

The Effects of Scene Category and Content on Boundary Extension

Katie Gallagher (katieg@mit.edu)

Benjamin Balas (bjbalas@mit.edu)

Jenna Matheny (jnmath@mit.edu)

Pawan Sinha (psinha@mit.edu)

Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology
Cambridge, MA 02139-4307, USA

Abstract

Boundary extension is a robust scene perception phenomenon in which observers erroneously remember seeing portions of a previously viewed scenes beyond the image boundaries. The effect is strong and reliable, and has been demonstrated for a wide variety of observers. Various studies have shown that boundary extension is distinct from amodal completion and subject to certain pictorial constraints that define an image as being a “scene.” We examine here how scene content influences boundary extension. We demonstrate that the presence of foreground objects appears to enhance the extension effect, and that scene complexity also plays a role in determining the magnitude of the extension effect. These results are discussed with regard to the scene context hypothesis of boundary extension and recent results in scene perception.

Introduction

When asked to recall a picture of a scene, observers have a strong tendency to remember seeing a wider-angle view of that scene than they actually did. (Intraub & Richardson, 1989) This phenomenon is known as “boundary extension.” (Figure 1).



Figure 1: An example of the boundary extension effect. Subjects viewing the left image tend to remember seeing a scene more closely resembling the right image later.

The extension effect is quite robust. Observers who are aware of the effect and told to guard against it persist in extending image boundaries (Intraub & Bodamer, 1993). Both short and long presentation times support the effect, including display conditions designed to approximate

saccadic viewing of a scene (Intraub, Gottesman, Willey, & Zuk, 1996). In terms of developmental stability, boundary extension has been found across a range of age groups encompassing young children to older adults (Candel, Merckelbach, Houben, & Vandyck, 2004; Seamon, Schlegel, Hiester, Landau, & Blumenthal, 2002). Overall, the tendency to remember details beyond the boundaries of a scene is widespread and very reliable. In terms of its generality, recent results indicate that the effect is not restricted to the visual modality. Haptic exploration of tactile scenes by both blind and sighted individuals results in extension of non-visual boundaries (Intraub, 2004).

The most prominent explanation for the boundary extension effect is the scene context hypothesis. This theory suggests that observers form expectations of scene content for regions beyond picture boundaries according to some perceptual schema (Intraub & Bodamer, 1993). It should be noted that this is not the same as suggesting that subjects default to a prototypic view of a scene. Experiments with close-up, prototypic, and wide-angle scenes have demonstrated that observers do not simply remember a scene prototype (Intraub & Berkowitz, 1996). The observers’ version of the remembered scene is augmented with new content generated by this schema, leading to an extension effect (Ohara & Kokubun, 2002). During picture viewing, observers continuously create a sketch of scene details beyond the boundaries and complete truncated surfaces.

Though powerful, the scene context hypothesis does not address the question of how the visual content of an image affects the amount of boundary extension that observers experience. Previous studies have demonstrated that certain image conditions do not support the effect, but these conditions are related to the appearance of an image as a “scene” rather than a picture. For example, neither line-drawings of objects on blank backgrounds nor photographic objects placed on blank backgrounds elicited boundary extension (Gottesman & Intraub, 2002). Similarly, object boundaries (formed when one object circumscribes another in an image) do not give rise to extension (Gottesman & Intraub, 2003). In both cases, the lack of a perceived continuous world beyond the imposed boundary may be to blame. This explanation does not speak directly to visual processing, but rather to more abstract notions of how observers understand a picture as a depiction of some

natural scene with rich spatial information. Examinations of boundary extension related to emotional valence of images are closer to our goal of understanding how visual content affects the extension effect. We note in particular a recent result suggesting that high-anxiety individuals are less susceptible to boundary extension when viewing emotionally arousing scenes (Matthews & Mackintosh, 2004). This reduction did not appear in normal observers, however (Candel, Merckelbach, & Zandbergen, 2003), indicating that the change in effect size may have had more to do with the emotional state of the observer than the image content.

