Mental schemata and the limits of perception

James T. Enns & Erin Austen

University of British Columbia

Peterson, M. A., Gillam, B., Sedgwick, H. A. (expected publication date: 2006). *In the Mind's Eye: Julian Hochberg on the Perception of Pictures, Film, and the World.* NY: Oxford University Press.

Address correspondence to:

J. T. Enns

Department of Psychology

University of British Columbia

2136 West Mall

Vancouver, B.C., Canada V6T 1Z4

email: jenns@psych.ubc.ca

phone: 604-822-6634

fax: 604-822-6923

Julian Hochberg deserves full credit for giving serious consideration to one of the most obvious and yet overlooked limitations of the human visual system. This is the fact that vision is not uniformly detailed over the field of view. Or, as he liked to put it, "...unlike objects themselves, our perception of objects is not everywhere dense..."(Hochberg, 1978). He was referring to the fact that retinal cones are concentrated near the fovea, so as to maximize spatial resolution for only the central 1-2 degrees around the point of fixation. This limitation means that in order to view any scene in detail, a series of eye movements and fixations must be made. It also implies that our subjective experience of a wide-angled field of view that is high in detail is an illusion, as many textbook demonstrations now attest. Yet, despite this knowledge, formal theories of vision have for the most part paid very little attention to it until rather late into the twentieth century. Even David Marr's (1982) Vision, the book often credited with the greatest influence on vision science in the past quarter century, is remarkably silent on the consequences of this fundamental limitation.

Vision researchers have no one to blame for this oversight but themselves. Almost 40 years ago, Hochberg summarized an extensive research program in which this fundamental limitation of vision played a starring role. First presented in 1966 as an invited address at the meetings of the American Psychological Society, and later published as a book chapter titled "In the mind's eye" (1968), it laid out some of the consequences of this uneven density of vision.

First, Hochberg pointed out that any analysis of shape and scene perception must be conducted on the mental representation or image in the mind's eye (he called it a schematic map or a mental schema) rather than on the stimulus itself. This is for the simple reason that every shape or scene enters the mind of the observer through a series of piecemeal views. If there is a picture of an extended scene, it resides solely in the observer's mind. This insight alone rules out all Gestalt analyses of the stimulus, and other coding schemes for that matter, in which the analysis begins with a physical description of the extended stimulus. The proper units of analysis must be sought in the visual brain and mind.

Second, Hochberg proposed that vision be subdivided functionally into two qualitative steps: one involving perception in a momentary glance (the processes at work during a single fixation) and the other involving perception of the schematic map (the integration of multiple glances). It was this second step that he proposed influenced perception so as to make it unique in each individual and in each moment of time, through the forces of attention, learning, and development. Simply put, if the scene is formed only in the mind, then the mind can have maximum influence at the level of the scene rather than at the level of the glance.

Third, he introduced the idea that perception in the momentary glance works recursively with the perception of a schema in a mutual relationship. For example, the processes of perception in a glance might automatically evoke hypotheses (learned or innately acquired) about what lies outside the scope of spatial resolution on that glance, leading to a bias in where the next glance will be placed and what might be expected to be found there. In turn, having an expectation of a particular schema in a scene would lead to the testing of specific hypotheses in subsequent glances. This recursive aspect to Hochberg's theory made it very clear that vision in almost every circumstance is influenced as much by what lies in the mind as what lies in the eyes of the beholder.

As has been repeated many times in history, the mere fact that these ideas were laid out in a compelling way did not in itself ensure that the larger community of vision science would embrace them. In this paper I will summarize some recent research we have conducted that is still aimed at establishing Hochberg's points that "… our perception of objects is not everywhere dense," and that expectations have a direct influence on what we see in a momentary glance.

