measures of visual consciousness are about. *Atten Percept Psychophys* 2013;**75**:1406–26.

Supplementary Material

to

Contextual cueing of visual search is associated with greater subjective experience of the search display configuration

Bernhard Schlagbauer[1,2,*,†], Manuel Rausch[1,2,3,*], Michael Zehetleitner[1,3], Hermann J. Müller[1,4], and Thomas Geyer[1]

[1]Department Psychologie, Ludwig-Maximilians-Universität München, Munich, Germany

[2]Graduate School of Systemic Neurosciences, Ludwig-Maximilians-Universität München, Planegg-Martinsried, Germany

[3]Katholische Universität Eichstätt-Ingolstadt, Eichstätt, Germany

[4]School of Psychology, Birkbeck College, University of London, London, United Kingdom

[*]These authors contributed equally to this work

†To whom correspondence should be addressed. E-mail: bernhard.schlagbauer@psy.lmu.de

Supplementary Methods

Staircase Procedure

To adjust SOAs individually for each participant, an adaptive 1-up 2-down staircase procedure with the same task was applied. The staircase started with an initial SOA value of 600 ms and was adjusted stepwise until the first reversal point was reached (i.e., a correct trial after an error trial, or vice versa). During the first four reversals, SOA values were modified by step sizes of four frame durations (~48 ms at 85 Hz) to step sizes of one frame duration (~12 ms) for the last eight reversals. The SOA step size was doubled (increased) following an error response trial. The staircase procedure stopped after the 12 reversal points. The whole procedure was repeated four times in order to account for procedural-learning effects and also to allow participants to become familiar with the task. From the last staircase procedure, the mean of the last six reversals was calculated and taken as SOA for part 1 (the localization task) of the experiment. The average SOA obtained by this procedure was 492 ms (SD = 117 ms). Only non-repeated search displays were used in the staircase procedure.

Bayesian statistics

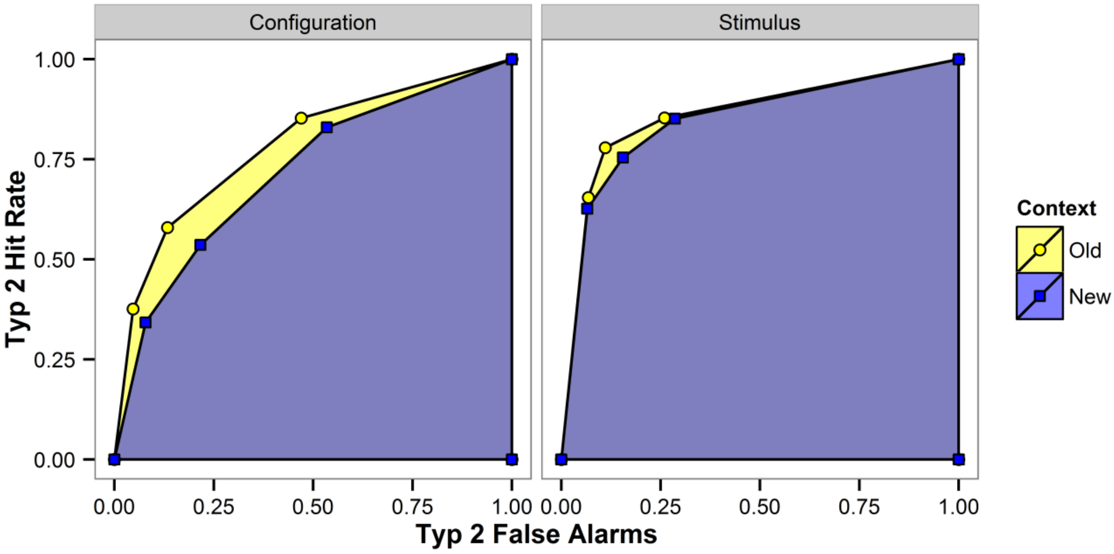
In general, orthodox statistics may bias in the investigation of implicit learning, since absence of evidence is often misinterpreted as evidence for implicit processing. A solution to this problem is provided by Bayesian statistics (Vadillo, Konstantinidis, & Shanks, 2015), because Bayes factors quantify evidence for both the null and the alternative hypothesis (Dienes, 2015; Rouder, Speckman, Sun, Morey, & Iverson, 2009). A Bayes factor indicates how the prior odds in favor of one over another hypothesis need to be adjusted in light of the data: For example, a Bayes factor of 10 for the hypothesis that there is an effect (H_1) vs. the hypothesis the effect is null (H_0) means that, according to the available data, researchers should increase their belief about the odds of the two hypotheses by a factor of 10.