In the current study, we ask whether or not simple changes in scene category and content give rise to changes in the occurrence and magnitude of boundary extension. Rather than investigating complex relationships between object and image boundaries, we turn to intuitive categories of scenes such as “indoor” v. “outdoor” scenes. Computational studies of scene classification suggest that scene categories such as these (as well as finer-grained categories such as “urban” and “mountain”) can be discriminated by means of relatively low-level features. (Oliva & Torralba, 2001; Torralba & Oliva, 2003) Likewise, the presence of a foreground object in a scene introduces dramatic differences in the global distribution of oriented edges compared to a background-only scene. (Johnson & Olshausen, 2003) In both cases, differences in low-level luminance structure and differences in the way attention is distributed across an image may result in varying amounts of boundary extension across scene categories.

In Experiment 1, we test this hypothesis by measuring the amount of extension for indoor and outdoor scenes that either contain a prominent foreground object or contain no such focal point. In terms of purely visual content, indoor scenes differ from outdoor scenes in terms of their “spatial envelope” (Torralba & Oliva, 2003), meaning that in an objective sense they are distinct image categories. Indoor and outdoor scenes are also interesting to study in the context of boundary extension because in an abstract sense, indoor scenes are enclosed within fixed borders (such as walls, ceilings, and floors). We hypothesize that boundary extension may take place to a lesser extent after viewing indoor scenes because of this enclosure. Subjects’ tendency to remember wider-angle views of a scene may be lessened by their implicit knowledge that they are viewing a scene that is circumscribed by real-world boundaries. The presence of objects in a scene also alters raw visual content and higher-level aspects of scene perception. In particular, we suggest that placing a foreground object in a natural scene may draw attention to that object at the expense of time spent viewing the background. Under these circumstances, we expect that less boundary extension would take place. “Tunnel vision” may dominate in such situations, as the observers’ memory of object details may elicit post-viewing “zooming in” to interesting object features.

In Experiment 2, we examine the additional influence of background and foreground salience. We utilize matched scenes of varying complexity that contain either chairs or people seated in chairs to test the hypothesis that allocation of attentional resources across the image may affect the extension process. Human faces and bodies are extremely salient visual stimuli, and contain a high level of detail that observers are very sensitive to. By contrast, chairs are a close match to the overall shape of a seated human, but lack the same level of interesting detail that we find in the biological stimulus. Given this, we expect that scenes with faces in them may give rise to smaller amounts of boundary extension due to a “zooming in” effect brought on by enhanced attention to the foreground object. Similarly, we expect that highly complex backgrounds that draw attention may give rise to increased boundary extension. Increased attention to the background of a scene may enhance the drive to generate new content beyond the image boundaries.

In both experiments, we use a very simple paradigm to quantify the amount of extension that takes place. Subjects are shown a series of target images at a relaxed but steady pace. After this initial presentation, subjects are shown a series of test images that consist of veridical, zoomed-in and zoomed-out versions of each original image. They are then asked to select the image that they believe they were previously shown. In this manner, we may assess both the presence and the degree of extension for the image categories we consider in our two experiments.

Experiment 1

In this first experiment, we investigate the degree of boundary extension after viewing indoor and outdoor scenes, as well as the effect of a prominent foreground object on the phenomenon.

Subjects

29 subjects (14 men, 15 women) were recruited from the MIT undergraduate population for this task. Subject age ranged from 18-24 years) All subjects had normal or corrected-to-normal vision, and were naïve to the purpose of the experiment.

Stimuli

A series of 24 images were collected from various sources for this experiment. Scenes depicted outdoor or indoor scenes that either contained one clear foreground object or no such focal point.

Eight test images were created by cropping and resizing each original image to construct “zoomed in” versions. Digital zooming (as opposed to true optical zooming) has been shown to have no significant influence on the extension effect (Wan & Simons, 2004), and was thus employed here for convenience.

Procedure

Subjects were seated at a comfortable viewing distance from the monitor, and were told that they were about to view series of images. Their instructions were to study each image carefully as they would be asked to perform a “difficult memory task” after viewing each image once.

The full set of images was displayed serially in a random order using Microsoft PowerPoint. Each image was displayed for 10 seconds before it was replaced by the next image in the sequence.

Following exposure to all target images, subjects were then told that they would view the 8 test images corresponding to each original image (Figure 2). Their goal was to select the test image that they believed was a perfect match to the target image they had already viewed. Each set of test images was presented in order from widest-angle view (1) to narrowest (8). The target image was always either image 4,5, or 6, allowing subjects room to make errors in either direction. Subjects were allowed as much time as necessary to choose which test image they believed corresponded to the original target.