The work was conducted in the context of what has come to be called change-blindness (Rensink, 2002; Simons & Levin, 1997; see also contributions in this volume by Hayoe and by Simons & Levine). In brief, interrupting an observer's view of a scene by an eye blink, a brief flicker in the image, a brief occlusion by a passing object, or a change in viewing position, can render the observer insensitive to changes in the location and identity of many objects that are not at the current focus of attention. Note that these effects should not come as any surprise to those following Hochberg's work, since he pointed out that precisely this sort of 'blindness' occurs when observers look at one of the Dutch artist, M.C. Escher's, impossible buildings. Hochberg's experiments showed that observers were slow and inaccurate in detecting inconsistencies in the details seen in separate glances of these drawings, provided that the information seen within each glance was internally consistent and structurally coherent. According to Hochberg, what it takes to detect the inconsistency in the separate glances is the formation of a very specific schema regarding the three-dimensional structure of a portion of the scene, presumably acquired from the first of the two glances. This schema must then be compared with the information acquired in a second glance at another region of the scene. The problem is that such specific schemata are not the default ones we use in exploring scenes of buildings and landscapes. Rather, we simply assume that the physical objects we see will be globally coherent, even in pictures.

One of the key ingredients for effective modern demonstrations of 'change blindness' is that the changes made to the image (by the experimenter) do not disturb the observers' general schema of the scene. Just as observers of Escher's drawings are likely to become aware of the inconsistencies once they test a perceptual schema against the details of a momentary glance, so too the successful detection of change in a scene requires observers to focus "attention" in one way or another on the location in which the change is occurring. On this all change blindness researchers seem to agree (Rensink, 2002; Simons & Levin, 1997). But within this broad agreement on the importance of attention, there is a considerable range of opinion on what is meant by focusing "attention." The same term has been applied to objects that are of central interest to the observer (Rensink, 2000; Rensink et al, 1997), to scene locations in which a local visual transient or a unique color has recently been presented (Rensink et al, 1997; Scholl, 2000), and to increased spatial certainty about where changed objects will appear (Rensink, 1999; Smilek, Eastwood & Merikle, 2000).

Our first question was whether the perception of an object is really as rich and detailed as is assumed by many modern attention researchers (Duncan, 1984; Egly, Driver & Rafal, 1994, Wolfe, 1994; Wolfe, Klempen & Dahlen, 2000). In this research, there is an oft-repeated claim that paying attention to an object "binds" all the features of that object, at least momentarily, thereby allowing the observer to have equal access to the entire bundle of visual features associated with an object (e.g., its relative location in the scene, its shape, color, texture, motion trajectory, etc). The strongest evidence for this claim comes from what has been called the two-object cost: reporting two visual attributes of a scene can be accomplished more rapidly and accurately when the two attributes are linked to the same attended object than when they are associated with two or more objects over which attention must be divided or rapidly switched (Baylis, 1994; Davis et al., 2000; Duncan, 1984).

In a first study (Austen & Enns, 2000) we began by differentiating the spatial distribution of attention (Is attention aimed narrowly or distributed widely over space?) from the expectations of the observer (Is an expected object in the fine details or in a more extended pattern?). We used the flicker method of change detection (Rensink, 2002) in which the observer's task is to view two alternating frames of a display and to indicate as soon as possible whether any item is changed in its identity from frame to frame.

The stimuli were compound letters, as shown in Figure 1a: at the local level they were multiple copies of the same small letter; at the global level these smaller letters formed the shape of a larger letter. Observers were informed that from trial to trial the number of compound letters would vary randomly between 1, 3, or 5. They were also told that the changed letter would be found either at the local level (the small letters that form the large letter) or at the global level (the large letter formed by the spatial configuration of the smaller letters). The cycling displays consisted of a view of one display for 200 ms, separated by a blank interval of 200 ms, and then a view of the alternate display, followed by the same blank interval. The cycling display ended when the observer pressed a key to indicate whether the two displays contained 'no change' or a 'change.'

