Boundary extension: Insights from signal detection theory

Zili Liu

Department of Psychology,
University of California Los Angeles, CA, USA

Xiaoyang Yang

Department of Psychology,
University of California Los Angeles, CA, USA

Helene Intraub

Department of Psychology,
University of Delaware, Newark, DE, USA

After viewing a scene, people often remember having seen more of the world than was originally visible, an error referred to as boundary extension. Despite the large number of studies on this phenomenon, performance has never been considered in terms of signal detection theory (SDT). We report two visual memory experiments that allowed us to explore boundary extension in terms of SDT. In our experiments, participants first studied pictures presented as close-up or wide-angle views. At test, either the identical view or a different view (a closer or wider angle version of the same scene) were presented and participants rated the test image as being the same or different than before on a 6-point scale. We found that both discrimination sensitivity and bias contributed to the boundary extension effect. The discrimination sensitivity difference was at least 28%, and its presence refuted the hypothesis that boundary extension was due solely to participants' response bias to label test pictures as more wide-angled. Instead, our results support the idea that participants' responses reflect false memory beyond the view (i.e., a more wide-angle view of the world).

Introduction

The boundary extension effect

After viewing a photograph of a natural scene, participants tend to remember having seen more of the world than was shown, as if the boundaries of the view had extended outward in memory (*boundary extension*). Discovered in the context of long-term memory for scenes (Intraub & Richardson, 1989), boundary extension can also occur rapidly enough to be present across a saccadic eye movement (e.g., Intraub & Dickinson,

2008). This constructive memory error is interesting for two reasons. First, participants remember seeing information that had no visual-sensory correlate in the stimulus. Second, although an error with respect to the photograph, boundary extension anticipates the continuation of the scene, predicting upcoming layout in the world (Gottesman, 2011; Intraub, 1997).

One theoretical explanation of boundary extension is provided by the multisource model of scene representation (Intraub, 2010, 2012). The model assumes two stages. In the first stage, visual scene information is perceived and rapidly elicits top-down processing that supports construction of the anticipated continuation of the view. These processes include amodal continuation of surfaces (Fantoni, Hilger, Gerbino, & Kellman, 2008; McDunn, Siddiqui, & Brown, 2014) and amodal completion of any objects that may be occluded by the boundary (Michotte, 1954), as well as expectations and constraints from rapid scene classification (e.g., Greene & Oliva, 2009) and object-to-context associations (Bar, 2004). Thus a multisource scene representation is constructed that reflects the visual information as well as its likely surrounding context. At test, when participants attempt to remember the viewed region alone, misattribution of the mentally constructed continuation of the view to vision causes boundary extension. The most frequently used measure of boundary extension is a 5-point rating scale that requires participants to indicate if a test view is the same, more close-up, or farther away than before (Intraub & Richardson, 1989). Our research question here was whether in terms of signal detection theory (SDT), boundary extension as measured by the rating scale is due to criterion bias, or discrimination sensitivity, or both.

To illustrate the uncertainty as to whether, in terms of SDT, criterion bias, discrimination sensitivity, or both, underlie boundary extension, consider the following

experiment by Park, Intraub, Yi, Widders, and Chun (2007). Pairs of stimulus images were created in which a more close-up view and a more wide-angle view of the same natural scene were created. In the study phase of the experiment, participants were shown one image from each pair such that half of these images were close-up views and half were wide-angle views. In the subsequent test phase, half of the pictures were the identical views as before, and half were replaced with the alternate views. Participants rated whether the test scene was closer, the same, or more wide-angled than the original view on the boundary rating scale (a 5-point scale).

Their results yielded the typical patterns diagnostic of boundary extension. When the same images were presented at study and at test, close-close (C-C) and wide-wide (W-W), participants rated the test pictures as more close up than before (indicating boundary extension in memory). On trials in which study and test views did not match, namely close-wide (C-W) and wide-close (W-C) trials, a critical asymmetry was observed, consistent with prior behavioral studies (Hubbard, Hutchison, & Courtney, 2010). When the first picture shown was a close-up (eliciting boundary extension in memory), and the second picture was a more wide-angle view of the same scene, the magnitude of the perceived change in the views was less than when the order was reversed. That is, although the same pair of pictures was presented, their order of presentation affected how similar the two images appeared to be. The typical interpretation of this asymmetrical pattern is that, in the C-W case, memory for the first picture includes boundary extension, causing it to more closely approximate the test picture, than in the W-C case.

However, this experiment was not designed and analyzed in terms of SDT, so that it remains unknown whether the effect was due to discrimination sensitivity or bias, or both. In contrast to the explanation of boundary extension described earlier, it may be that this error has little to do with constructed scene representations in memory. Instead, it may be that observers tend to rate a remembered photograph as being more wide-angled on the scale, without there being any change to the internal representation whatsoever. In other words, in studies of boundary extension that rely on the closer–farther rating scale, the apparent boundary extension effects could be interpreted as due to bias in SDT terms without involving discrimination sensitivity changes. The aim of the current study was to use SDT to determine if boundary extension can be explained solely in terms of response bias.