A strength of Bayes factors is that in contrast to a non-significant result, which does not allow for deducing that the data supports the H_0, a Bayes factor is also informative whether the result is just inconclusive or whether the data supports the null hypothesis. In other words, Bayes factors can provide evidence of absence.

Using receiver operating characteristics to quantify the relation between visual experience and task accuracy

When assessing the relationship between subjective reports of visual experience and trial accuracy, it is necessary to control for the criteria participants apply when they select a specific category on the discrete scale. It can be assumed that subjective experiences of low-level stimuli such as the search target or of display configurations vary gradually, i.e. the experience is absent in some trials, weak in other trials, and strong in the remainder of trials (Windey, Vermeiren, Atas, & Cleeremans, 2014). To select one rating category of the four-point scales in our experiment, participants need to compare their experience with three criteria that delineate the rating categories. For example, on an experience scale with the options “none”, “vague”, “almost clear” and “clear”, participants might respond “vague” when their subjective experience falls between that criterion separating the response “none” from “vague” and that criterion separating “vague” from “almost clear”. When the relation between experience and trial accuracy is quantified, the obtained estimate should be independent from these criteria. Importantly, the mathematical problem is the same as discussed in the context of type 2 signal detection theory (Fleming & Lau, 2014; Galvin, Podd, Drga, & Whitmore, 2003; Maniscalco & Lau, 2012); the difference is that subjective reports in type 2 signal detection theory are introduced to the subject as being about the accuracy in the task response (Galvin et al., 2003), while subjective reports of experience do not bear a semantic relation to task performance. At a consequence, the relation between experience and trial accuracy can be quantified the area under the type 2 receiver operating characteristics (ROC). It should be noted that ROC curves can be seen as a measure of performance of a binary classifier independent of classification criterion closely related to the Wilcoxon test of ranks (Fawcett, 2006); as a consequence, it is #not required that the subjective report is about the accuracy of the task response as it is the case in type 2 signal detection theory. Alternative measures of association are either strongly influenced by criterion setting, e.g. gamma correlations (Masson & Rotello, 2009) and logistic regression slopes (Rausch, Müller, & Zehetleitner, 2015; Rausch & Zehetleitner, 2017) or not applicable for the present four quadrant localization task, e.g. meta-d' (Maniscalco & Lau, 2012).

To construct type 2 ROC curves, type two hit rate and type 2 false alarms are computed. The type 2 hit rate h is defined as proportion of high experience trials when the participant is correct and type 2 false alarms rate f as the proportion of high experience trials when the participant is incorrect (Galvin et al., 2003). However, with four category reports, each two adjacent rating categories can be treated as criterion that separates high from low experience. For instance, there is a liberal criterion that assigns low experience = 1 and high experience = 2 – 4, then there is a higher criterion that assigns low experience = 1 and 2 and high experience = 3 and 4, and so on. For each split of the rating data, h and f are calculated and plotted to obtain a type 2 ROC curve (see Supplementary Fig. 1). The area under the type 2 ROC curve can then be used as a measure of metacognitive sensitivity and was computed using the function *somers2* implemented in the R library Hmisc (Harrell, 2014). An area under the curve of 1 indicates perfect sensitivity to differentiate between correct and incorrect trials, whereas tan area is of .5 indicates chance level.



Supplementary Figure 1. Receiver operating characteristics as a function of context (old vs. new) in the configuration (left panel) and stimulus (middle panel) condition in epoch 2.

Supplementary Results

Localization performance

Accuracy in the first epoch

In an alternative approach to analyze possible differences between the experimental groups, we also compared localization performance in the first epoch of localization task. The three groups were compared pairwise with the Bayesian equivalent of a two-sided t-test with the same prior as we used for the post-hoc tests. Two out of three possible pairings showed that there was substantial evidence that overall localization performance was identical between groups [configuration vs. stimulus: BF_{10} = .23; configuration vs. control: BF_{10} = .22], while the comparison of the stimulus and the control condition yielded some indication that there was of no difference between groups [stimulus vs. control: BF_{10} = .335].

