The Spatial Representation of Natural Scenes

Helene Intraub
Department of Psychology
University of Delaware
Newark, DE 19716

Abstract

A scene is never perceived in its entirety. The input for scene comprehension is always a partial view. Whether looking at a picture or scanning the visual environment, the observer is faced with partial views. An Extension-Normalization Model is offered to account for the ability to perceive scenes based on partial information. The model, proposes an extrapolation process that occurs during perception and then influences spatial memory -- yielding a memory distortion referred to as "boundary extension" (Intraub & Richardson, 1989). Research testing the implications of the model for spatial memory, imagination, and the integration of eye movements during visual scanning will be discussed.

"Boundary extension" refers to the viewer's tendency to remember seeing information that was not in a photograph of a scene, but that was likely to have existed just outside the camera's field of view (Intraub & Richardson, 1989). This distortion of the remembered picture-space is evident in subjects' drawings of previously presented scenes. They tend to depict a more wide-angle view of the scene than was actually shown in the photograph, reducing the size of main objects and extending the background accordingly (Intraub & Richardson, 1989; Intraub, 1992; Intraub & Bodamer, 1993; Legault & Standing, 1992).

Boundary extension is also evident in tests of recognition memory. For example, when faced with the same picture again, subjects tend to rate it as depicting more of a close-up view than before, thus indicating that they remember the original picture as having shown a more wide-angle view (Intraub & Richardson, 1989; Intraub, Bender, & Mangels, 1992; Intraub, 1992; Intraub & Bodamer, 1993). The same distortion was also observed when subjects were asked to physically move boundaries on a test picture to indicate their remembered location (Nyström, 1993). All three types of tests demonstrate that viewers remember having seen a wider expanse of the scene than had actually been depicted in the photograph. The effect is so robust, that even when subjects were informed of the phenomenon in advance and tried to prevent it, they were unsuccessful in doing so (Intraub & Bodamer, 1993).

The research described in this report is part of an ongoing project to develop a model of the spatial representation of scenes, referred to an the Extension-Normalization Model (Intraub et al., in press). The model has implications for picture memory, spatial imagination, and the integration of eye fixations during visual scanning. It is characterized by two types of schematic structures. The first, called the "perceptual schema", is thought to be a mental representation of the expected layout of a partially specified scene. It is thought to be similar to the "mental schema," proposed by Hochberg (1978, 1986) to account for the integration of successive views during visual perception. The perceptual schema, which extrapolates expected scene structure, is activated when the viewer encounters a partial view. The expected scene structure outside the picture's boundaries is so fundamental to comprehension that it becomes incorporated in the observer's mental representation of the picture, thus causing boundary extension in memory.

The strength of the effect has been shown to depend, in part, on picture view: with close-ups yielding the greatest degree of boundary extension and wide-angle pictures the smallest, or none at all (Intraub et al., 1992). A possible explanation for the effect of picture view is that the most predictive part of the schema is that area immediately surrounding an attended object. In a close-up, there is relatively little uncertainty about what is likely to exist just outside the picture's boundaries. In more wide-angle views of the object, more of the predictable area is represented within the picture itself. In a sense the most informative part of the schema is redundant with the information contained in the wide-angle picture. For example, consider a tight close-up of a pen on a desk. There is very little uncertainty that more desk-top exists just outside the picture's boundaries. The more wide-angle the view, the less certain that the desk-top won't end just outside the boundaries. This is true whether the background is a desk-top or a field of grass. As a result, close-ups yield a greater sense of expectancy and a greater degree of boundary extension, than do more wide-angle views of the same scene.

The second schematic structure involves the formation of an "average" representation that is derived from the picture set. Over time, the remembered spatial layout of the scene changes toward this average view -- a process referred to as "normalization". The representations are remembered as looking more like the average view (i.e., regression toward the mean).