----------- Insert Figure 1 about here -------------

In experiment 1 (Austen & Enns, 2000) the changing letter was equally often at the local and global levels. The results showed first that when attention was spatially distributed among three or five compound letters, rather than being focused on only one letter, changes were detected more readily at the global level. This is consistent with a general bias in perception favoring the global or 'gist' level of structure when attention is distributed among multiple objects (Navon, 1977). Faced with the task of dividing attention among multiple objects it appears that perception first provides access to the crude outline or gist of a scene before narrowing its scope to specific details of a single object. From Hochberg's perspective, one might say that this is consistent with perception having access to a mental schema prior to having access to the details in a momentary glance (Biederman, Rabinowitz, Glass, & Stacey, 1974; Hochstein & Ahissar, 2002).

A second finding of experiment 1 was that when there was only a single compound letter in a display, changes in letter identity were detected equally well at the local and global levels of structure. This finding, at first blush, appears to concur with the popular modern view that a single attended object is richly represented in the mind of the observer, following only a single brief glance (Duncan, 1984).

However, the results of experiment 2 (Austen & Enns, 2000) held a surprise for this interpretation. In this experiment, the perceptual expectations of the observer were systematically varied without influencing their response tendency. In a global bias condition, 75% of the trials involved a change to one of the global letters while the remaining 25% of those trials involved letter changes at the local level. A local bias condition involved the complementary arrangement: 25%

changes were to global letters and 75% were to local letters. Under these conditions, observers were both faster and more accurate in detecting changes at the level that was most likely to change. Most important, this was true even when there was only a single item to attend in order to detect the change. It must be emphasized that this expectation bias on perception was not associated with any response bias: change versus no change trials were still equally likely, just as in experiment 1. This finding therefore strongly suggests that the number of objects that are attended is not the only limiting factor on perception; the nature of the mental schema that the observer tests when they first view the attended object is also important (Archambault O'Donnell C., & Schyns, 1999; Lamb, Pond, & Zahir, 2000). Even a single, visually attended, compound letter is not represented in the mind such that its local and its global structure are equally accessible for report.

But how general is this conclusion? Should we draw the conclusion from these results that attended objects are generally not represented as richly and in as much detail as predicted by research on object-based attention? After all, compound letters could be criticized for having little ecological or social significance as 'objects.' If anything, they are among the most arbitrary and artificial stimuli that could possibly be tested. Objects in the natural world are noted for their redundancy, not for their independently variable attributes at different levels of analysis. Moreover, the constituents of these stimuli, letters of the alphabet, are extremely over-learned. Perhaps attended objects are richly represented when a more natural and less familiar set of stimuli are tested.

A second study aimed at these questions (Austen & Enns, 2003) tested the perception of human faces with an experimental design very similar to the study using compound letters (Austen & Enns, 2000). Faces were chosen for several reasons. First, they are stimuli with unique social and biological significance for humans, who are experts at making the subtle visual discriminations required to identify them. Observers' expertise with faces should ensure that they are processed efficiently even when the particular faces seen are not well known. Second, face perception depends on "configural processing," meaning that the perception of the individual constituents and features of a face (e.g., particular nose shape, emotional expression) are highly dependent on the presence of other constituents and features (e.g., mouth shape, three-dimensional structure of the face). This strongly biases the test in favor of finding evidence of a "rich representation" of an attended face. If face perception turns out to be every bit as expectation-dependent as the perception of compound letters, we would be forced to conclude that their perception in a glance is also not richly detailed. Even the details of a single glance may not be "everywhere dense" (Peterson & Gibson, 1991; Peterson & Hochberg, 1983).

The stimulus set we used is shown in Figure 1B. It consisted of four faces derived from the combinations of two individuals (person 1 and 2) posing with each of two emotional expressions (happy and sad). We again conducted two experiments, one in which change in the faces was equally likely to occur in the identity or in the emotional expression of the face (Experiment 1), and one in which the change was more likely to occur in either the identity or the emotional expression of the face (Experiment 2).