Defining discrimination sensitivity

The measure of discrimination sensitivity that is familiar to most people is d', defined as the intercenter distance between two Gaussian distributions of equal standard deviation σ, and normalized by this σ. Naturally, in order for d' to be definable, the two distributions (noise and signal) need to be Gaussians and their variances need to be identical. While it is difficult to verify whether or not the two distributions are indeed Gaussian, it is possible to verify a weaker version of this assumption. Namely, if the two distributions are assumed to be Gaussian, then the receiver operating characteristic (ROC) in the Z coordinate space is a straight line. If the two distributions, in addition, share the same variance, then this straight ROC line in the Z coordinate space has a slope of one (Wickens, 2001).

In order to test both assumptions, namely two Gaussian distributions with equal variance versus two Gaussian distributions with unequal variance, fitting data to a straight line is needed. It should be pointed out, however, that ordinary linear (or nonlinear) regression is unsuitable for this fitting. This is because ordinary curve fitting is based on the assumption that the independent variable is exact or observed without measurement error and all the uncertainty is from the dependent variable. But this assumption is invalid in ROC curve fitting. When fitting the hit rate with the false alarm rate, both variables are subject to measurement errors. Consequently, an ordinary regression gives rise to incorrect and, in fact, systematically biased estimates. For the linear fitting, unless the data are nearly error free, the regression will yield too small a slope and consequently will overestimate σ (Wickens, 2001). For example, when measurement errors in the data are considerable, this bias can lead to the rejection of the equal variance assumption when in fact the assumption is appropriate. Such systematically biased parameter estimation is known as the attenuation bias. In nonlinear models the direction of the bias is more complex.

In statistics, the errors-in-variables model (Griliches & Ringstad, 1970) is used to account for measurement errors in the data. The total least square (Golub & van Loan, 1980) is the technique to take into account the measurement errors in both variables. Recall that in ordinary curve fitting, since the data in the *x*-dimension are assumed exact or measured free of error, the residual error only represents the distance along the *y*-dimension between a datum point and the fitted curve. However, in the total least square method, a residual represents the distance between a datum point and the fitted curve measured along a direction in both *x*- and *y*-dimensions. In fact, if both variables are measured with the same unit, then the residual error is the shortest distance between the datum point and the fitted curve. That is, the residual vector is perpendicular to the tangent of the curve. This is called two-dimensional Euclidean regression (Stein, 1983). In our

case, it is legitimate to assume that the hit and false alarm rates are measured in the same units (in the Z-space or hit and false alarm rate space). This is because the labeling of noise and signal is arbitrary in our case, so the labeling of a hit and a correct rejection is also arbitrary. Given that the false alarm rate = 1 – correct rejection rate, the hit and false alarm rates can be reasonably assumed to share the same units. As a result, the total least square fitting provides an appropriate method for fitting the data.

Fitting a straight line in the Z-space, nevertheless, has to deal with the following problem. Because $Z(0)$ and $Z(1)$ are both undefined, correction is needed when a participant's hit or false alarm rate is 1 or 0. Conventionally, the value $1/(2n)$ is subtracted from 1 or added to 0 in order for the corresponding Z-values to be definable (where $2n$ is usually the total number of trials in the experiment). However, this correction is arbitrary because there is no principled reason as to why this correction factor should be $1/(2n)$, but not $1/n$ or $1/(4n)$ for example.

In order to avoid this problem of arbitrary correction, an alternative method is to fit the ROC in the hit and false alarm rate space. The fitting is nonlinear, but does not suffer from the infinity problem above and hence needs no arbitrary corrections. Another advantage of fitting the ROC in this space is that, even if the equal variance assumption is violated, the area under the ROC still serves as a valid measure of discrimination sensitivity.

To summarize, it is desirable to fit a straight line ROC in Z-space. This is because fitting a straight line is always simpler than a nonlinear curve, and the fitted linear slope serves as a direct test of the equal variance hypothesis. However, curve fitting in Z-space suffers the potential problem of infinity when the hit or false alarm rate is 0 or 1, because the standard correction method is arbitrary. As an alternative, fitting the ROC in the rate space avoids the infinity problem. With this fitting, the area under the ROC is in itself a valid measure of discrimination sensitivity. In the current study, we employed ROC fitting in both spaces in order to look for converging evidence for our hypothesis testing.