Normalization in memory is a well-established occurrence (e.g., Bartlett, 1932; Gibson, 1969). In fact, subjects will remember having seen the average item in a set, even when it was never actually presented (Franks & Bransford, 1971). In terms of memory for picture boundaries, depending on a particular picture's relation to the average view, its boundaries will either extend or restrict as they normalize. This means that relatively close-up views in the set will yield boundary extension, but relatively wide-angle views will yield boundary restriction. The strength of normalization is determined by the heterogeneity of the set. (A truly homogeneous set would yield no normalization.)

The perceptual schema exerts greater influence on the representation immediately, and normalization increases in influence over time. This can account for the counterintuitive finding that boundary memory is more accurate after a 2-day delay, than minutes after presentation (Intraub, et al., 1992). What appears to be happening is that within minutes, boundary extension occurs for close-up and more wide-angle pictures, but after a delay that is sufficient for normalization to occur, close-ups tend to yield boundary extension, whereas wide-angles tend to yield boundary restriction.
The normalization process, therefore, served to temper the original boundary extension effect, thus yielding what appeared to be more accurate mental representations of the pictures.

If the proposed perceptual schema is indeed a fundamental knowledge representation that underlies scene comprehension, it will have implications for picture memory, imagination, and perhaps the integration of eye fixations. Research into the possible use of the schema in each of these three areas will be summarized in the following sections.

Implications for Memory of Objects and Scenes

An important caveat of the model is that the perceptual schema will only be activated by partial scene information. In contrast, normalization should occur for all types of visual-spatial stimuli. For pictures of scenes, this results initially in boundary extension, which over time is tempered by normalization. For stimulus displays that do not depict a scene, only a normalization pattern would be expected. Put more directly, the same object, that appears in the same location within the picture-space, should be remembered as having been a different size (with respect to the picture space) depending on whether or not it was presented within the context of a scene. For example, a picture of a basketball on a gym floor initially would be expected to yield boundary extension, whereas a picture of the same basketball on a blank white background (i.e., a picture with no scene structure), would not. Both types of stimuli, however, would be subject to normalization over time.

In a recent test of this prediction, we presented 122 undergraduates at the University of Delaware with one of three types of stimuli: photograph-scenes, outline-scenes or outline-objects (Gottesman & Intraub, 1993). All three stimulus types were based on the same 16 scenes in which a single object was presented on a simple real world background and was photographed in both a close-up version and a relatively wide-angle version. These original photographs served as the photograph-scenes. In the outline-scenes, the main objects and the background shown in the photograph were traced in black ink on white paper. In the outline-objects, only the main objects were traced and the background was left blank. Therefore for each stimulus type the main object was the same and was in the same location in the picture-space. The question was whether memory for objects in scene contexts would result in overall boundary extension, whereas memory for the same objects without scene context would not. Memory for both photograph-scenes and outline-scenes was tested. Up until this experiment, boundary extension research had only been conducted using photographs. Therefore, in addition to the major goal of the experiment, this comparison allowed us to determine if simple outline representations of the scenes would yield the same phenomenon.

The subjects were randomly assigned to one of the 3 stimulus conditions. Each group of subjects was presented with 16 pictures for 15 seconds each. Half the pictures were presented in the close-up version, and half in the wide-angle version, counterbalanced across subjects. They were told to remember the size and location of everything in the picture space.

Following this, the same 16 pictures were presented. Half the pictures were presented in the same version as before: Close-up tested by a close-up (CC) and wide-angle tested by a wide-angle (WW). The other half were presented in the alternate version: Close-up tested by a wide-angle (CW) and wide-angle tested by a close-up (WC). Subjects rated each picture on a 5-point scale, as "same" (0), "slightly closer-up then before, object slightly bigger" (-1), "much closer-up, object much bigger" (-2), "slightly more wide-angle, object slightly smaller" (1) or "much more wide-angle, object much smaller" (2). Note that boundary extension would result in the object being remembered as taking up a smaller proportion of the picture-space than had actually been the case, and boundary restriction would result in the object being remembered as taking up a larger proportion.

