Routledge Taylor & Francis Group

Visual Cognition

ISSN: 1350-6285 (Print) 1464-0716 (Online) Journal homepage: http://www.tandfonline.com/loi/pvis20

The role of context on boundary extension

Ezgi Mamus & Aysecan Boduroglu

To cite this article: Ezgi Mamus & Aysecan Boduroglu (2018) The role of context on boundary extension, Visual Cognition, 26:2, 115-130, DOI: 10.1080/13506285.2017.1399947

To link to this article: https://doi.org/10.1080/13506285.2017.1399947

Published online: 23 Nov 2017.

VISUAL COGNITION, 2018
VOL. 26, NO. 2, 115–130
https://doi.org/10.1080/13506285.2017.1399947

Routledge
Taylor & Francis Group

The role of context on boundary extension

Ezgi Mamus and Aysecan Boduroglu

Department of Psychology, Bogazici University, Istanbul, Turkey

ABSTRACT
Boundary extension (BE) is a memory error in which observers remember more of a scene than they actually viewed. This error reflects one's prediction that a scene naturally continues and is driven by scene schema and contextual knowledge. In two separate experiments we investigated the necessity of context and scene schema in BE. In Experiment 1, observers viewed scenes that either contained semantically consistent or inconsistent objects as well as objects on white backgrounds. In both types of scenes and in the no-background condition there was a BE effect; critically, semantic inconsistency in scenes reduced the magnitude of BE. In Experiment 2 when we used abstract shapes instead of meaningful objects, there was no BE effect. We suggest that although scene schema is necessary to elicit BE, contextual consistency is not required.

ARTICLE HISTORY
Received 12 October 2016
Accepted 24 October 2017

KEYWORDS
Scene perception; boundary extension; contextual consistency

Boundary extension (BE) is a memory error in which observers remember more of a scene than they are shown, particularly the areas which fall outside the physical boundaries of a view (Intraub & Richardson, 1989). While BE is typically described as an error, it is believed to have heuristic value. BE reflects a mostly accurate prediction of the natural continuation of a scene, and viewing durations as short as 42 ms have been shown to induce BE (Intraub & Dickinson, 2008; Intraub & Richardson, 1989). Intraub (2012), in her multisource model of scene perception, has specifically argued that BE may support scene perception (Intraub, 2012; Intraub & Dickinson, 2008). Given the frequent interruptions of the visual input by blinks and saccades, BE may be functional in facilitating and speeding up the integration process of successive sensory input, leading to a more continuous perceptual experience of the visual surroundings. While early findings from scene memory tasks have demonstrated that BE was limited to scene memory and did not occur when objects were presented on blank backgrounds (Gottesman & Intraub, 2002; Intraub, Gottesman, & Bills, 1998), more recent research showed that BE was similarly elicited for basic objects on blank backgrounds (Konkle & Olivia, 2007) and for abstract stimuli on blank backgrounds – manipulations that eliminated context and schematic information from images (McDunn, Siddiqui, & Brown, 2014). Given these discrepant findings, in this study we investigated the necessity of object and context related conceptual information on BE.

Intraub and Richardson (1989) initially introduced the BE phenomenon based on data gathered from experiments using both a picture drawing and a picture recognition task. When viewers were asked to draw previously studied scenes, they drew parts of the scenes that were not seen in the original stimulus. For instance, they completed the cropped objects in the studied picture. In a separate experiment, participants studied both close-up and wide-angle scenes. After a two-day delay, they were asked to rate whether the test scene was identical, closer or further from the studied scenes. Participants rated both close-up and wide-angle test scenes as being more close-up than before, but close-up scenes were rated as closer than wide-angle scenes, leading to greater BE magnitude. Subsequent research by Intraub and colleagues has repeatedly shown that BE is only associated with images involving scenes or elements that can be construed as scenes (Gottesman & Intraub, 2002; Intraub et al., 1998). For instance, Intraub et al. (1998) presented photographs of objects in coherent scene contexts, line drawings of those natural scenes or line drawings of single objects on blank white backgrounds. Participants were instructed to remember the objects and their sizes. The results revealed that participants in scene conditions (either natural or line drawings) showed

CONTACT Aysecan Boduroglu aysecan.boduroglu@boun.edu.tr Bogazici Universitesi, Psikoloji Bolumu, Bebek, Istanbul 34340, Turkey

BE but this error did not occur in the blank background condition. When participants were instructed to imagine a coherent context for the blank backgrounds, BE occurred as if the objects were presented in natural scenes. Thus, the presence of a real or imagined natural context stimulated probable scene layout which was invisible at the time of viewing. Gottesman and Intraub (2002) showed that photographs of objects on blank backgrounds elicited BE whereas line drawings of the same objects did not. They suggested that real photographs as opposed to line drawings are more likely to induce a partial image of a continuous scene (for a critical review of this study, see McDunn, Brown, Hale, & Siddiqui, 2016). In her unpublished thesis, Gottesman (1998) created non-organized scenes by cutting a scene into six pieces and jumbling it up; she found BE even for these jumbled scenes. However, even though Gottesman (1998) argued that these scenes were jumbled, breaking a continuous background (such as rocks on the beach) into six pieces and shuffling them around did not completely eliminate scene schema. Based on all these findings, Intraub and colleagues concluded that observer's interpretation of a background as a continuous surface rather than background details might be critical in eliciting BE.

The multisource model of scene perception was proposed to account for this set of findings on BE (Intraub & Dickinson, 2008). According to this model, various sources of top-down and bottom-up information are integrated to internally simulate the world that humans cannot see as a whole within a single view. Together with visual sensory input, expectations and restrictions arising from context, knowledge and our prior experience construct an internal spatial framework (Intraub, 2012). The model states that BE occurs because observers erroneously rely on the representation they have internally generated instead of their sensory experience during scene memory tasks. Therefore, BE could be considered as a source monitoring error (Gagnier & Intraub, 2012; Intraub, Daniels, Horowitz, & Wolfe, 2008). For instance, Intraub et al. (2008) showed that BE for scene memories was higher when observers attended to an unrelated visual search task in addition to picture viewing. They inferred that decreased attention during picture viewing increased the difficulty of differentiating the internal amodal representation and experienced visual input. Moreover, Gagnier and Intraub (2012) found that line drawings of scenes elicited much more BE than coloured photographs of the same scenes. They concluded that reduced quality of visual input might increase the source monitoring error for the line drawings because the sensory experience becomes more similar to the imperfect internal representation.

Despite both the empirical evidence and the models indicating that BE is driven by contextual processing of scenes (Gottesman, 1998; Gottesman & Intraub, 2002; Intraub et al., 1998), recent studies have suggested the contrary. For instance, Konkle and Olivia (2007) demonstrated that participants distorted their memories of single objects presented on blank backgrounds. Specifically, they showed that whenever displays consisted of objects not matching participants' idiosyncratic preferences for object size, participants remembered the objects in a normalized (extended or contracted) fashion. Thus, the explanation offered by Konkle and Oliva did not necessitate the presence of a contextual background for BE. More critically, a recent study by McDunn et al. (2014) demonstrated that BE occurred even in abstract scenes consisting of irregular polygons in which semantic context and scene schema were eliminated. There were four conditions. In the first two, they presented the irregular polygons on random-dot backgrounds that either conveyed a sense of occlusion or not; in the latter two the polygons, either as filled-in-grey or as black-outline objects were presented on white blank backgrounds. In all four conditions, they reported BE and concluded that high-level information might not be necessary to elicit BE. Furthermore, they argued that placing objects on random-dot backgrounds might have created "some sense of a spatial context" (p. 374). This may have been particularly true given the visual contrast between the irregular polygon and the random-dot background; the polygon was likely perceived as a foreground object on a 3D space. When Hale, Brown, McDunn, and Siddiqui (2015), presented the same scenes from McDunn et al. (2014) with extremal edges indicating scene boundaries, BE was either eliminated or reduced. These results altogether led the authors to conclude that the amodal extrapolation of scene background played a critical role in BE. Given the limited amount of evidence suggesting that displays do not need to be contextual, we wanted to further test the necessity

of background context on BE by manipulating semantic consistency within displays.