Summary of the experiments

To anticipate, we conducted two visual memory experiments. Instead of the typical rating scale in boundary extension research, we used a one-interval, six-scale old–new rating design that was an extended version of a yes–no experiment. The rating method used the following scales: "sure old," …, "guess old," "guess new," …, and "sure new," where "old" and "new" referred to whether the test image was the same

as or different from the study image. This six-scale rating method allowed us to obtain an ROC function per participant.

In Experiment 1, participants first studied pictures of close-up or wide-angle views of natural scenes. To follow the Park et al. (2007) design as closely as possible, at test, either the identical view would be presented (C-C and W-W), or the alternate view (C-W and W-C). Participants rated the test picture as looking the same as before or different. In Experiment 2, we changed the design by blocking the trials such that, in a given block, the study pictures were either all close or all wide. Thus, when a test picture differed from its studied counterpart, the test picture was always a wider view or a closer view (depending on block). This design allowed a cleaner SDT application because, in each block, there were only two distributions that were clearly defined: close-close (C-C) and close-wide (C-W), or wide-wide (W-W) and wide-close (W-C). This design allowed a more straightforward application of SDT. The compromise was that wide-close (W-C) trials were no longer in the same block as close-wide (C-W) trials, which was the design used in Experiment 1. With different designs, both experiments sought converging evidence regarding whether there was any discrimination sensitivity difference between C-C versus C-W and W-W versus W-C conditions.

We found in both experiments that discrimination sensitivity, as measured either in d' or area under ROC, was different between those two conditions. There was also a corresponding bias change. Both the sensitivity difference and bias difference were consistent with and therefore contributed to the boundary extension effect.

Experiment 1: A one-interval old–new rating experiment

In this experiment, we aimed to keep the design as similar as possible to the Park et al. (2007) design, which was introduced earlier. The major difference is that the original response categories of closer, same, and wider were changed to standard old–new rating responses, namely, same or different, each in three levels of certainty. Consequently, the experimental data could be analyzed in standard SDT terms.

Stimuli

The stimuli were 121 pairs of color photographs. These included many of the same single-object scenes as in Park et al. (2007) and others of the same kind. Each pair consisted of a closer and a wider angle view of the same single object on a natural background. The resolution was
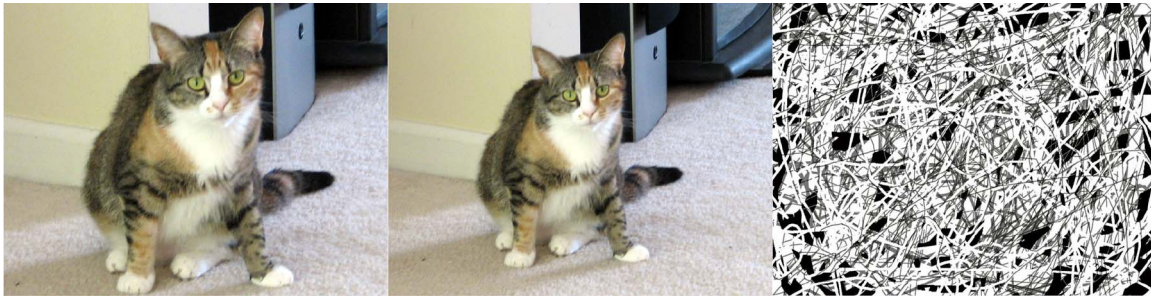
Figure 1. Example of a stimulus pair stimuli, illustrating a close-up scene (left) and a wide-angle scene (middle). The picture on the right is the mask used in the experiments.

640×480 pixels. Figure 1 shows an example of a stimulus pair. Each image was $21° \times 16°$ in visual angle and was presented at the center of the display.

Procedure

For any given participant, 108 pairs from the 121 available pairs were randomly selected as the experimental stimuli. The experiment consisted of three blocks. Each block had a study and a test phase, and used 36 of the 108 pairs of the images. In the study phase, 36 images of different scenes were shown, 18 were wide and 18 were close images. The presentation began with a 1-s green fixation dot at the center of the screen, and then an image was presented for 0.5 s. This image was followed by a 0.5-s image mask, and then by a white fixation dot for 4 s. The cycle continued until all 36 images were shown. The participants were instructed to spread their attention across each image and remember it in as much detail as possible, including the objects, their layout in the scene, and the background. The participants were informed that the background was as important to remember as the foreground object and they were instructed to try to remember the image photographically.

After the first study phase and prior to the test phase, participants were shown an example of a closer and a wider image of the same scene (similar to those in Figure 1) to illustrate how a test image may differ from its study counterpart image. During the test phase, half of the studied images (nine closed and nine wide images) were shown as old images. The wider or closer counterpart images of the other half of the studied images were shown as new images. Each test image was shown with unlimited time, and with a 6-point rating scale underneath. Below the scales were the following words: "Sure Old" (−3), "Guess Old" (−1), "Guess New" (+1), and "Sure New" (+3).