Accuracy in the second epoch

To investigate whether the report condition had an influence on the magnitude of contextual cueing, we conducted an analysis of localization accuracy restricted to the second epoch of the localization task. There was a main effect of display type [BF_{10} = 4,629], and inconclusive evidence regarding a main effect of report condition [BF_{10} = .74]. The latter effect mirrors the interaction between epoch and report condition observed in the original analysis, indicating differences in observers' general improvement in the localization task. Importantly, there was no interaction of display type and report condition [BF_{10} = .14], providing further evidence for the idea that the concurrent assessment of observers' conscious reports and their localization performance did not impact on the contextual cueing effect.

Association between verbal reports and localization accuracy

Accuracy and reaction times as covariates

Assuming that type 2 ROC curves also depend on observers objective task performance, the effect of repeated context on the association between localization performance and verbal reports could be just a by-product of the behavioral improvement in the search task (Fleming & Lau, 2014; Galvin et al., 2003; Maniscalco & Lau, 2012). Recall that the analysis localization performance was not conclusive whether localization performance differed between groups. It is thus possible that the differences between group in objective task performance have contributed to ROC measures. As a first check, we repeated the analysis of type 2 ROC curves with accuracy and reaction time as covariates. Although the localization task was unspeeded, reaction times were included in this analysis to control as many objective task parameters as possible. The results of this ‘control’ analysis are identical to those from the ‘original’ analysis of ROC curves, suggesting that repeated contexts increased the area under the ROC curve of subjective reports about experience of the display configuration (the effect emerged only in epoch 2 of the experiment). Specifically, in the control analysis, only the three-way interaction between display type, report condition, and epoch was substantial [BF_{10} = 4.84]. There was anecdotal evidence for a main effect of report condition [BF_{10} = 2.49] and the interaction of report condition and epoch [BF_{10} = .55]. Regarding all other effects, there was substantial evidence for the null hypothesis [BF_{10} < .30].

Analysis of a homogeneous subset of observers

As a second check, we performed an analysis of a subset of observers, in which we included only observers whose localization performance was within one standard deviation of the sample mean. This resulted in the exclusion of 11 subjects, reducing the sample standard deviation from 11% to 4%. The subset analysis was done in an attempt to examine the effects of context memory in participants who displayed more consistent accuracies across report conditions, being able to further minimize potential effects of (between-group differences in) observers objective task performance on their ROC curves. After exclusion of 11 observers, mean localization accuracy was 75.0% in the configuration, 75.8% in the stimulus and 74.4% in the control condition. Performing then the analysis on this very homogenous set of observers revealed two theoretically important effects, replicating those from the analysis of the entire set of observers: (1) there was a main effect of display type in the analysis of observers' localization performance, suggesting a reliable context effect in all three report conditions/groups. (2) there was a reliable three-way interaction of display type, epoch, and report condition in the analysis of observers' association between subjective reports and localisation accuracy (suggesting that the area under the ROC curve increased over time in the configuration condition). In more detail, the (control) analysis of localization accuracy revealed substantial effects of display type [BF_{10} = 7189], epoch [BF_{10} = 165520], report condition x epoch [BF_{10} = 3.35], and display type x epoch [BF_{10} = 6.73]. Regarding all other effects, there was substantial evidence for the null hypothesis [BF_{10} < .22]. Regarding the (control) analysis of ROC curves, the only substantial effect was the three-way-interaction of display type x epoch x report condition [BF_{10} = 7.70]. All other effects were inconclusive with anecdotal evidence for an effect of report condition [BF_{10} = 1.56], display type [BF_{10} = 2.35], epoch [BF_{10} = 2.02] and report condition x epoch [BF_{10} = 2.31], while there was anecdotal evidence for the null hypothesis in the case of report condition x display type [BF_{10} = 0.59] and epoch x display type [BF_{10} = 0.53].

Tables

Table 1. Mean values and standard deviations of the dependent variables localization accuracy, verbal reports, and area under the ROC curves (= sensitivity) in the localization task.