In order to directly test the impact of contextual information on BE, we specifically focused on how object and context consistency impact scene processing. Both empirical studies and computational models have illustrated that scenes are processed rapidly and in a holistic fashion, and identification of scene gist is known to contribute to object identification (for a review see Bar, 2004). Since the early work by Potter (1975), it has been known that target objects could be detected in scenes viewed as short as 125 ms. More recently, Oliva and colleagues demonstrated that scenes can be accurately categorized in even shorter durations, around 40 ms (Greene & Oliva, 2009). Critically, contextual consistency has been shown to facilitate object perception (e.g., Biederman, Mezzanotte, & Rabinowitz, 1982; Davenport, 2007; Davenport & Potter, 2004; Palmer, 1975). Inconsistency between the target and the context retarded processing of scenes and decreased accuracy in scene categorization even when these scenes were flashed for as short as 26 ms (Joubert, Rousselet, Fize, & Fabre-Thorpe, 2007). The advantage of using contextually consistent and inconsistent scenes is that in both cases the background context is continuous and meaningful. However, the perception of the latter group of scenes are not likely to be facilitated by an additive effect of coherence between object and background because the scene elements might have different and independent semantic associations (e.g., Biederman et al., 1982; Davenport & Potter, 2004; Palmer, 1975; Potter, 1976). For instance, inconsistent scenes might support two separate internal representations rather than one coherent internal representation. If this is the case, for inconsistent compared to consistent scenes, there would be less similarity between the richer sensory experience and two internal representations and consequently less source monitoring error and less BE.

A secondary goal of the present study was to determine how individual differences in visual and spatial imagery were related to BE. Image quality and attentional processes have been shown to affect BE (Gagnier & Intraub, 2012; Intraub et al., 2008). Recently, Munger and Multhaup (2016) investigated whether observers' elaboration of sensory details in images increased BE via a boost in source memory error. They hypothesized that visual and spatial imagery might be positively correlated with the magnitude of BE because observers high in imagery might have richer internal representations of scenes and, therefore, make more source memory error. In six experiments, they asked participants to imagine additional smells, sounds and visual details while viewing scenes. Additionally, participants were given self-report imagery questionnaires: Vividness of Imagery Questionnaire, VVIQ (Marks, 1973) in Experiment 1A&B and 2A&B; and Object Spatial Imagery Questionnaire, OSIQ (Blajenkova, Kozhevnikov, & Motes, 2006) in Experiment 3A&B. However, Munger and Multhaup (2016) found that elaborating on any additional sensory details did not predict the magnitude of BE. Furthermore, visual imagery was not correlated with the BE magnitude. In one of the six experiments (Experiment 3A), they showed that individuals higher in spatial imagery had greater BE, but this was shown only for one out of the four trial types (CC trials). It must be noted that the task Munger and Multhaup used in that experiment was highly spatial in nature and required participants to adjust the scene vantage point via mouse clicks. Given the inconclusive results of this study, we wanted to explore the issue of imagery ability on BE. Since the theoretical framework put forward by Intraub and colleagues suggests that BE may be linked to source memory errors, and the link between high imagery ability and greater source memory error (e.g., Markham & Hynes, 1993), we wanted to explore whether individual differences in visual and spatial imagery contributed to BE. Furthermore, in addition to using questionnaire-based measures, we wanted to explore whether standardized tests measuring spatial visualization ability – specifically mental rotation performance – may be linked to the magnitude of BE observed.

To sum up, in the present study we conducted two experiments to examine the necessity of context and scene schema in BE. In the first experiment, we compared BE across consistent and inconsistent scenes. To determine whether we would be able to replicate BE for objects on blank backgrounds, we tested a separate group of participants in the no-background condition. Inclusion of such a condition also allowed us to test whether individual differences in BE contributed to picture ratings. Results of the first experiment revealed there was a BE difference between consistent

and inconsistent scenes partially in line with our expectations. We also found BE in the no-background condition. In Experiment 2, we wanted to further test the necessity of scene context by fully eliminating scene context and continuity. Therefore, in Experiment 2 we took an approach similar to that employed by McDunn et al. (2014) and presented viewers with complex multi-component filled-in objects on blank grey backgrounds. In both studies we also asked participants to fill the Object-Spatial Imagery and Verbal Questionnaire (OSIVQ) and complete a performance-based measure of spatial visualization ability.

Experiment 1

Method

Participants

A total of 120 undergraduate students at Bogazici University participated in the experiment in exchange for extra credit in their psychology courses. There were 80 participants ($M = 20.11$ years (18–24), $SD = 1.21$, 78% female) in the scene condition and 40 participants ($M = 20.33$ years (18–25), $SD = 1.76$, 75% female) in the no-background condition. They all had normal or corrected-to-normal vision and provided written informed consent.

Materials

Picture rating task. The picture rating task consisted of two between subject conditions. One group of participants was asked to rate studied scenes and a separate group of participants were asked to rate single objects presented on blank backgrounds (no-background condition). In the scene condition, there were both consistent and inconsistent scenes presented in a mixed order (see Figure 1). In both scene and no-background conditions, participants were asked to study the presented stimuli for a future memory test and, immediately following the study phase, they completed the test phase in which they had to rate whether a test picture was in the same perspective, closer up or further away than the studied one. Before the study phase, participants were instructed to try to remember the pictures in as much detail as possible. In the scene condition, participants were specifically instructed to pay attention equally to the objects, the background and the layout in the pictures. In the no-background condition, participants were asked to remember the object in as much detail as possible.

In the study phase, we presented each picture for 15 s. Images in the test phase were presented in the same order as in the study phase. Studied pictures were either close-up or wide-angle images. In the test phase, half of the images were shown identical to their studied versions and the remaining half was presented from the alternative perspective of the study versions (i.e., close-up pictures were presented in wide-angle or vice versa). Therefore, there were four trial types in the test phase: close-up study and test (CC), wide-angle study and test (WW), close-up study and wide-angle test (CW) and wide-angle study and close-up test (WC).

In the test phase, participants were instructed to rate each test picture on a 5-point scale. Specifically, they were asked to indicate the position of the camera as being "much closer-up" (−2), "slightly closer-up" (−1), "same" (0), "slightly more wide-angle" (1) or "much more wide-angle" (2). They also reported their confidence in their rating on 4-point scale, with 0 indicating no memory of the picture, (1) "not sure", (2) "pretty sure" and (3) "sure". Before the test phase, participants completed two practice trials with both close-up and wide-angle versions of the pictures in succession. In the scene condition, the first practice trial included a consistent picture which was shown as close-up at test, and the second practice trial included an inconsistent picture which was shown as wide-angle at test. In the no-background condition, participants also had two practice trials in the first one, the test object was presented as close-up; in the second one, it was presented as wide-angle. After completing the practice trials, participants were given further explanations if there was any indication that they misunderstood the instructions.

Pictures. We prepared two sets of 40 (total 80) coloured natural images using Adobe® Photoshop® CS6. Each image had a central foreground object pasted onto a background scene that was either semantically consistent or inconsistent with the object. We collected these objects and the backgrounds separately from available data sets (Davenport & Potter, 2004) and online services (e.g., Google images). Objects consisted of individuals, animals or inanimate man-made things; scenes included both indoor and outdoor environments.