In this design, when a test image was new, it was either wider or closer than its studied image counterpart. This is the standard "Detection of Signals with Different Amplitudes" that was defined in Macmillan

and Creelman (2005, p. 189). Specifically, an old test image is a "no" trial, and a new test image is a "yes" trial. The research question here was whether d' for the wider study condition was smaller than for the closer study condition.

Technically, though, d' may not be definable, because the signal and noise distributions may not be Gaussian, or they may be Gaussian but having different variances. That was why a rating experiment, rather than a binary old–new experiment, was used to obtain a full ROC function. In this way, the Gaussian and equal variance assumptions could be separately verified.

It should be noted that randomization of image assignment was applied whenever possible. For example, for each participant, 108 pairs of images were randomly selected from the 121 pairs available. The assignment of "old" and "new" images was also randomized for each participant. As a result, any possible d' difference between the wider and closer study conditions could not be due to any particular images, so long as the number of participants was reasonably large.

It should also be noted that in the case of the first block of trials, the specific nature of the memory test was not divulged to the participants until the time of the test, whereas in the two blocks that followed, participants had full knowledge of the precise nature of the test that would follow. It took about 45 min for the participants to complete all three blocks.

Apparatus and participants

The display was a 17-in. Dell E773c CRT monitor, with a resolution of 1024×768 pixels, 32-bit color, and 85-Hz refresh rate. The images were rendered using MatLab (The MathWorks, Inc., Natick, MA) and Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). The experiment was conducted in a dim room lit from indirect natural light. The viewing distance was 57 cm.

Twenty-four undergraduate students from UCLA participated for course credits. Their recruitment adhered to the Helsinki Declaration.
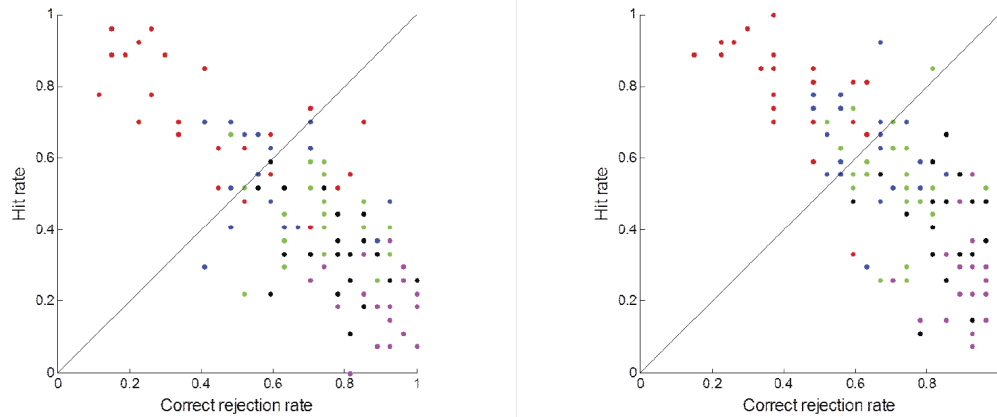
Figure 2. Hit rate and correct rejection rate scatter plots of the six-scale rating data from all 24 participants. A hit here is defined as correctly responding "new," and a correct rejection as correctly responding "old" to a test image. The five groups of data are color coded from red, blue, green, black, and magenta as decision criterion was moved from left to right along the *z*-axis, in the direction from noise to signal. For example, the red dots represent the responses when the criterion was set at the leftmost Z position, between −3 ("surely old") and −2. Left: The C-C condition was noise and C-W was signal. Right: The study-test condition of W-W was noise and W-C was signal.

Results

Figure 2 shows scatter plots of all participants' six-scale ratings, which are color coded as follows. Moving the decision criterion from left to right along the Z-axis, the five data groups are coded in red, blue, green, black, and magenta. Hit and correct-rejection rates are plotted so that the green data (when the criterion was between −1 and +1) directly inform how biased the participants were. In order to understand the results in the context of boundary extension and SDT, a different representation of the data is shown in Figure 3 that may be more intuitive.

We computed discrimination sensitivities (d' and area under ROC) by assuming that the noise and signal distributions were both Gaussians. Without loss of generality, we assumed that the W-W distribution was $N(0, 1)$, and the W-C distribution was $N(\mu, \sigma)$. Similarly, we assumed two distributions for the C-C and C-W. The question was whether the two discrimination sensitivities thus separately obtained, measured in either d' or area under ROC, were the same. It should be noted that whether the W-W and C-C distributions were identical in shape is unknown, but was irrelevant here. This is because in SDT calculations, the noise distribution is always normalized to be $N(0, 1)$.