Figure 2: The eight test images created for stimulus C in our task. The target image in this case was image #4.

Results

The difference between the location of the selected test image and the original target image on the 1-8 spectrum defined previously was determined for each item. Boundary extension was quantified by subtracting the zoom level of the selected test image from the zoom level of the target image. The average amount of extension by condition was determined for each subject, and a 2x2 repeated measures ANOVA was conducted with scene category (indoor v. outdoor) and foreground object (present v. absent) as factors. A graph of the results is displayed in Figure 3.

Our first hypothesis prior to analysis was that the circumscribed world boundaries implicit in an indoor scene

would yield less boundary extension than outdoor scenes. Second, we expected less boundary extension to occur when viewing images containing a foreground object, due to attention being disproportionately allocated to the focal point of the scene. As evidenced by the mean values in Figure 3, neither of these hypotheses are confirmed by our data. Our ANOVA yields only a main effect of object v. no-object, such that scenes containing a focal foreground object result in significantly more boundary extension. ($p < 0.05$) Scene category had no effect, and there was no significant interaction between the two factors.

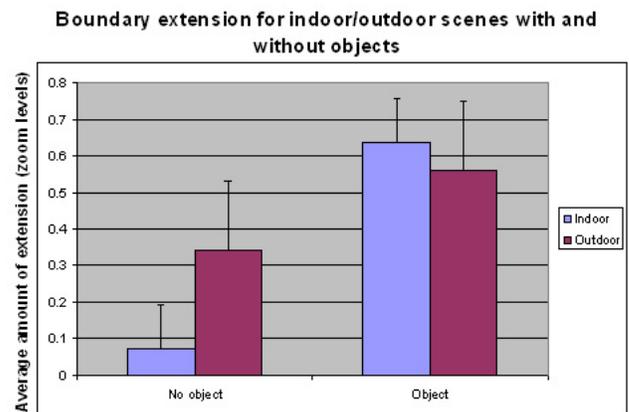


Figure 3: Mean extent of boundary extension in terms of zoom levels for indoor/outdoor scenes with a foreground object or without.

Discussion

Contrary to our hypothesis, there was no significant difference in the amount of boundary extension for indoor v. outdoor scenes. More importantly, there was actually more boundary extension for images that contained a clear foreground object. We expected that a “zooming in” effect might take place due to increased attention to the focal point of the scene, but we see the reverse effect instead.

Could both of these results be an extension of the finding that wide-angle views produce less boundary extension than close-ups? It may seem that scenes containing objects might be necessarily considered close-ups due to the central positioning of the object, but we believe our stimuli to be quite varied in this regard. Our focal objects were depicted at a range of scene depths, from very close to quite far away. Moreover, they were also depicted in environments ranging from panoramic to relatively claustrophobic, with no clear item effects emerging.

Second, our objects were not truncated by the image boundary in any of the target images. Therefore, any object completion processes would not have been able to enhance the extension effect given the target scenes as input. Given both of these qualities of our stimulus set, we suggest that the result reported in this first experiment represents a true influence of scene content on boundary extension. The

presence of objects may operate independently of other factors that influence extension, such as wideness of viewing angle.

To explain this result, we suggest a modification of our initial attentional hypothesis for “zooming in” effects. Our naïve assumption regarding the role of attention in boundary extension was that increased attention to a particular object in a scene would lead to a “tunnel memory” for details of that object. The emphasis on object detail could lead to the selection of test images in which those details were very apparent, i.e. “zoomed in” images. However, let us consider what processes might apply to the background of an image which contains a focal object. If attention is indeed biased towards the foreground, it stands to reason that fewer details about the background are retained. We believe this might lead to two outcomes, each of which may enhance boundary extension.

First, under the scene context hypothesis it is assumed that a sketch of what scene details lie beyond truncated boundaries is continuously formed during picture viewing. We suggest that when few background details are gathered at initial viewing, the impoverished information leads to an increase in the amount of visual content that is filled-in. Second, given that the background is not well attended to in these circumstances, errors between the extended image and the original image will be far less apparent. The combination of an overactive scene completion mechanism with an inability to error-check the finished product may result in increased amounts of boundary extension.