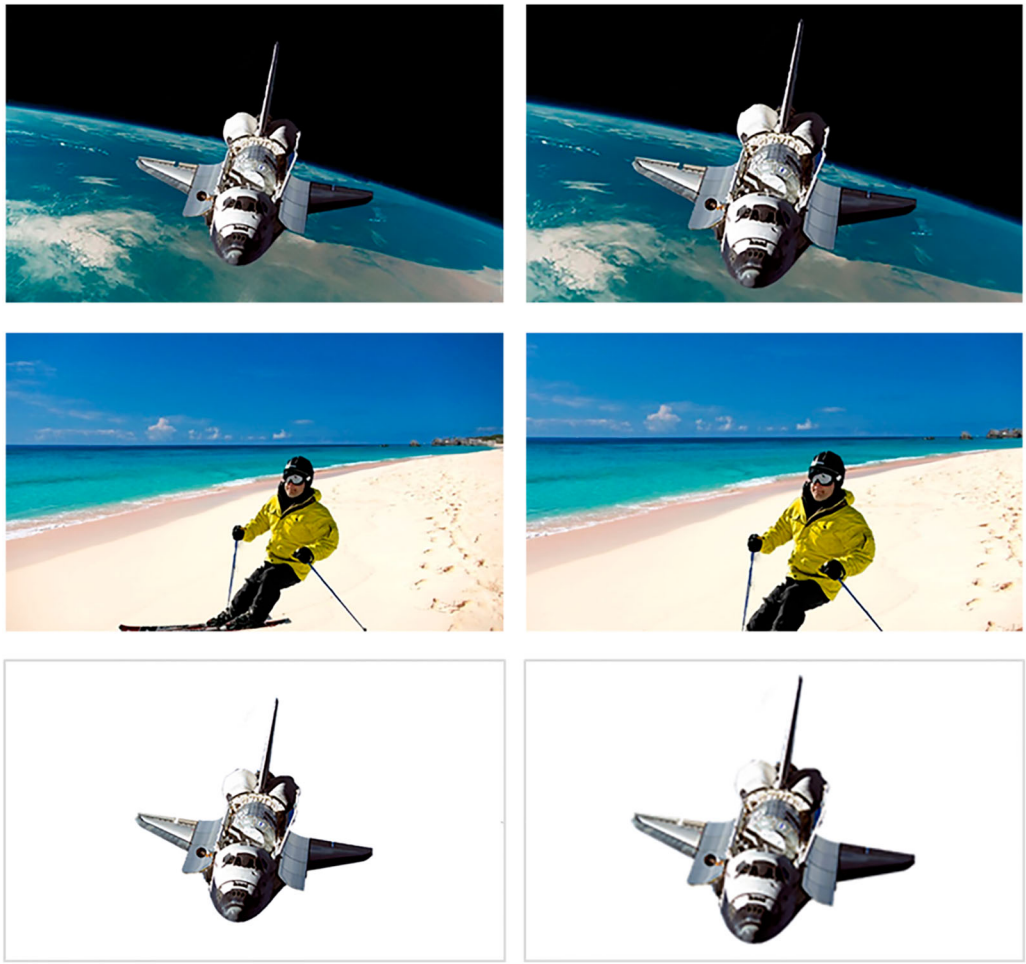

Figure 1. Example images in Experiment 1. Wide-angle (left) and close-up views (right). Top, consistent scene; middle, inconsistent scene; bottom, no-background condition.

To create consistent images, we pasted foreground objects onto semantically consistent background scenes. We carefully considered the geometrical appropriateness in scenes (e.g., size and/or position of scene elements) when we were creating consistent stimuli. We created three options for each one of 40 semantic categories (e.g., forest, underwater, etc.). These scenes were assessed in a pilot study in our previous work (Mamus, Boduroglu, & Gutchess, 2015). Using a scale from 1 to 7, 73 naive participants rated pictures in terms of semantic consistency and image quality. We selected 40 images from different categories that were rated above 4 in terms of both semantic consistency and image quality to be used in the study. To create 40 inconsistent images, we exchanged the foreground objects between two semantically consistent images. The position and the size of each object remained the same in both the semantically consistent and inconsistent versions of images. All final images were in JPEG format, and had a width of 750 pixels and height of 450 pixels. The stimuli were presented on a grey screen.

Close-up and wide-angle versions of each image were created based on the method used by Intraub and Dickinson (2008). Close-up versions were prepared by enlarging the wide-angle images by 20% and then by cropping the new image to its original wide-angle size. Thus, the sizes of the images remained the same across both versions, but wide-angle views had more background and smaller objects (see Figure 1 for examples).

For the scene condition, we randomly separated all 80 consistent and inconsistent images into two 40 image sets such that semantic consistency and close-up/wide-angle versions of the images were equally distributed. A foreground object or a background scenery never occurred more than once in each set. For the no-background condition, we used

40 images having same objects; the scenes of these images were replaced with white blank backgrounds. The no-background condition included close-up and wide-angle objects (see Figure 1).

PEBL matrix rotation task. The PEBL matrix rotation task (MRT) is a computerized version of the MRT in UTC-PAB Test developed by Perez, Masline, Ramsey, and Urban (1987). This task measures spatial rotation ability and short-term perceptual memory (Mueller & Piper, 2014). In this task participants saw a series of 6 by 6 cell matrices. Each matrix had six illuminated cells. Participants had to determine if the second matrix was a 90° (either left or right) rotated version of the first matrix. Participants had to press the left shift key if two matrices were the same; otherwise they had to press the right shift key on the keyboard. They received two practice trials together with the instructions. Accuracy and response time (RT) were measured over 20 test trials.

Object-Spatial Imagery and Verbal Questionnaire. Participants were asked to complete the OSIVQ (Blazhenkova & Kozhevnikov, 2009). This questionnaire consists of 45 items and three subscales. Participants were asked to rate each item on a 5-point Likert scale from 1 (totally disagree) to 5 (totally agree). They indicated their preferences for colourful, pictorial and high-resolution images for the object imagery scale, for schematic images, spatial relations amongst objects and spatial transformations for the spatial imagery scale, and verbal explanations and writing abilities for the verbal imagery scale. We calculated reliability scores based on our data coming from both experiments ($n = 206$). Cronbach's alphas were .84, .80 and .83 for object imagery, spatial imagery and verbal thinking, respectively.

Picture questionnaire. We also asked each participant to rate the pictures they studied via a computer survey. In the scene condition, participants rated each image in terms of semantic consistency (1 = highly inconsistent; 7 = highly consistent) and whether the elements making the scene consistent/inconsistent captured their attention (here onwards referred to as distinctiveness) (1 = highly indistinctive; 7 = highly distinctive). In the no-background condition, they were only asked to indicate how distinctive each object was.

Procedure

Participants individually sat in a dimly-lit room approximately 57 cm away from a 17-inch CRT monitor (60-Hz refresh rate) which was set to a resolution of 1280 × 1024 pixels. Each participant was randomly assigned to the scene or the no-background condition. Participants in the scene condition were randomly assigned to one of two sets. The procedure was the same for both the scene and no-background conditions except the nature of the pictures presented. The experiment started with the picture rating task and, immediately after, continued with the MRT, the OSIVQ, the picture questionnaire and demographic questionnaires. The total duration of the experiment was about 50 minutes.

Results

For each participant, we calculated the average BE ratings separately for all trial types. For the scene group, there were two critical variables: consistency (2-levels) and viewing condition (4-levels), resulting in eight trial types. For the no-background group, only viewing condition was manipulated resulting in four trial types. We determined whether there were any individuals with extreme responses by calculating the Cook's distance for all trial types. In the scene condition, seven participants' average ratings were at extreme in at least two out of eight trial types. In the no-background condition, two participants' average ratings were at extreme in at least two out of four trial types. After the exclusion of participants with extreme ratings, we were left with 73 participants in the scene condition and 38 participants in the no-background condition. Furthermore, we examined participants' confidence in their response to detect cases referring to no memory of the picture. Across all trial types, only on 2% of the trials participants reported that they did not remember the picture. This pattern was similar across both the scene and the no-background conditions. After the exclusion of these trials the pattern of the results did not change. Therefore, all trials were included in the analyses.

Boundary extension

We first present results from the scene condition, and then from the no-background condition. Finally, we discuss findings regarding confidence ratings from both conditions.

To determine whether our set of stimuli elicited BE, we first looked at the identical viewing conditions (WW and CC). For identical scenes, ratings below zero indicate extension. Therefore, we conducted one-sample *t*-tests (criteria = 0) and found that for both the WW ($M = -.16$, $SD = .36$) and the CC ($M = -.43$, $SD = .38$) conditions, people remembered the scenes as more extended, $t(72) = -3.88$, $p < .001$, $d = .44$; $t(72) = -9.57$, $p < .001$, $d = 1.13$, for WW and CC, respectively. In addition, participants were able to notice correctly when the view of the pictures were changed at test; they rated the scenes closer in the WC condition ($M = -1.04$, $SD = .46$), $t(72) = -19.49$, $p < .001$, $d = 2.26$, and wider in the CW condition ($M = .46$, $SD = .42$), $t(72) = 9.18$, $p < .001$, $d = 1.10$. Thus, participants followed the instructions of the task and correctly used the scale.