We first fitted ROC in the Z-space using each participant's rating data (Figure 4). The mean R^2 for the linear fitting was 0.90. With the quadratic term added, 6% additional variance could be accounted for. Given that the linear fitting accounted already for 90% of the variance, we concluded that linearity was acceptable for the 24 participants' data. The average σ

calculated from the linear slope for the C-W distribution was 1.17, which was only marginally different from one, $t(23) = 1.88$, $p = 0.07$). The average σ for the W-C distribution was 1.18, which was also only marginally different from one, $t(23) = 1.83$, $p = 0.08$; all t tests in this paper were two-tailed. Because of the marginal significance, we decided to also compute the area under the ROC in addition to d' to calculate discrimination sensitivities. The areas were 0.63 and 0.68 for close and wide study image conditions, respectively. The difference was statistically significant, $t(23) = 2.52$, $p < 0.02$. If we assume that d' was definable and ignore the
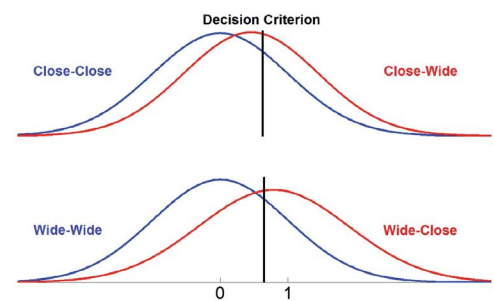

Figure 3. Top: The recovered noise (C-C) and signal (C-W) distributions for close-studied images, and the participants' decision criterion. Bottom: The corresponding distributions and criterion in the case of wide studied images. Along the horizontal z-axis, the two noise distributions are centered at the origin per convention. The two decision criteria were located at approximately the same location. Nevertheless, the signal distribution in the case of close studied images was closer to the noise distribution, resulting both in a smaller d' and a statistically significant bias as compared to the case of wide studied images, where the bias was not significant.
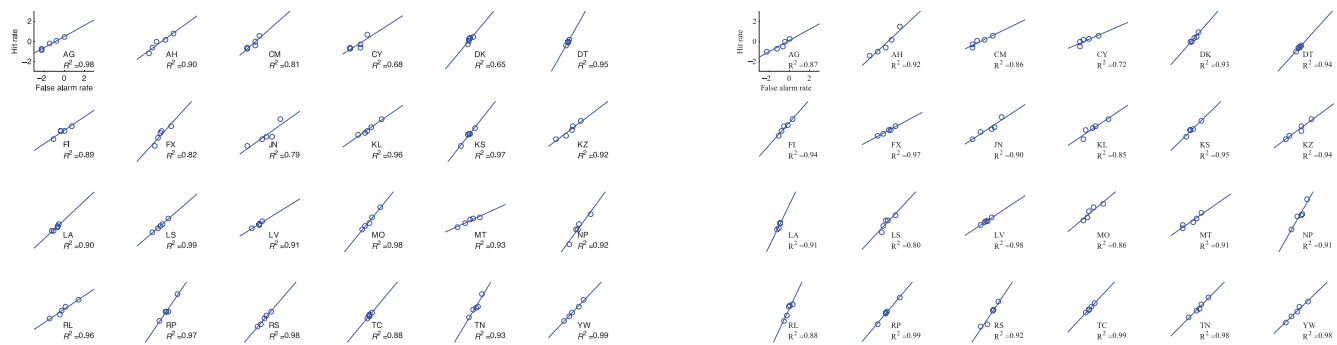
Figure 4. Linear fittings in the Z-space for each of the 24 participants. Left: When close-up pictures were studied. Right: When wide-view pictures were studied.

marginal difference above, the d' values were 0.43 and 0.77, giving rise to a significant difference between them, $t(23) = 2.73$, $p = 0.009$. In other words, the drop of sensitivity d' that was presumably due to boundary extension was 44%. Figure 3 illustrates these two pairs of distributions.

Next we calculate the decision criteria for close and wide study images, respectively. We define the bias-free criterion as the intersection between the signal and noise distributions. In the case of close studied images, these two distributions correspond to C-C and C-W distributions. The bias-free criterion obtained from the hit and false alarm rate space fitting was $Z = 0.21$, and was $Z = 0.28$ from the Z-space linear fitting. The actual criterion coordinate calculated from the participants' false-alarm rates was $Z = 0.63$. (Here, a false alarm was defined by assuming that the decision criterion was in the middle of the six-scale.) There was therefore indeed bias, $t(23) = 3.45$, $p = 0.002$, or $t(23) = 2.91$, $p = 0.008$, in that a wider test image was more likely to be considered as the same as the closer studied image, in agreement with the boundary extension effect.