In Experiment 2, we test this notion by examining whether object salience can influence the magnitude of the boundary extension effect. We hypothesize that very salient foreground objects (such as faces) should give rise to still more extension of image boundaries. This hypothesis is based on our suggestion that the diversion of attentional resources from the background to the foreground results in increased activity of filling-in processes and a lower fidelity representation of the background. We extend this attentional account to suggest that making the background of a scene more complex (and therefore interesting to the observer) will counteract this phenomenon, leading to less boundary extension. That is, when attention is drawn to the background, a higher-fidelity representation of the scene is maintained and there is less of a drive to generate additional content due to increased viewing of the background.

Experiment 2

In this experiment, we adopt the same procedure employed in Experiment 1 to investigate the effect of background and object salience on boundary extension.

Subjects

29 subjects (17 men, 12 women) were recruited from the MIT undergraduate population to participate in this experiment. Subject age ranged from 18-25 years.

Stimuli

36 digital images were taken specifically for this experiment. In each location, a chair was included in the center of the image. For half of the images, an individual was depicted seated in the chair. The remainder of the images depicted the chair alone.

Additionally, the locations in which pictures were taken were chosen to conform to three pre-conceived levels of background complexity. Scenes were placed in these categories according to recent results concerning the perceptual dimensions of scene complexity (Oliva, Mack, Shrestha, & Peeper, 2004). One-third of our images were taken in “simple” settings (low background clutter), one-third in “medium” complexity settings, and the remaining one-third in “complex” settings. Naïve subjects provided complexity ratings for our stimuli. Target images are displayed in Figure 4.

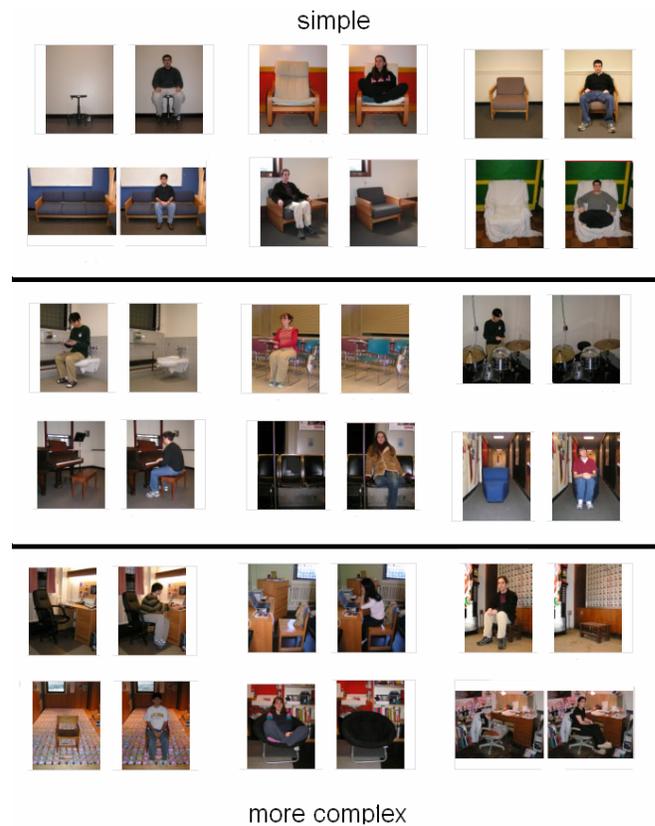


Figure 4: Example stimuli ranging from simple (top row) to complex (bottom row). We also display companion images containing a chair and a seated individual.

Procedure

Subjects were given the same instructions as outlined in Experiment 1. Afterwards, they viewed 18 of the 36 images created for this task in random order at a rate of 10 seconds

per image. The image set was split in half for each subject so that no subject viewed an image of a particular location both with and without a seated individual. This was done to eliminate any interference effects between target images. As before, all targets were presented at a zoom level of 4, 5, or 6. At test, all 8 images were displayed in order of increasingly close-up views.

Results

The difference between each subject's selected test image and the original target's position on our 1-8 scale of zoom levels was calculated for each trial. The average differences for all 6 conditions were calculated for each subject, and a 3x2 repeated -measures ANOVA was carried out. A bar graph of our results is displayed in Figure 5.