To determine whether contextual inconsistency between a foreground object and its background scene may have affected the occurrence and/or the magnitude of BE, we conducted a 2 (consistency: consistent, inconsistent) × 4 (viewing condition: WW, CC, WC, CW) repeated measures ANOVA. The results revealed that there was a main effect of consistency, $F(1,72) = 4.30$, MSE = .118, $p = .042$, $\eta_p^2 = .06$. Semantically consistent scenes ($M = -.32$, $SD = .27$) elicited more BE than semantically inconsistent scenes ($M = -.26$, $SD = .29$), $d = .25$. There was also a main effect of viewing condition, $F(3, 216) = 210.72$, MSE = .266, $p < .001$, $\eta_p^2 = .75$. As can be seen in Figure 2, the difference across viewing condition was similar to previous findings demonstrating that the CC condition elicited more BE than the WW condition, $p < .001$ (e.g., Intraub & Dickinson, 2008; McDunn et al., 2014).

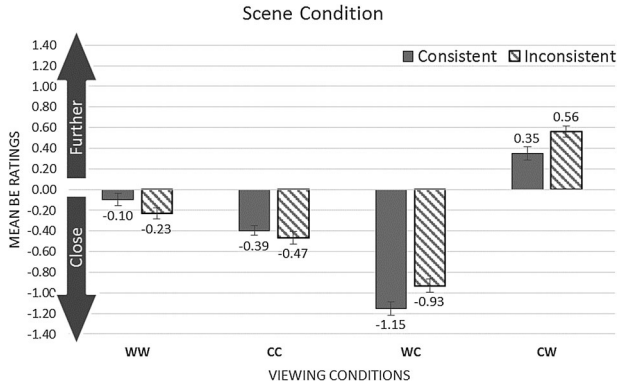

Figure 2. The effect of consistency on mean BE ratings for different viewing conditions. Error bars indicate standard error of mean.

More critically, these main effects were qualified by a significant interaction effect between consistency and viewing condition, $F(3, 216) = 6.41$, $p < .001$, $\eta_p^2 = .08$. This interaction was because consistent scenes elicited much greater BE compared to inconsistent scenes, but this difference emerged only from the different viewing conditions, in which the view of the pictures was changed. There was no significant difference between consistent and inconsistent scenes in the identical WW and CC conditions, *p*s > .05; but consistent scenes were rated much closer-up than inconsistent scenes in the different viewing conditions, $t(72) = -3.01$, $p = .004$, $d = .40$, $t(72) = -3.27$, $p = .002$, $d = .41$, for WC and CW conditions, respectively.

Since it was previously argued that people in the WC condition were disproportionately more biased to rate a picture as closer than a CW picture further (e.g., Dickinson & Intraub, 2008; Gottesman & Intraub, 2002), we wanted to test whether there was such an asymmetry in our data. In the WC condition, when the scene was already closer, correct recognition of the change would require negative ratings. So, negative ratings receding away from zero could be taken to indicate BE. On the other hand, in the CW condition, since the second scene is already extended, correct performance would require a positive rating. Therefore, in this condition, positive ratings approaching zero could be taken as indicating a tendency for BE. To test this idea, we followed the approach taken by Gottesman and Intraub (2002) who compared the absolute values of the ratings for the WC and CW conditions. To test whether the magnitude of ratings in these conditions were asymmetrical in our data, we transformed the ratings in the WC condition to a positive range by multiplying by (−1). A paired-sample *t*-test revealed that the responses in the WC and CW conditions were significantly different and thus asymmetrical, $t(72) = 8.54$, $p < .001$, $d = 1.32$. This overall pattern suggests that, in addition to the identical viewing condition, in the different viewing conditions there was a tendency towards BE; this tendency was stronger in the WC condition.

To determine whether BE was also elicited for the no-background condition, we conducted one-sample *t*-tests (criteria = 0). As can be seen in Figure 3, the results revealed that BE was significant in the CC condition ($M = -.46$, $SD = .48$), $t(37) = -5.84$, $p < .001$, $d = .96$; but not in the WW condition ($M = -.03$, SD

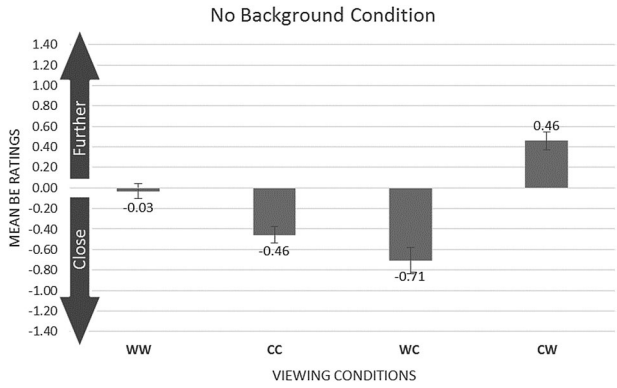

Figure 3. Mean BE ratings in the no-background condition. Error bars indicate standard error of mean.

= .45), $t(37) = -0.43$, $p = .666$. As in the scene condition, participants' ratings reflected that they correctly noticed that the view of the pictures were changed at test (for the WC condition $M = -.71$, $SD = .79$, $t(37) = -5.53$, $p < .001$, $d = .90$; and the CW condition $M = .46$, $SD = .53$, $t(37) = 5.35$, $p < .001$, $d = .87$). To test whether any asymmetry existed between different viewing conditions, we transformed the ratings in the WC condition to a positive range by multiplying by (−1). A paired sample *t*-test revealed that the responses in the WC and CW conditions tended to be asymmetrical in the expected direction, although this tendency was not significant, $t(37) = -1.76$, $p = .086$, $d = .37$.

Previous work had repeatedly shown that participants were highly confident in their ratings even though those ratings reflected BE (e.g., Gottesman & Intraub, 2002; McDunn et al., 2014). We wanted to determine whether the confidence and picture rating data were also dissociated in our sample. In both the scene ($M = 2.28$, $SD = .38$) and the no-background condition ($M = 2.32$, $SD = .25$), confidence ratings were above midpoint (range: 1–3) and there was no difference between the conditions ($p = .400$), replicating earlier findings (Gottesman & Intraub, 2002; McDunn et al., 2014). We also examined confidence ratings between consistent and inconsistent scenes across the four viewing conditions. A 2 (consistency: consistent, inconsistent) × 4 (viewing condition: WW, CC, WC, CW) repeated measures ANOVA revealed that there was a main effect of consistency, $F(1, 72) = 9.24$, $p = .003$, $\eta_p^2 = .11$ and a main effect of viewing condition, $F(3, 216) = 14.73$, $p < .001$, $\eta_p^2 = .17$, further qualified by a significant interaction, $F(3, 216) = 3.16$, $p = .025$, $\eta_p^2 = .04$. Participants were more confident in their response for consistent scenes ($M = 2.32$, $SD = .38$) compared to inconsistent scenes ($M = 2.25$, $SD = .40$), $d = .18$. The results indicated that although participants' task performance was better (as suggested by less BE), they were less confident in their responses for the inconsistent scenes. Post-hoc comparisons showed that confidence ratings were highest in the WC condition both for consistent (WW: $M = 2.23$, CC: $M = 2.29$, WC: $M = 2.52$, CW: $M = 2.24$) and inconsistent scenes (WW: $M = 2.11$, CC: $M = 2.21$, WC: $M = 2.36$, CW: $M = 2.28$), compared to the other three conditions (*p*s < .001 for all comparisons). This is in line with previous findings (e.g., Czigler, Intraub, & Stefanics, 2013). However, these results also revealed that, even in the WC condition, ratings were less for inconsistent scenes compared to consistent scenes ($p < .001$).