In the case of wide studied images, the bias-free criterion obtained from the rate-space fitting was $Z = 0.52$, and was $Z = 0.55$ from the Z-space linear fitting. The actual criterion calculated from the false alarm rates was $Z = 0.65$. This bias was not statistically significant, $t(23) = 1.48$, $p = 0.15$, or $t(23) = 1.12$, $p = 0.27$. It is interesting to note that the criterion locations in the two cases were very similar to each other ($Z = 0.63$ and 0.65). This makes sense because the two conditions were randomly mixed, so that it was perhaps impossible to hold two separate criteria. From this single criterion perspective, boundary extension amounted to a relative shift of the signal to the noise distribution in the condition of close study photos.

This result also raised the following question. During each test block, there were four distributions: C-C, C-W, W-W, and W-C. We the experimenters separated these four distributions into two halves (C-C and C-W, W-W and W-C), in order to calculate discrimination sensitivities. This way of separation had an assumption

in it, which needed verification. The next experiment addressed this issue.

Experiment 2: A one-interval old-new rating experiment with close and wide study images in separate blocks

A potential criticism of the last experiment was that the four conditions, C-C, C-W, W-W, and W-C, were all intermingled in a test block. The experimenters, not the participants, separated them into C-C and C-W, and W-W and W-C pairs in order to calculate discrimination sensitivities and biases. In the current experiment, close and wide images were studied in separate blocks. As a result, in a test block, there were only two distributions. They were either C-C and C-W, or W-W and W-C. Consequently, discrimination sensitivity could be calculated in a standard one interval rating design.

Experimental design

The experiment was very similar to the last experiment, except there were four blocks of study and test. The study images in two of the four blocks were all close images. During test, half of the images were identical to those in the study (C-C), and half were the wider counterparts (C-W). The study images in the remaining two blocks were all wide images. During test, half of the images were identical to those in the study (W-W), and the remaining were the closer counterpart images (W-C). The sequence of the first three blocks, which used the same 108 image pairs as in Experiment 1, was randomized from one participant to the next. In the fourth block, 36 additional pairs from author Intraub's laboratory were added. Each of these images

was 332×332 pixels, which was $11° \times 11°$ in visual angle. For half of the participants, this block used close images as study images. For the other half of the participants, this block used wide images as study images.

Participants

One hundred fifty-six new UCLA undergraduate students were similarly recruited as in Experiment 1. They were naive to the purpose of the experiment.

Result

In the interest of space, we report here directly the d' and area under the ROC as discrimination sensitivity measures. As in Experiment 1, the unbiased decision criterion should be in the middle of the six-scale rating, and we calculated the d' for each participant and for close and wide study images, separately. The mean d' (0.76) for close study images was smaller than that for wide study images d' (0.92), $t(164) = 2.56$, $p = 0.011$. The decrease of d' in the C-W condition relative to the W-C condition, presumably due to boundary extension, was 28%. The mean bias for the C-W condition was 0.35, and that for the W-C condition was 0.13, and the difference was highly significant, $t(164) = 7.40$, $p = 6.54E-12$. Figure 5 shows the results.

Next, we calculated the area under the ROC. The mean ROC area for the close study images was 0.65, and that for the wide study images was 0.69. This difference was statistically significant, $t(164) = 3.70$, $p = 0.0003$. This result was qualitatively consistent with the d' analysis above, indicating that the boundary extension effect was in part accompanied by discrimination sensitivity change. The d' analysis also indicated that bias was also partially responsible for the boundary extension effect.

Discussion

In this study, we modified the typical rating procedures used to test boundary extension so that we could conduct an SDT analysis of boundary extension. The key modification was to change from the typical viewing-distance rating to a measure of old or new, so that SDT could be applied. The six-scale rating, as opposed to the binary old–new, was not critical although it provided the means to measure the ROC. In both experiments, we were able to replicate boundary extension effects for photographs of natural scenes. In terms of SDT alone, in Experiment 1, when the test
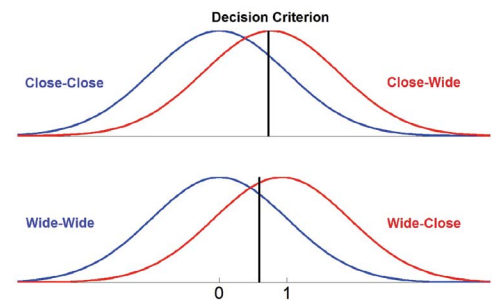

Figure 5. Schematic representation of the noise and signal distributions in the condition when the study images were close views (top, C-C and C-W), and when the study images were wide views (bottom, W-W and W-C). The participants were more biased in the condition on the top than on the bottom (the vertical blue lines). That is to say, after a close image had been studied, when its wider counterpart image was shown in test, participants were more often calling the wider image as same (or old) than different (or new). This phenomenon is the standard boundary extension effect. Note that the two distributions were closer to each other on the top than on the bottom, indicating that there was also sensitivity decrease in boundary extension.

view was either the same or different from the studied view, and participants had to rate the test image as being exactly the same or different (wider or closer), we found that the discrimination sensitivity was smaller when close-up views were studied than when wide-angle views were studied. Accompanied with this reduced discrimination sensitivity, there was also a bias. Both the bias and the reduced sensitivity influenced the boundary extension effects. In comparison, when wider views were studied, the decision criterion was little biased.