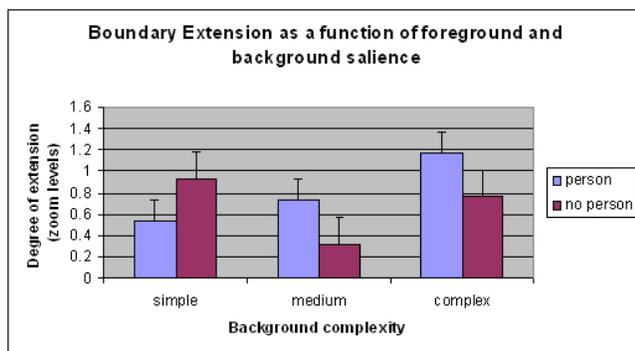


Figure 5: The amount of boundary extension when viewing scenes of varying background complexity, both with and without people in the foreground.

Our ANOVA reveals a main effect of background complexity ($p < 0.05$), and an interaction between the background complexity and the presence of a face. ($p < 0.05$) There was no main effect of the presence of a face in the scene ($p > 0.4$).

We consider these results in relation to the hypotheses formulated after viewing the results of Experiment 1. Had our initial theory been correct, we would have expected the most extension for simple scenes that contained faces. This condition should have maximized the amount of attention allocated to the foreground and minimized the amount of attention allocated to the background. By our initial account this should have enhanced the need to generate background content, while leaving no high-fidelity representation of the background against which to check errors in reconstruction. The confluence of these factors should have led to very large amounts of boundary extension, but instead failed to produce much at all. Indeed, when scenes contained human faces we note a monotonic increase in the amount of extension as background complexity increases, running completely counter to our proposed account of the extension process.

In the absence of faces, we find equally puzzling behavior. A U-shaped function relating complexity to extension is evident here, with medium complexity scenes giving rise to the smallest amounts of extension. Clearly our initial guess concerning the role of object and background complexity is mistaken. The actual interaction of foreground and background content with the boundary extension effect is obviously far more complicated than our naïve intuitions.

Discussion

Contrary to our expectations based on the attentional account we formulated at the end of Experiment 1, we do not find any more extension for images that contain a highly salient object that likely draws attention. We do find, however, an effect of background complexity and a potentially puzzling interaction between the scene background and foreground. These findings suggest that the interaction between boundary extension, attention, and memory for objects and scenes is quite complex.

At present, we can only speculate on what processes are actually at work when observers view these scenes. We present here a discussion of several alternative factors that may be playing an important role in determining the degree to which image boundaries are extended.

First, we explicitly assumed that our manipulations of foreground object identity and background complexity were analogous to manipulating the allocation of attention across the image. While we still believe that human faces and bodies are likely to draw more attention than chairs, it is unclear to what extent we are justified in assuming that more cluttered backgrounds draw more attention and/or viewing. If we are mistaken in our understanding of how attention is influenced by background complexity, our expectations for Experiment 2 may be quite skewed. Future studies could make use of eye-tracking data to ensure that our notions of how subjects distribute saccades across the image surface are correct. Alternatively, change blindness tasks could be employed as a means of measuring how much attention is being directed towards the background and the foreground of images that will be used in an extension task. In both cases, an attentional account of the data would be greatly bolstered by more rigorous data concerning exactly how attention is directed towards the foreground and the background of a scene.

A second possible influence on the extension effect studied in Experiment 2 is that of scene predictability. In our attentional account of the data, we assumed that the only important function of the visual content of the scene was to direct attention towards or away from different image regions. However, given that we are discussing a process of reconstruction (the filling-in of scene content beyond the boundaries) it is also important to consider how visual content might make reconstruction more or less difficult. "Simple" scenes are very boring and should not draw much attention, yet it is also very easy to predict what should be

visible beyond the image boundary. By the same token, one could imagine that “complex” scenes would be very unpredictable. However, we note that many of our complex backgrounds are composed of highly textured elements, like books or tiles. While these backgrounds contain many more edges and more “clutter” than simple backgrounds, they are also predictable in their own way. Could this explain the U-shaped function relating complexity to extension in scenes without people? Perhaps, although we are still at a loss to incorporate these ideas with the results obtained from images containing human faces. In very complex scenes, viewers may be more willing to accept novel objects introduced with a wider view. All of these explanations are at best highly speculative, but at least point to possible future studies to disentangle the relationship between foreground and background image content and boundary extension.