Individual differences in imagery

Participants completed the MRT and the OSIVQ. We examined whether imagery abilities as measured by the behavioural task and the self-report measures explained the magnitude of BE. For each participant, we calculated the scores for each one of the three subscales of the OSIVQ by averaging their responses for those subset items. To determine whether there was a correlation between visual and spatial imagery scores and BE ratings, we conducted Pearson correlations separately for CC and WW conditions for scenes (collapsed across consistent and inconsistent) and objects (no-background condition). In the identical viewing conditions, since negative ratings for these conditions always indicated BE, we collapsed ratings across CC and WW. There were no significant correlations in any of the analyses, all *r*s < .036, all *p*s > .212. We also looked at extreme groups (as in Munger & Multhaup, 2016) in the scene condition by focusing on the ratings of the top and bottom quartiles in object and spatial imagery. However, we did not find any difference in the BE ratings for the identical viewing conditions when participants were grouped in terms of their object imagery, $t(36) = 0.158$, $p = .875$, or their spatial imagery, $t(33) = -0.856$, $p = .398$. The same approach was taken for the no-background condition; we again did not find any imagery differences in the BE ratings for the identical viewing conditions when participants were grouped in terms of their object or spatial imagery, $t(18) = 0.539$, $p = .597$, $t(18) = 0.620$, $p = .543$, respectively. We also calculated participants' percent accuracy

and RT in the MRT. The BE ratings did not correlate with MRT performance measures in any of the conditions, all $rs < .29$, all $ps > .075$.

Perceived distinctiveness and image saliency on BE ratings

We created our scenes by pasting an object on top of a consistent background; inconsistent scenes were created by swapping the foreground objects between consistent scenes. To ensure that our approach to creating consistent and inconsistent scenes indeed resulted in such scenes, we asked participants to rate the scenes they had studied earlier. Participants' average ratings for consistent and inconsistent scenes were 6.67 ($SD = .31$) and 1.65 ($SD = .53$), respectively. Then, we wanted to determine whether consistent and inconsistent scenes differed in how much they captured one's attention as atypical and whether this contributed to the BE difference across scenes. We took a two-fold approach. First, we asked each participant to rate whether the elements making the scene consistent/inconsistent captured their attention (i.e., distinctiveness). Participants rated inconsistent scenes ($M = 5.6$, $SD = .45$) as more distinctive than consistent scenes ($M = 2.9$, $SD = .55$, $p < .001$). However, these differences in distinctiveness did not correlate with BE ratings in either consistent or inconsistent scenes, $r(40) = -.070$, $p = .670$; $r(40) = -.130$, $p = 423$, respectively.

Second, we also wanted to ensure that the consistent and inconsistent scenes created did not involve differential bottom-up cues. We computed saliency maps for all consistent and inconsistent scenes based on measures in the Graph-Based Visual Saliency (GBVS) algorithm (Harel, Koch, & Perona, 2006). To understand whether inconsistency between object and background created any boost in saliency scores, we calculated the object/background saliency ratio for both consistent and inconsistent scenes by dividing object saliency value by the remaining peripheral area saliency value. We checked whether saliency map scores differed between consistent and inconsistent scenes. There was no difference between consistent and inconsistent scenes in terms of object saliency, $t(78) = -0.100$, $p = .921$ and object/background ratio, $t(78) = -0.352$, $p = .726$, rendering it unlikely that low-level image characteristics lead to BE.

Discussion

In line with our predictions, we found that BE existed in both identical and different viewing conditions, and there was a small yet significant effect of scene consistency on the BE magnitude. Semantically consistent scenes elicited more BE than semantically inconsistent scenes, but this difference in the size of BE between consistent and inconsistent scenes came from the judgments in the different viewing conditions. In the different viewing conditions, consistent pictures were more likely to be rated as extending outwards than not. This was revealed by a more negative judgment in the WC condition and a positive yet closer to zero judgment in the CW condition for consistent than inconsistent pictures. We also found that BE was present for close-up scenes but not for wide-angle scenes in the no-background condition. In addition, we found an asymmetrical tendency towards BE between the WC and CW trials in the no-background condition, but this tendency was not significant. We replicated earlier findings demonstrating that close-up scenes elicited greater BE than wide-angle scenes did (e.g., Intraub & Dickinson, 2008). We also found a similar pattern in the no-background condition (e.g., Gottesman & Intraub, 2002). We also replicated earlier findings showing that confidence ratings were highest in the WC trials than the other three trials for all scenes (e.g., Czigler et al., 2013). Our results also revealed that confidence ratings were less for inconsistent scenes compared to consistent scenes. Critically, we were able to demonstrate that neither the perceived distinctiveness nor the differences in saliency maps of our images contributed to BE ratings.

Importantly, our results suggest that BE does not require the consistency between the foreground object and the background context. When there is a meaningful scene, independent of whether it is contextually consistent or not, people extend scenes. Considering these findings, we do not rule out the prediction account of the multisource model, which indicates that contextual knowledge helps individuals to create internal representations based on learned categories (Intraub, 2012). We merely state that contextual consistency is not necessary for prediction but that when scenes are consistent this allows more prediction and consequently leads to greater extension at least in different viewing conditions.

Although BE observed in the no-background condition may seem contradictory to the hypothesis of the multisource model, BE had been shown to be present in earlier studies presenting objects on blank backgrounds as well. Similar to our findings, both Gottesman and Intraub (2002) and McDunn et al. (2014) did not find BE in the WW condition for their stimuli on the blank background, but found BE in the CC condition. Furthermore, Gottesman and Intraub (2002) found a non-significant asymmetrical tendency between WC and CW trials for the photographs of real objects on the blank background. As they argued, even a pretence of continuity may be sufficient to elicit extension (Gottesman & Intraub, 2002). Therefore, pictures of real objects might lead blank backgrounds to be interpreted as part of a continuous scene even though they lack enriching details. Besides, it has also been shown that just the imagination of a coherent scene around single objects may be sufficient to trigger BE (Intraub et al., 1998).

Even though there has been empirical demonstrations that imagery-based instructions have resulted in BE in no-background conditions (Intraub et al., 1998), the individual differences evidence linking imagery abilities and BE magnitude has been equivocal at best (e.g., Munger & Multhaup, 2016). In this experiment, we found that individual differences in object and spatial imagery skills had no influence on BE ratings. There was also no relationship between BE ratings and mental rotation scores. Similar to the current results, Munger and Multhaup (2016) did not find any consistent relationship between BE magnitude and visual imagery. They reported that spatial imagery scores were correlated with BE, but only for CC trials and only in one of the six experiments (Experiment 3a). However, as we discussed earlier, the memory task they employed was spatial in nature and required participants to adjust the viewpoint of a test scene to match a studied one. Thus, the relationship they report may have been partly driven by the visuospatial demands of their memory task. While our results do not necessarily suggest that there is no link between imagery and BE, the tools we have utilized may not have been sufficient. Furthermore, individuals may not be strategically utilizing their visual and spatial imagery abilities while engaging in the memory task we employed. Alternatively, it may be possible that there is no or little relationship between BE and imagery abilities. A more direct test of this would be by comparing BE for scenes at different level of memorability.

Finally, we showed that no major perceptual or stimuli characteristics contributed to the difference in BE ratings across consistency manipulations. We used a bottom-up approach where we looked at potential differences in saliency maps across consistent and inconsistent scenes and showed that there were no differences across these scenes that could contribute to differences in BE ratings. Although participants reported inconsistent pictures more distinctive compared to consistent pictures, no effect of perceived distinctiveness was found. Thus, we ensured that the results did not reflect the effect of a potential confound in low-level stimulus characteristics. Even though consistent and inconsistent scenes were similar in low-level characteristics, it is possible that there were high-level differences in these scenes which might have modulated attention and memory and consequently impacted BE. In fact, we believe that this was the likely mechanism.