Although these results are straightforward in SDT terms, their implications in terms of inferring the nature of scene representations in memory remain an open question. This is because SDT is based on functionally characterizing uncertainties when an observer attempts to behaviorally categorize stimuli, regardless of the mechanisms involved. SDT is therefore agnostic about the biological or psychological origins of the sensitivity and response criterion. Nevertheless, a widespread misconception exists that equates the bias to decision or response bias, suggesting that the bias is necessarily high level in nature. Indeed, even in Green and Swets (1988), it was said that "the main purpose of the application of decision theory is to separate the detectability of the signal, a sensory process, from the decision criterion of the subject, a response or motivational process" (p. 180). However, as Georgeson (2012) pointed out, bias could be also perceptual, or lower level. He used the motion aftereffect as an example, an effect whose perceptual nature is usually unquestioned. The

motion aftereffect is nevertheless characterized psychophysically as a shift of the psychometric function, whose slope does not change. This shift is indistinguishable from response bias. Hence, this example illustrates that bias in SDT is not necessarily equivalent to high-level decision or response bias. Although the meaning of bias remains an open question, the meaning of discrimination sensitivity is less controversial. In the specific context of boundary extension, the sensitivity difference is perhaps directly related to the information content of the memory representation of studied pictures. In other words, inclusion of the additional content beyond the boundaries into the representation of a studied picture, we speculate, makes subsequent matching to the test picture more error prone. In what follows, we elaborate on this speculation, in the context of a scene representation model.

The theoretical explanation of boundary extension described in the Introduction, (multisource model; Intraub, 2010, 2012) assumes two stages. In the first stage, visual scene information is perceived and elicits top-down processes such as amodal perception beyond the boundaries, as well as expectations and constraints from contextual scene classifications. Among the several top-down sources discussed is amodal perception of surfaces beyond the boundaries (Fantoni et al., 2008), as well as amodal completion of any partially visible objects occluded by a boundary (Michotte, 1954). In the context of amodal completion, Lu and Liu (2008, 2009) used an experimental technique similar to the current study to investigate memory representations of objects and scenes, and their results were consistent with the multisource model, as follows. Participants first viewed faces that were partially occluded by small squares. They subsequently viewed either identical images or similar images but the occluding squares were smaller, such that additional face area was revealed. The participants determined that the latter images were more similar to the study images than those test images that were in fact identical to the study images. This difference was reflected in discrimination sensitivity.

Here is another example to illustrate how amodal perception of both types can contribute to the scene representation. When a photo of a natural scene is viewed, the photo necessarily has a boundary, making the scene limited in spatial expanse. The multisource model assumes that when a scene with limited spatial expanse is viewed, it is analogous to viewing the scene through a window with the surrounding scene being occluded. The memory system automatically fills out the expected continuation of the scene beyond the boundary via amodal perception, and available general knowledge about natural scenes and the natural interactions we have with the world (e.g., moving one's head/body to reveal more about what is seen through a

window). For example, grassland should continue with similar texture statistics, and a partially visible object at the boundary should be a complete object. Therefore, if the traditional amodal completion is considered as spatial interpolation, boundary extension may be analogously considered as spatial extrapolation.

In this sense, according to the multisource model, the memorized scene is not a photographic replica; it is a scene representation that is expanded beyond the initially visible, but artificial, boundary. In Stage 2 of the model, the observer makes a decision about how much of the scene had actually been visible, and in so doing, misattributes the highly constrained continuation of the view as having been present in the initial view. We now start from this hypothesis to interpret results in Experiments 1 and 2. From the outset, another assumption is needed. Based on the multisource model, we assume that the representation of the studied, but not the test view, scene has an extended boundary. This is because a test picture is visible during the test, providing unambiguous stimulus information, whereas the studied picture is in memory, so not visible. Our behavioral results bear this out. We observed the asymmetrical response pattern in C-W and W-C conditions that is diagnostic of boundary extension, and supports the notion that boundary extension occurs in memory for the studied view, not in perception of the test picture. Otherwise we would not have obtained an asymmetrical pattern.

In Experiment 1, the false alarm rates for the C-C and W-W conditions were comparable, as shown in the data (0.28 and 0.27), probably due to a single decision criterion. However, in the W-C condition, the perceived difference became larger because the wide studied images became even wider. In comparison, in the C-W condition, the perceived difference became smaller because the close studied images became wider. In effect, this means that the W-C distribution moved away from the W-W distribution, whereas the C-W distribution moved toward the C-C distribution. As a result, discrimination sensitivity was higher for the wide than for the close studied images.