The pattern of results supported the extension-normalization model. For pictures of objects in scenes, overall boundary extension was obtained (close-ups yielded a large significant degree of boundary extension and wide-angles yielded no directional distortion). In contrast, pictures of objects without scene structure showed no overall directional distortion (close-ups and wide-angle pictures yielded boundary extension and boundary restriction of equivalent magnitude). This latter pattern is a typical normalization pattern. For those stimuli that would not be expected to activate the perceptual schema, (the "extension" component of the model), all that was left to influence the memory was normalization.

Table 1 shows the mean boundary scores (based on the 5-point scale described previously) for pictures that were presented in the same version during the presentation and the test, that is the "CC" and "WW" pictures. Negative scores indicate boundary extension and positive scores indicate boundary restriction. Confidence intervals were constructed around all the means, Those that are significantly different than zero ("same") are indicated in Table 1.

Table 1. Mean boundary rating for "CC" and "WW" pictures (range -2 to +2).

Picture	Picture Version	
Type	Close-up	Wide-angle
Photograph	-.26**	.00
Outline-scene	-.38**	-.04
Outline-object	-.13*	.13*

** p<.001
* p<.01

Orthogonal planned comparisons were conducted, comparing: a) the photograph-scene condition with the outline-scene condition and, b) the two scene conditions (photograph-scene and outline-scene) to the non-scene condition (outline-object). This analysis for the "CC" and "WW" pictures showed that there was no difference between mean boundary scores for photographs-scenes and outline-scenes, $F(1,119) = 1.53$, but that overall, boundary extension was greater for pictures with scene structure than for those without, $F(1,119) = 12.14$, $p<.001$. In fact, those without scene structure showed no overall extension. As in previous research, overall, there was more boundary extension for close-up pictures than for wide-angle pictures, $F(1,119)=31.24$, $p<.001$. There were no significant interactions, $F<1$.

The responses to the distractor pictures also support the Extension-Normalization Model. One of the diagnostic patterns indicating boundary extension is an asymmetry in responses to the distractors. Given the same pair of pictures, if the close-up is the target and the wide-angle version is the distractor, boundary extension of the target often causes the subjects to accept the wide-angle distractor as the picture that was shown before. On the other hand, if the wide-angle picture is shown first, and some extension occurs, the close-up distractor will seem to be even more of a close-up. Table 2 shows the mean boundary scores for pictures that were presented in different versions during presentation and test, i.e. the "CW" and "WC" pictures. As may be seen in the table,

Table 2. Mean boundary rating for "CW" and "WC" pictures (range -2 to +2).

Picture Type	Picture Version	
	CW	WC
Photograph	1.21	-1.45
Outline-scene	.88	-1.24
Outline-object	1.42	-1.43

this asymmetry was obtained for the scenes ($p<.01$). It is important to note that in contrast, there was no asymmetry in the responses to distractors in the outline-object condition.

Implications for Imagination of Scenes

Both behavioral and neuropsychological evidence have been offered to support the position that visual imagery and visual perception share common systems (e.g., Farah, 1988; Finke, 1985; and Shepard, 1984). If the proposed perceptual schema is a knowledge representation that underlies scene comprehension, the same mental structure may underlie imagination of scenes. Imagination requires knowledge about the expected layout of the to-be-imagined scene. Thus far, we had attempted to activate the perceptual schema by manipulating the "bottom-up" information in the display (i.e., scene structure vs. no scene structure). In the following experiment, conducted in collaboration with Amy Bills, we attempt to activate the perceptual schema by instructing the subject to imagine scene structure.

The stimuli were the outline-objects, that had been created by tracing objects from close-up and wide-angle views of the same scenes (described in the previous section). Subjects in the control condition were asked to remember each object's size and position in the picture space (as in the previous experiment). In addition to this instruction, subjects in the imagine-scene condition, were given a description of the photograph-scene that the object had been traced from and were asked to mentally "superimpose" the described scene onto each outline drawing. If imagination activates the perceptual schema, then boundary extension should occur in the imagination condition, but not in the control condition. In the control condition there is no partial scene presented or imagined that should activate the schema and cause boundary extension in memory.