Experiment 2

In Experiment 1, we found that contextual inconsistency in scenes decreased the magnitude of BE and people extended scenes even when main objects did not fit a scene schema. We believe that the larger BE for consistent scenes could be explained by the multisource model; consistent scenes may have helped viewers create more schema-consistent, detailed and rich representations which in turn may have been hard to distinguish from the sensory experience. This may have contributed to a confusion of source and thus to BE. We also found BE in the no-background condition, replicating previous findings (Gottesman & Intraub, 2002). As discussed before, this may have been due to the ease of imagining a real object on a continuous background. In other words, in both inconsistent scenes and the no-background condition, the presence of a central object may have led people to represent visual continuity of the background; that might have been sufficient to create extension. To eliminate all such possible cues and scene schema completely, in Experiment 2 we generated meaningless shapes based on the real objects in the first experiment (see Figure 4). These meaningless shapes were multi-component and complex but at the same time similar in size

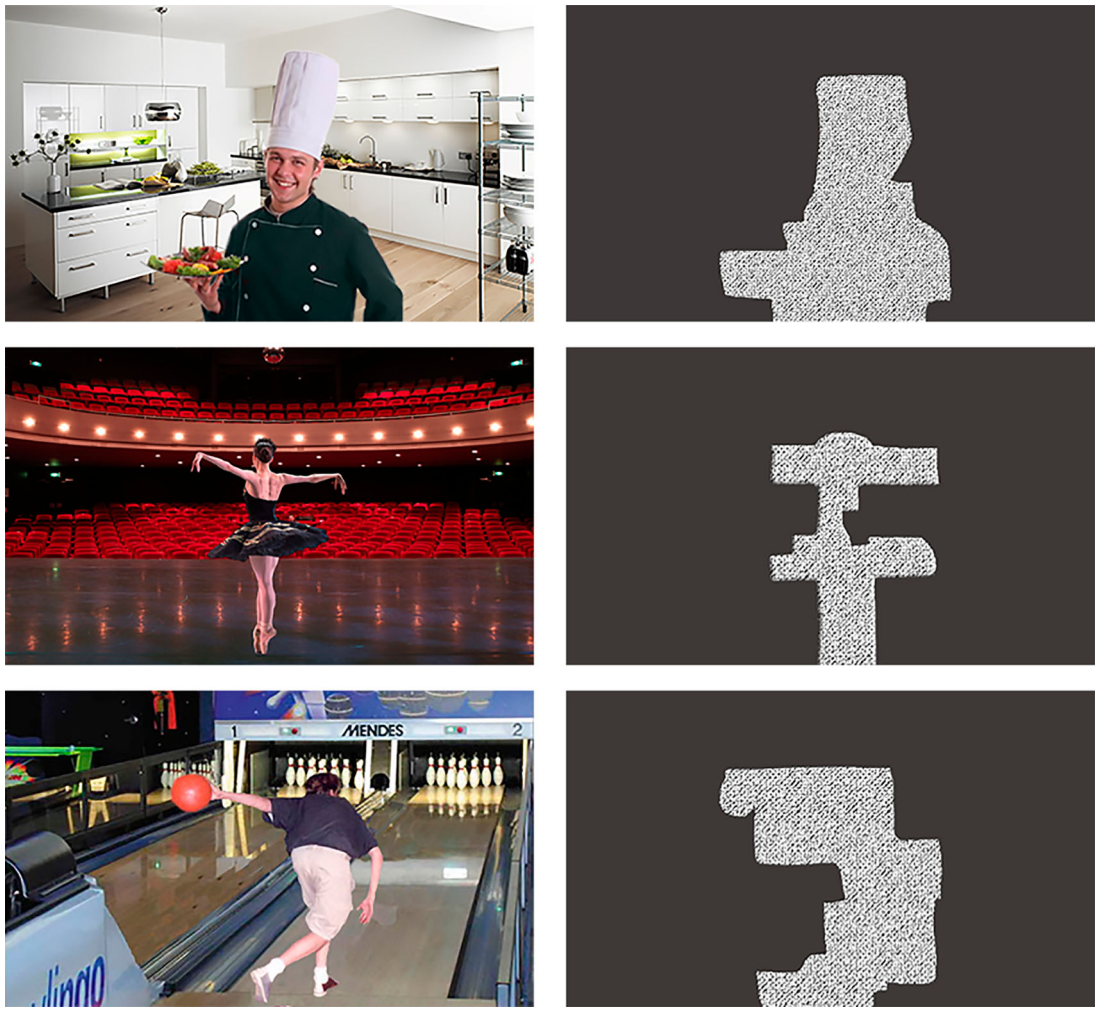

Figure 4. Example images from Experiment 1 (left) and their distorted counterparts used in Experiment 2 (right).

and shape with the central objects in Experiment 1. This approach is akin to the approach taken by McDunn et al. (2014); the most critical differences between our stimuli and theirs was that the objects in our stimuli were more complex and multi-component (as judged by the examples provided in Figure 2 in McDunn et al., 2014) and were anchored to the bottom of the image, unlike theirs that were placed in the middle of the display (allowing for perceptions of floating objects in space). Also, our objects were filled-in on grey backgrounds; they instead had filled-in objects on white backgrounds. Thus, in our stimuli, the contrast between the foreground object and background was not as salient. Furthermore, the anchoring of the central object to the bottom of the image may have rendered a perception of an object floating in space less likely; rather, our stimuli (as can be seen in Figure 4) may have been more likely to elicit the 2D pictorial representation of an object as opposed to being in 3D space.

Given the differences between our study and McDunn et al. (2014), we did not necessarily expect to replicate their findings of BE with abstract shapes. Rather, we predicted that the lack of a meaningful object (or background) may eliminate BE. We also further explored the relationship between different visual and spatial imagery, and visuospatial performance measures and BE.

Method

Participants

A total of 86 Turkish undergraduate students (M = 20.13 years (18–24), SD = 1.2, 75% female) at Bogazici University participated in the experiment in exchange for extra credit in their psychology courses. All had

normal or corrected-to-normal vision and provided written informed consent.

Materials

Picture rating task. Same as the task in Experiment 1, except the nature of the images. There were two between subject conditions which were the grey background and the white background conditions. Both conditions contained semantically meaningless shapes with the only difference being the colour of the background. We presented each image for 15 s in the study phase. Images in the recognition test were presented in the same order as in the study phase. As in Experiment 1, there were four trial types in the recognition test: CC, WW, CW and WC.

Pictures. The stimuli consisted of semantically meaningless filled shapes that were created from slightly distorted contours of the same target objects in the first experiment (see Figure 4 for examples). These patterns were filled by black-and-white texture and were displayed centrally on a blank white or grey background. A total of 40 images were prepared with their close-up and wide-angle versions in a similar manner with the first experiment. All images were prepared using Adobe® Photoshop® CS6. They were in JPEG format, and had a width of 750 pixels and height of 450 pixels.

The MRT and the OSIVQ were administered similarly as in Experiment 1. In the picture questionnaire, participants answered two questions regarding the images presented in the memory task. First, we asked the participants to rate the images on a 7-point Likert scale in terms of distinctiveness (1: highly indistinctive; 7: highly distinctive). Second, we asked them to report whether the shapes in the images evoked any real objects while viewing the images and, if so, they were asked to name them.

Procedure

The procedure was the same as in the first experiment, except for the nature of the pictures presented. Participants individually sat in a dimly-lit room approximately 57 cm away from a 17-inch CRT monitor (60-Hz refresh rate) which was set to a resolution of 1280×1024 pixels. Each participant was randomly assigned to the grey background or the white background condition.[1] The procedure was the same for both conditions. The experiment started with the picture rating task and, immediately after, it continued with the MRT, the OSIVQ, the picture questionnaire and demographic questionnaires. The total duration of the experiment was about 50 minutes.

Results

Data was analysed similarly to Experiment 1. For each participant, we calculated the average BE ratings separately for all trial types. There was only the viewing condition variable, and therefore four different trial types. We determined whether there were any individuals with extreme responses by calculating the Cook's distance for all trial types. One participant's average ratings were at extreme in at least two out of four trial types; after the exclusion of this participant, we were left with a total of 85 participants. Furthermore, we examined participants' confidence in their response to detect cases referring to no memory of the picture. Across all trial types, on average participants reported that they did not remember the picture on 8.5% of the trials. After the exclusion of seven participants whose "don't remember" trials were above 20%, the pattern of the results did not change. Therefore, all trials were included in the analyses.