	Repeated				**Non-repeated**			
	Epoch 1		*Epoch 2*		*Epoch 1*		*Epoch 2*	
Configuration	Accuracy:	Mean = 74.2% SD = 8.6%	Accuracy:	Mean = 83.2% SD = 8.3%	Accuracy:	Mean = 71.0% SD = 9.5%	Accuracy:	Mean = 76.2% SD = 8.8%
	Reports:	Mean = 2.50 SD = 0.52	Reports:	Mean = 2.68 SD = 0.57	Reports:	Mean = 2.43 SD = 0.51	Reports:	Mean = 2.54 SD = 0.50
	Sensitivity:	Mean = 0.73 SD = 0.15	Sensitivity:	Mean = 0.77 SD = 0.16	Sensitivity:	Mean = 0.74 SD = 0.16	Sensitivity:	Mean = 0.72 SD = 0.16
Stimulus	Accuracy:	Mean = 74.0% SD = 4.9%	Accuracy:	Mean = 77.9% SD = 15.5%	Accuracy:	Mean = 74.3% SD = 5.3%	Accuracy:	Mean = 72.0% SD = 14.8%
	Reports:	Mean = 2.84 SD = 0.36	Reports:	Mean = 3.01 SD = 0.40	Reports:	Mean = 2.85 SD = 0.32	Reports:	Mean = 2.87 SD = 0.37
	Sensitivity:	Mean = 0.83 SD = 0.07	Sensitivity:	Mean = 0.83 SD = 0.12	Sensitivity:	Mean = 0.81 SD = 0.07	Sensitivity:	Mean = 0.83 SD = 0.08
Control	Accuracy:	Mean = 71.3% SD = 12.3%	Accuracy:	Mean = 74.2% SD = 11.6%	Accuracy:	Mean = 69.8% SD = 10.2%	Accuracy:	Mean = 70.7% SD = 10.7%

Table 2. Mean values and standard deviations of reaction times in the discrimination task.

	Repeated		Non-repeated	
Configuration	Mean = 933 ms	SD = 153 ms	Mean = 1015 ms	SD = 174 ms
Stimulus	Mean = 963 ms	SD = 186 ms	Mean = 1023 ms	SD = 162 ms
Control	Mean = 1038 ms	SD = 219 ms	Mean = 1131 ms	SD = 261 ms

References

Dienes, Z. (2015). How Bayesian statistics are needed to determine whether mental states are unconscious. In M. Overgaard (Ed.), *Behavioural Methods in Consciousness Research* (pp. 199–220). Oxford: Oxford University Press.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*, 861–874. http://doi.org/10.1016/j.patrec.2005.10.010

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, *8*(443), 1–9. http://doi.org/10.3389/fnhum.2014.00443

Galvin, S. J., Podd, J. V, Drga, V., & Whitmore, J. (2003). Type 2 tasks in the theory of signal detectability: Discrimination between correct and incorrect decisions. *Psychonomic Bulletin & Review*, *10*(4), 843–876.

Harrell, F. E. J. with contributions from C. D. and many others. (2014). Hmisc: Harrell Miscellaneous. R package version 3.14-5.

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, *21*(1), 422–30. http://doi.org/10.1016/j.concog.2011.09.021

Masson, M. E. J., & Rotello, C. M. (2009). Sources of Bias in the Goodman – Kruskal Gamma Coefficient Measure of Association : Implications for Studies of Metacognitive Processes, *35*(2), 509–527. http://doi.org/10.1037/a0014876

Rausch, M., Müller, H. J., & Zehetleitner, M. (2015). Metacognitive sensitivity of subjective reports of decisional confidence and visual experience. *Consciousness and Cognition*, *35*, 192–205. http://doi.org/10.1016/j.concog.2015.02.011

Rausch, M., & Zehetleitner, M. (2017). Should metacognition be measured by logistic regression? *Consciousness and Cognition*, *49*, 291–312. http://doi.org/10.1016/j.concog.2017.02.007

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–37. http://doi.org/10.3758/PBR.16.2.225

Vadillo, M. A., Konstantinidis, E., & Shanks, D. R. (2015). Underpowered samples, false negatives, and unconscious learning. *Psychonomic Bulletin & Review*, *23*(1), 87–102. http://doi.org/10.3758/s13423-015-0892-6

Windey, B., Vermeiren, A., Atas, A., & Cleeremans, A. (2014). The graded and dichotomous nature of visual awareness. *Philosophical Transactions of the Royal Society B*, *369*, 20130282. http://doi.org//10.1098/rstb.2013.0282