As in the research, with Gottesman, in the control condition, there was no overall boundary extension. However, in the imagination condition, recognition memory for the same pictures of objects yielded boundary extension. Both the close-up and wide-angle pictures were remembered with significant degrees of boundary extension, with close-ups yielding a larger degree of extension than did the wide-angle views. This pattern is the typical pattern of boundary memory errors that is obtained when observers actually view pictures of scenes.

The results show that although both groups of subjects saw exactly the same objects, memory for the object's size depended on whether or not the subject had imagined that object in a scene context. Furthermore, in the imagination condition, the typical response asymmetry associated with boundary extension was obtained in the distractor conditions (CW & WC). No such asymmetry was obtained in the control condition. These results support the hypothesis that the perceptual schema is a representation of scene structure that can be activated either by perception or imagination of partial views.

Implications for Integration of Eye Fixations

The Extension-Normalization Model is not limited to the perception and memory of pictures. It has implications for other conditions in which the input to the system is a partial view. The world we perceive is made up of complex interrelated collections of objects and backgrounds. It is important to remember that at no moment in time does an observer perceive a scene in its entirety. As discussed earlier, a picture presents a partial view; one that is bounded by the picture's edges. Similarly, during visual scanning, each fixation delivers a partial view of a continuous visual environment to the observer. If the proposed perceptual schema is a fundamental representation of expected scene structure, then it may possibly play a role in the integration of the partial views that are conveyed to visual/cognitive system with each new eye fixation.

Recent advances in the study of eye movements, suggest that our ability to perceive a continuous visual world in spite of the fact that the input to the visual/cognitive system is a series of discrete eye fixations (each providing us with a partial view of a scene), suggest that an abstract spatial layout may underlie integration (e.g., Irwin, Brown and Sun, 1988; O'Regan, 1992; Rayner & Pollatsek, 1992). If the perceptual schema is to be considered as a possible candidate for the spatial structure underlying integration, then we would expect to see its effect, not only on pictures presented for multisecond durations, but for pictures presented briefly enough to simulate single eye fixations. Because all previous research used multisecond presentations and retention intervals of at least several minutes, the early time-course of the phenomenon was not known. It was possible that contrary to the perceptual schema hypothesis, boundary extension may take time to build up in memory, and that it may not be immediately apparent following perception.

To test these possibilities, Intraub et al. (in press), tested boundary memory for pictures that had been shown for only 250 msec each interspersed with multisecond interstimulus intervals, and for successive presentations of briefly presented pictures (simulating the rapid changes associated with visual scanning). In both cases, boundary extension was obtained.

Most relevant to the issue of integration across fixations is the latter experiment. In that case, they presented triads of digitized color photographs to subjects. In each triad, pictures were presented at a rate of 3/sec, followed by a 1 sec masked interval and a repetition of one of the pictures. The repeated picture served as the test stimulus and remained on the screen while the subjects rated it using the same 5-point boundary scale described earlier. Subjects tended to rate the same pictures as being "more close-up than before" indicating that they remembered them as having shown more of the scene. Regardless of the serial position of the to-be-tested picture in the triad, significant degrees of boundary extension were obtained, with scores associated with each of the three positions ranging from -.37 to -.41.

The fact that boundary extension was obtained within 1 second of glimpsing a picture indicates that the schema is very rapidly activated and incorporated into memory.

Boundary extension clearly did not require large memory sets or long intervals to become evident.This supports the possibility that the perceptual schema is a knowledge structure that underlies, not only picture perception and imagination, but perhaps the integration of successive eye fixations.