Boundary extension

To determine whether our meaningless stimuli elicited BE, we first conducted one-sample *t*-tests (criteria = 0) for the identical viewing conditions in which pictures were presented in the same view (WW and CC). For identical scenes, ratings below zero would indicate extension whereas ratings above zero would indicate restriction. As expected, there was no BE in the identical viewing conditions ($M = .19$, $SD = .37$), $t(84) = 4.74$, $p < .001$, $d = .51$ and ($M = -.09$, $SD = .45$), $t(84) = -1.82$, $p = .072$, $d = .20$ for WW and CC conditions, respectively. In addition, participants were able to correctly identify when the view of the pictures were changed at test; they rated the scenes closer in the WC condition ($M = -.47$, $SD = .51$), $t(84) = -8.44$, $p < .001$, $d = .92$ and wider in the CW condition ($M = .43$, $SD = .46$), $t(84) = 8.60$, $p < .001$, $d = .93$ (see Figure 5). Thus, as in Experiment 1, participants followed the instructions of the task and correctly used the scale.

As we discussed in greater detail in Experiment 1, there may be some asymmetry in BE magnitude across the CW and WC conditions (Dickinson & Intraub, 2008; Gottesman & Intraub, 2002). In the WC

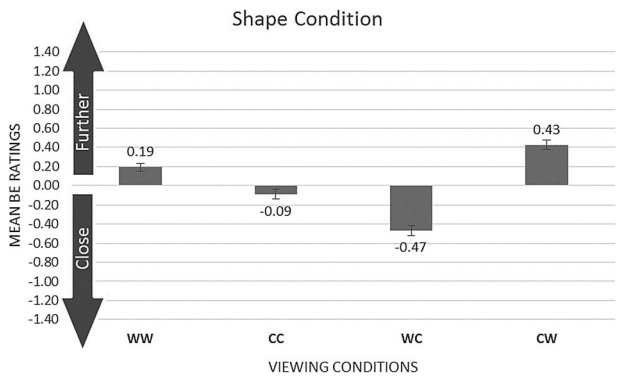

Figure 5. Mean BE ratings in the shape condition. Error bars indicate standard error of mean.

condition, negative ratings receding away from zero could be taken to indicate BE. On the other hand, in the CW condition, positive ratings approaching zero could be taken as indicating a tendency for BE. As in Experiment 1, to test whether any asymmetry existed between different viewing conditions, we transformed the ratings in the WC condition to a positive range by multiplying by (−1). A paired-sample *t*-test revealed that the responses in the WC and CW conditions were not different, thus symmetrical $t(84) = 0.54$, $p = .590$.

Previous work had repeatedly shown that participants were highly confident in their ratings even though those ratings reflected BE (e.g., Gottesman & Intraub, 2002; McDunn et al., 2014). Similarly, in Experiment 1 we found that participants' confidence decreased when their task performance was better (as suggested by less BE). Thus, we compared the groups across the experiments to determine how participants' confidence ratings would change when there was no BE. To test whether the participants' confidence in the shape condition differed from confidence in the scene and no-background conditions, we conducted a one-way ANOVA. The results showed that participants in the shape condition ($M = 1.94$, $SD = .41$) were less confident in their responses than participants in the scene ($M = 2.28$, $SD = .38$) and no-background conditions ($M = 2.32$, $SD = .25$), $F(2,193) = 19.73$, $MSE = .135$, $p < .001$, $\eta_p^2 = .17$. Thus, regarding the confidence pattern, we replicated our finding in Experiment 1.

Individual differences in imagery

As in Experiment 1, we measured participants' imagery skills by applying the MRT and the OSIVQ. For each participant, we calculated the scores for each of the three subscales of the OSIVQ by averaging their responses for those subset items. To determine whether BE ratings in the identical viewing conditions differed regarding imagery scores, we conducted Pearson correlations. However, the BE ratings in the identical viewing conditions correlated with neither object nor spatial imagery where $r(85) = -.015$, $p = .895$ and $r(85) = -.171$, $p = .118$, respectively. We also calculated participants' percent accuracy and RT in the MRT. The results showed that percent accuracy was negatively correlated with BE ratings in the identical viewing conditions $r(85) = -.324$, $p = .003$ but RT did not $r(85) = -.032$, $p = 772$. Thus, participants with higher spatial visualization ability as measured by mental rotation accuracy were more likely to extend scenes in the direction of BE even though their ratings did not indicate BE.

Discussion

In line with our predictions, semantically meaningless stimuli on blank background did not elicit BE in the identical viewing condition. We demonstrated that when the foreground object did not make it possible to imagine a continuous context, BE was unlikely. Interestingly, we also found that participants who showed better spatial rotation ability were more likely to extend pictures.

Our results supported the multisource model but these were in conflict with McDunn et al. (2014)'s findings which showed BE even in abstract images consisting of irregular polygons. However, as we highlighted in the introduction to Experiment 2, there were critical methodological differences between the two studies. Most critically, they argued that the placement of objects on random-dot backgrounds might have helped observers by creating "some sense of a spatial context" (p. 374). The findings of Hale et al. (2015) supported this possibility. However, in our study, the objects were always placed at the bottom and centrally and almost all of our objects intersected the bottom boundary of the images. Thus, such placement may have acted as an anchor that prevented extension. It must be noted that in previous studies BE was similarly elicited for meaningful objects cropped or non-cropped at scene boundaries (e.g., Intraub & Bodamer, 1993); in those instances there were meaningful objects that easily triggered a

scene schema and helped viewers establish rich internal representations. In other words, cropping may have different consequences for meaningful objects versus meaningless abstract shapes.

The lack of BE in Experiment 2 could also be explained by a normalization process. Recently, McDunn et al. (2016) demonstrated in their Experiment 2 that scene normalization (as opposed to BE) when central salient objects were presented on identical and relatively uniform backgrounds (a basketball on a table covered in a black tablecloth in front of a wall). They argued that when the backgrounds were identical across different scenes, differentiating scenes based on background information may have become challenging and instead participants may have created a normalized, summary representation of the background contexts presented. Thus, when there was no distinguishing spatial context, BE was eliminated. In this regard, our Experiment 2 might be thought to be similar in that the filled-in objects were always presented on blank grey backgrounds. The fact that we find no evidence of BE with our abstract objects in Experiment 2 may also be due to such a normalization process. However, at this point, it is not possible for us to determine with these results whether the lack of BE is due to an anchoring effect as described above or a normalization effect reported by McDunn et al. (2016). Future research is necessary to understand possible mechanisms eliminating BE.

General discussion

To our knowledge, this study is the first to test the role and the necessity of context information on BE phenomenon by violating scene consistency and focusing on scene schema. In Experiment 1 we examined how semantic consistency between object and background would influence BE. Along with consistent scenes, we used inconsistent scenes that have been known to decrease accuracy in scene categorization and object perception (Biederman et al., 1982; Davenport, 2007; Davenport & Potter, 2004; Joubert et al., 2007). In Experiment 2 we aimed to eliminate scene schema completely from images, and thus we created images including distorted filled shapes (i.e., distorted contours of the same target objects in Experiment 1). The results showed that although scene schema is necessary to elicit BE, contextual consistency is not required but can influence BE under some conditions (e.g., WC vs CW).

Our findings are consistent with the multisource model. According to this model, scene schema and contextual knowledge are defined as two of the main contributors to BE; these factors play a critical role in helping individuals create internal representations (Intraub, 2012). We supported this model by demonstrating that when scene schema and semantic meaning were eliminated from the images, as in Experiment 2, BE did not occur. We showed that scenes with semantic inconsistency and no-background objects, which less likely supported scene continuity, also elicited BE, but the magnitude of BE for these two types of scenes was less.