Given that the close and wide study images were randomly interleaved, given that all test conditions were also randomly interleaved, and given that wider and closer viewing angles were relative terms, it is sensible for the participants to hold a single decision criterion location. This location was not statistically different from the bias free location when wide images were studied. When close images were studied, the similar location of the decision criterion was now different from the bias free location. We do not know if these bias results were simply due to the differential movements of the two signal distributions, or whether the positioning of the decision criterion played its own independent role.

Experiment 2 confirmed that even when the four distributions C-C, C-W, W-W, and W-C were not intermixed but separated to allow straightforward application of the standard SDT, the same qualitative results held.

In conclusion, by making only small changes to the traditional methodology used to test boundary extension, we were able to use SDT to examine the functional nature of boundary extension in terms of sensitivity and bias. This analysis allowed us to test an alternative to the multisource model in which boundary extension can be fully accounted for by response bias. Namely, human observers might have a general tendency to label remembered natural photographs as providing a wider viewing angle, without changing the mental representation of the scene whatsoever. Our results rejected this alternative hypothesis, by demonstrating that boundary extension could not be due simply to response bias in a SDT sense. Importantly, the results from both experiments support the same single hypothesis—that the representation of the studied view contains a wider viewing angle than the physical view. Thus, the small changes in our experimental design were crucial because they made it possible to reject an alternative hypothesis about the cause of boundary extension by using SDT.

Keywords: boundary extension, signal detection theory, visual memory, scene recognition, bias

References

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, 5, 617–629, doi:10.1038/nrn1476.

Brainard, D. J. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433–436.

Fantoni, C., Hilger, J. D., Gerbino, W., & Kellman, P. J. (2008). Surface interpolation and 3D relatability. *Journal of Vision*, 8(7):29, 1–19, doi:10.1167/8.7.29. [PubMed] [Article]

Georgeson, M. (2012). Sensory, perceptual and response biases: The criterion concept in perception. *Journal of Vision*, 12(9):1392, doi:10.1167/12.9.1392. [Abstract]

Golub, G., & van Loan, C. (1980). An analysis of the total least squares problem. *Society of Industrial and Applied Mathematics Journal on Numercial Analysis*, 17, 883–893.

Gottesman, C. V. (2011). Mental layout extrapolations prime spatial processing of scenes. *Journal of Experimental Psychology: Human Perception & Performance*, 37, 382–395, doi:10.1037/a0021434.

Green, D. M., & Swets, J. A. (1988). *Signal detection theory and psychophysics*. Huntington, NY: Robert E. Krieger Publishing.

Greene, M. R., & Oliva, A. (2009). Recognition of natural scenes from global properties: seeing the forest without representing the trees. *Cognitive Psychology*, 58(2), 137–179.

Griliches, Z., & Ringstad, V. (1970). Error-in-the-variables bias in nonlinear contexts. *Econometrica*, 38(2), 368–370.

Hubbard, T. L., Hutchison, J. L., & Courtney, J. R. (2010). Boundary extension: Findings and theories. *Quarterly Journal of Experimental Psychology (Hove)*, 63(8), 1467–1494.

Intraub, H. (1997). The representation of visual scenes. *Trends in the Cognitive Sciences*, 1, 217–221, doi:S1364-6613(97)01067-X.

Intraub, H. (2010). Rethinking scene perception: A multisource model. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 52, pp. 231–264). Burlington: Academic Press.

Intraub, H. (2012). Rethinking visual scene perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3(1), 117–127.

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

Lu, H., & Liu, Z. (2008). When a never-seen but less-occluded image is better recognized: Evidence old-new memory experiments. *Journal of Vision*, *8*(7): 31, 1–9, doi:10.1167/8.7.31. [PubMed] [Article]

Lu, H., & Liu, Z. (2009). When a never-seen but less-occluded image is better recognized: Evidence from same-different matching experiments and a model. *Journal of Vision*, *9*(4):4, 1–12, doi:10.1167/9.4.4. [PubMed] [Article]

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide*. Mahwah, NJ: Lawrence Erlbaum Associates.

Michotte, A. (1954). *La perception de la causalite*. Louvain: Publications Universitaires de Louvain.

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

Park, S., Intraub, H., Yi, D. J., Widders, D., & Chun, M. M. (2007). Beyond the edges of a view: Boundary extension in human scene-selective visual cortex. *Neuron*, *54*(2), 335–342.

Pelli, D. G. (1997). The Videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

Stein, Y. (1983). *Two dimensional Euclidean regression*. Paper presented at the Conference on Computer Mapping, Herzelia, Israel.

Wickens, T. (2001). *Elementary signal detection theory*. New York: Oxford University Press.