Conclusions

Boundary extension is a very robust and common representational error that occurs when observers try to remember photographs of scenes. It raises interesting questions about the nature of the spatial representation of scenes. Three lines of research that test a developing model, called the Extension-Normalization Model, were briefly summarized. Object/Scene comparisons indicated that boundary extension occurs for displays that clearly depict scene structure (indicating a partial view), but does not occur for displays in which an outline object is presented on a blank background. When the perceptual schema was not activated, as predicted by the model, the only distortion that was observed was dictated by the remaining component -- normalization. However, these same nonscene pictures did yield boundary extension when subjects were instructed to imagine a scene during encoding. Apparently, imagination provided a top-down route for activating the perceptual schema. Finally, boundary extension was found to occur as soon as 1 sec following the offset of a briefly presented photograph. This is consistent with the possibility that boundary extension reflects processes involved in scene perception, and that it is reasonable to entertain the notion that the perceptual schema may underlie the integration of eye fixations.

Taken together, these results suggest that the visual/cognitive system is designed to understand partial views, regardless of whether these views are delivered by a single fixation, a picture, or an image of a scene. They suggest the hypothesis that the system interprets and remembers partial views by extrapolating beyond the boundaries of what currently can be seen. When remembering, the viewer tends to blur the distinction between what was actually seen, and what was understood during that extrapolation. Therefore, expectations of the structure that is likely to exist just beyond the boundaries is incorporated in the pictorial representation, causing viewers to experience boundary extension. This suggests that although scene perception is enhanced by a system that tries to integrate partial views by ignoring actual boundaries, the cost is in the accuracy of spatial memory for the specific partial view that was seen.

References

Bartlett, F. C. (1932). Remembering: A Study in Experimental and Social Psychology. London & New York: Cambridge University Press.

Gibson, E. (1969). Principles of Perceptual learning and Development. New York: Appleton-Century-Crofts.

Farah, M.J. (1988). Is visual imagery really visual? Overlooked evidence from neuropsychology. Psychological Review, 95, 307-317.

Finke, R.A. (1985). Theories relating mental imagery to perception. Psychological Bulletin, 98, 236-259.

Gottesman, C.V. & Intraub, H. Representational differences between pictures of objects and pictures of scenes. Thirty-fourth Annual Meeting of the Psychonomics Society, Washington D.C., November, 1993.

Hochberg, J. (1978). Perception (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Hochberg, J. (1986). Representation of motion and space in video and cinematic displays. In Boff, K. J., Kaufman, L., & Thomas, J. P. (Eds.), Handbook of perception and human performance (vol. 1, pp. 22:1-22:64). New York: John Wiley & Sons.

Intraub, H. (1992). Contextual factors in scene perception. In Chekaluk, E. & Llewellyn, K. R. (Eds.), The Role of Eye Movements in Perceptual Processes. Elsevier Science Publishers B.V.: Holland. (Pgs. 45-72).

Intraub, H., Bender, R. S., & Mangels, J. A. (1992). Looking at pictures but remembering scenes. Journal of Experimental Psychology: Learning, Memory and Cognition, 18, 180-191.

Intraub, H., & Berkowits, D. (In press). Beyond the edges of a picture. American Journal of Psychology.

Intraub H., & Bodamer, J. A. (1993). Boundary Extension: Fundamental Aspect of Pictorial Representation or Encoding Artifact? Journal of Experimental Psychology: Learning, Memory & Cognition, 19, 1387-1397.

Intraub, H., & Richardson, M. (1989). Journal of Experimental Psychology: Learning, Memory and Cognition, 15, 179-187.

Intraub, H., Gottesman, C. V., Willey, E. V., & Zuk, I. J. (In press). Boundary extension for briefly glimpsed pictures. Journal of Memory and Language.

Irwin, D. E., Brown, J. S., & Sun, J. (1988). Visual masking and visual integration across saccadic eye movements. Journal of Experimental Psychology General, 117, 276-287.

Legault, & Standing, L. (1992). Memory for size of drawings and of photographs. Perceptual and Motor Skills, 75, 121.

Nystrom, M. (1993). Is picture memory wide-angle? Psychological Research Bulletin, 33, 1-16.

O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. Canadian Journal of Psychology, 46, 461-488.

Rayner, K. and Pollatstek, A. (1992). Eye movements and scene perception. Canadian Journal of Psychology, 46, 342-376.