Conclusions

We have demonstrated that scene content can indeed affect the occurrence and magnitude of boundary extension. In particular, we find in Experiment 1 that the presence of objects in the foreground of an image produces larger amounts of extension. Likewise, the complexity (and therefore salience) of the background can also affect the degree to which boundary extension takes place. We suggest a complex relationship between attention, scene complexity, and boundary extension. Future efforts to establish the nature of this relationship could both elucidate the nature of the boundary extension phenomenon and provide insight into scene perception and encoding processes. The interaction between scenes and the objects they are composed of may prove to be exceptionally rich.

Acknowledgments

The authors wish to thank Erin Conwell, Richard Russell, and Yuri Ostrovsky for many helpful comments. We also thank the men and women of MIT’s East Campus for volunteering their time to serve as models for our stimulus set. BJB is supported by National Defense Science and Engineering Graduate Fellowship.

References

Candel, I., Merckelbach, H., Houben, K., & Vandyck, I. (2004). How children remember neutral and emotional pictures: boundary extension in children’s scene memories. *American Journal of Psychology*, *117*(2), 249-257.

Candel, I., Merckelbach, H., & Zandbergen, M. (2003). Boundary distortions for neutral and emotion pictures. *Psychonomic Bulletin and Review*, *10*(3), 691-695.

Gottesman, C. V., & Intraub, H. (2002). Surface construal and the mental representation of scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(3), 589-599.

Gottesman, C. V., & Intraub, H. (2003). Constraints on the spatial extrapolation in the mental representation of scenes: View-boundaries vs. object-boundaries. *Visual Cognition*, *10*(7), 875-893.

Intraub, H. (2004). Anticipatory spatial representation of 3D regions explored by sighted observers and a deaf-and-blind observer. *Cognition*, *94*(1), 19-37.

Intraub, H., & Berkowitz, D. (1996). Beyond the edges of a picture. *American Journal of Psychology*, *109*, 581-598.

Intraub, H., & Bodamer, J. L. (1993). Boundary extension: fundamental aspect of pictorial representation or encoding artifact? *J Exp Psychol Learn Mem Cogn*, *19*(6), 1387-97.

Intraub, H., Gottesman, C. V., Willey, E. V., & Zuk, I. J. (1996). Boundary extension for briefly glimpsed photographs: Do common perceptual processes result in unexpected memory distortions? *Journal of Memory and Language*, *35*, 118-134.

Intraub, H., & Richardson, M. (1989). Wide-angle memories of close-up scenes. *J Exp Psychol Learn Mem Cogn*, *15*(2), 179-87.

Johnson, J. S., & Olshausen, B. A. (2003). Timecourse of Neural Signatures of Object Recognition. *Journal of Vision*, *3*, 499-512.

Matthews, A., & Mackintosh, B. (2004). Take a closer look: emotion modifies the boundary extension effect. *Emotion*, *4*(1), 36-45.

Ohara, T., & Kokubun, O. (2002). Boundary extension as the effects of scene context on picture-memory. *Shinrigaku Kenkyu*, *73*(2), 121-130.

Oliva, A., Mack, M. L., Shrestha, M., & Peeper, A. (2004). *Identifying the Perceptual Dimensions of Visual Complexity of Scenes*. Paper presented at the The 26th Annual Meeting of the Cognitive Science Society, Chicago.

Oliva, A., & Torralba, A. (2001). Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. *International Journal of Computer Vision*, *42*(3), 145-175.

Seamon, J. G., Schlegel, S. E., Hiester, P. M., Landau, S. M., & Blumenthal, B. F. (2002). Misremembering pictures objects: people of all ages demonstrate the boundary extension illusion. *American Journal of Psychology*, *115*(2), 151-167.

Torralba, A., & Oliva, A. (2003). Statistics of Natural Image Categories. *Network: Computation in Neural Systems*, *14*(391-412).

Wan, X., & Simons, D. J. (2004). Examining boundary extension in recognition memory for a large set of digitally edited images. *Journal of Vision*, *4*(8), 872a.