We also replicated earlier studies by showing BE for real objects on blank backgrounds. While this finding seems to contradict predictions drawn from the multisource model, earlier studies had showed that real objects on blank backgrounds could elicit BE (e.g., Gottesman & Intraub, 2002). It has been argued even a pretence of continuity may be sufficient to elicit extension even though scenes lacked enriching elements such as background details (Gottesman & Intraub, 2002). Thus, Intraub and colleagues concluded that any interpretation of a background as a continuous surface is necessary in eliciting BE. Our findings on the relation between BE and imagery abilities may be in line with these claims. Visual spatial imagery and visuospatial ability become partly influential only when there is no or little scene element to induce interpretation of background as a continuous surface. Indeed, in Experiment 1 we found that, when individuals were presented with natural scenes and real objects, their imagery skills did not explain their BE ratings. However, in Experiment 2, when images consisted of semantically meaningless shapes, individuals with higher spatial visualization ability measured via mental rotation performance had a tendency to rate scenes more extended compared to ones with lower spatial ability. Future research is necessary to determine individual differences in visual and spatial imagery and visuospatial ability on BE. Evidence gathered by Munger and Multhaup (2016) and our results suggests that subjective measures of visual and spatial imagery ability may not be the best predictors of BE.

The results indicated that the consistent pictures were more likely to be rated as extended than inconsistent pictures in the different viewing conditions but not in the identical viewing conditions. We also showed that BE in the identical and the different viewing conditions might differ from each other in relation to scene consistency. Previous studies investigated the different viewing trials either in terms of asymmetry between the magnitude of BE in WC and CW trials, or as a task control (Dickinson & Intraub, 2008; Gottesman & Intraub, 2002; Intraub et al., 1998). Furthermore, our results suggest that it is also worthwhile to investigate how different viewing trials are deviating from identical trials in order to better understand observers' internal schemas.

While we are excited to highlight the role of semantic consistency on BE, we are aware that there may be some methodological limitations that necessitate further replications of the presented findings. When we were preparing the stimuli set for this study, we adapted a practice commonly used by researchers interested in scene consistency: a target object is pasted either onto a consisted (related) or inconsistent (unrelated) scene (e.g., Gutchess, Welsh, Boduroglu, & Park, 2006). However, we were not able to fully control factors like object size, distance depth cues and idiosyncratic preferences, all of which have been shown to influence BE (Bertamini, Jones, Spooner, & Hecht, 2005; Konkle & Olivia, 2007; Lukavský, 2014). Future research is necessary to determine if any one of these factors interact with the observed consistency effects on BE.

In summary, with two experiments we demonstrated that scene schema is necessary for BE existence, but contextual consistency is influential only on the magnitude of BE. We also suggest that spatial imagery skill may foster greater BE but only when picture elements do not persuade observers to imagine a continuous context. Further exploration of individual differences in imagery skills and differences of trial types is needed to illuminate how BE supports scene processing.

Note

1. We compared the results to discover if the colour of the blank background (white or grey) affects the BE ratings. There was no difference between the two colours; thus, we combined the data for the analyses.

Acknowledgments

We would like to thank Dr Alper Acik for help on saliency related analyses and Dr Ali Tekcan, two anonymous reviewers and Dr Jiri Lukavsky for comments on earlier versions of this manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Aysecan Boduroglu was supported by a BAGEP award by the Science Academy; TUBITAK [grant number 113K217].

References

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

Bertamini, M., Jones, L. A., Spooner, A., & Hecht, H. (2005). Boundary extension: The role of magnification, object size, context, and binocular information. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(6), 1288–1307.

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

Blajenkova, O., Kozhevnikov, M., & Motes, M. A. (2006). Object-spatial imagery: A new self-report imagery questionnaire. *Applied Cognitive Psychology*, *20*, 239–263.

Blazhenkova, O., & Kozhevnikov, M. (2009). The new object-spatial-verbal cognitive style model: Theory and measurement. *Applied Cognitive Psychology*, *23*(5), 638–663.

Czigler, I., Intraub, H., & Stefanics, G. (2013). Prediction beyond the borders: ERP indices of boundary extension-related error. *PloS One*, *8*(9), e74245. doi:10.1371/journal.pone.0074245

Davenport, J. L. (2007). Consistency effects between objects in scenes. *Memory & Cognition*, *35*(3), 393–401.

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

Dickinson, C. A., & Intraub, H. (2008). Transsaccadic representation of layout: What is the time course of boundary extension? *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 543–555.

Gagnier, K. M., & Intraub, H. (2012). When less is more: Line drawings lead to greater boundary extension than do colour photographs. *Visual Cognition*, *20*(7), 815–824.

Gottesman, C. V. (1998). *The role of the perceptual schema in the representation of visual scenes* (*Unpublished doctoral dissertation*). University of Delaware, Newark, USA.

Gottesman, C. V., & Intraub, H. (2002). Surface construal and the mental representation of scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 589–599.

Greene, M. R., & Oliva, A. (2009). The briefest of glances: The time course of natural scene understanding. *Psychological Science*, *20*(4), 464–472. doi:10.1111/j.1467-9280.2009.02316.x

Gutchess, A. H., Welsh, R. C., Boduroglu, A., & Park, D. C. (2006). Cultural differences in neural function associated with object processing. *Cognitive, Affective, & Behavioral Neuroscience*, *6*, 102–109.

Hale, R. G., Brown, J. M., McDunn, B. A., & Siddiqui, A. P. (2015). An influence of extremal edges on boundary extension. *Psychonomic Bulletin & Review*, *22*(4), 961–966.

Harel, J., Koch, C., & Perona, P. (2006). Graph-based visual saliency. *Advances in Neural Information Processing Systems*, 19, 545–552.

Intraub, H. (2012). Rethinking visual scene perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, *3*, 117–127.

Intraub, H., & Bodamer, J. L. (1993). Boundary extension: Fundamental aspect of pictorial representation or encoding artifact? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 1387–1397.

Intraub, H., Daniels, K. K., Horowitz, T. S., & Wolfe, J. M. (2008). Looking at scenes while searching for numbers: Dividing attention multiplies space. *Perception & Psychophysics*, *70*, 1337–1349.

Intraub, H., & Dickinson, C. (2008). False memory 1/20th of a second later: What the early onset of boundary extension reveals about perception. *Psychological Science*, *19*, 1007–1014.

Intraub, H., Gottesman, C. V., & Bills, A. J. (1998). Effects of perceiving and imagining scenes on memory for pictures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 186–201.

Intraub, H., & Richardson, M. (1989). Wide-angle memories of close-up scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 179–187.

Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision Research*, *47*, 3286–3297.

Konkle, T., & Olivia, A. (2007). Normative representation of objects: Evidence for an ecological bias in object perception and memory. *Proceedings of the 29th Annual Meeting of the Cognitive Science Society*. Lawrence Erlbaum Associates, 407–413.

Lukavský, J. (2014). Changes in boundary extension effect during spatial confinement. *Visual Cognition*, *22*(7), 996–1012.

Mamus, A. E., Boduroglu, A., & Gutchess, A. (2015, November). *Cross-cultural differences in conceptual scene processing*. Poster session presented at the Psychonomic Society Annual Meeting, Chicago, IL.

Markham, R., & Hynes, L. (1993). The effect of vividness of imagery on reality monitoring. *Journal of Mental Imagery*, *17*, 159–170.

Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology*, *64*, 17–24.

McDunn, B. A., Brown, J. M., Hale, R. G., & Siddiqui, A. P. (2016). Disentangling boundary extension and normalization of view memory for scenes. *Visual Cognition*, *24*(5-6), 356–368.

McDunn, B. A., Siddiqui, A. P., & Brown, J. M. (2014). Seeking the boundary of boundary extension. *Psychonomic Bulletin & Review*, *21*(2), 370–375.

Mueller, S. T., & Piper, B. J. (2014). The psychology experiment building language (PEBL) and PEBL test battery. *Journal of Neuroscience Methods*, *222*, 250–259.

Munger, M. P., & Multhaup, K. S. (2016). No imagination effect on boundary extension. *Memory & Cognition*, *44*(1), 73–88.

Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition*, *3*, 519–526.

Perez, W. A., Masline, P. J., Ramsey, E. G., & Urban, K. E. (1987). Unified tri-services cognitive performance assessment battery: Review and methodology, DTIC Document ADA181697.

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

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 509–522.