We also repeated all of our tests for both upright-faces and for upside-down faces, to ensure that observers were actually processing these stimuli as faces. For example, when faces are seen in an upside-down orientation, the identity of even well known faces is difficult to determine (Yin, 1969). In contrast, the perception of individual features is not influenced as strongly by turning the picture upside-down. It was important to include this kind of a control in our experiment, to ensure that observers were not "cheating" in our task by performing it in a way that bypasses the normal route for face perception. The results from the upside-down control conditions indicated that observers were indeed treating the images as faces.

The important result of the face study was that change detection in faces was strongly influenced by an observer's expectation regarding which aspect of the face was likely to change on a given trial. Observers detected an expected change in a face, both in identity and in emotional expression, more rapidly and accurately than an unexpected change in the same face. This was true not only when attention was distributed across five faces, as would be expected by almost all theories of perception on the grounds that detection of change in an individual face would require a time-consuming narrowing of attention to an individual face. But it was still equally true when observers were examining a single face, a face that was at the center of their gaze and could be attended single-mindedly because it was unaccompanied by any other stimuli. Again, it is important to emphasize that this result was not contaminated by any tendency to detect change in itself, since "change" and "no change" trials were equally represented in all conditions of the experiment. Yet, attending to a face did not permit equally efficient change detection of expected and unexpected changes (see also Schyns & Gosselin, 2003).

We interpret these two studies as extending Julian Hochberg's original observation that "…our perception of objects is not everywhere dense." That is, these results take us beyond the limitations that derive from having a fovea that registers only a small region of the visual field in much detail at any moment in time. These limitations of foveal vision are very real and must be overcome every time we view an extended scene. Yet, an even more fundamental limit on perception derives from the fact that our visual consciousness is able to focus on only one mental schema at a time. We believe that this way of framing the limitations of vision helps to clarify one of the puzzles of the change blindness literature. The puzzle is that although focused attention is necessary to detect change in a scene (Rensink et al., 1997), focused spatial attention is not itself sufficient for reliable change detection. Its lack of sufficiency can be seen in the cases of change blindness that have been reported even when foveal viewing of an object has been ensured and attention has been drawn to the object that is about to change (Levin & Simons, 1997; Simons & Levin, 1998). Indeed, the change-detection experiments just described are another instance of this lack of immunity to change blindness when viewing attended objects.

What we believe is missing in this way of talking about "attention" is the important distinction between merely orienting toward "a location" or "an object" versus attending to the "match between the contents of an expected mental schema for a location (an object) and the sensory information at that location (about that object)." Our view is that perception is fundamentally limited by the inherently serial nature of the schema matching process (Di Lollo, Rensink & Enns, 2000; Enns & Di Lollo, 2000; Kawahara, Enns & Di Lollo, 2002), not only by the practically serial nature of eye movements and attention over spatial locations. Support for this view comes from many recent experiments showing the severe, serial limitations on processing that remain when observers view rapid sequences of foveal images (Shapiro, Arnell & Raymond, 1997; Visser, Bischof & Di Lollo, 1999). The challenge for vision research in the future will be to try to understand the basis of this limit. Why is it that the massively parallel and interconnected neural system called the brain is only able to bring one mental schema to consciousness at a time? Although we do not have an answer at present, it is important to acknowledge the large contribution that Julian Hochberg has made in bringing modern researchers to the point where they are even able to pose this question.

References

Archambault, A., O'Donnell, C. & Schyns, P.G. (1999). Blind to object changes: When learning the same object at different levels of categorization modifies its perception. Psychological Science, 10, 249-255.

Austen, E., & Enns, J. T. (2000). Change detection: Paying attention to detail. Psyche, 6, 11.

Austen, E. & Enns, J.T. (2003). Change detection in an attended face depends on the expectation of the observer. Journal of Vision, 3, 64-74.

Baylis, G.C. (1994). Visual attention and objects: Two-object cost with equal convexity. Journal of Experimental Psychology: Human Perception and Performance, 20, 208-212.

Biederman, I., Rabinowitz, J. C., Glass, A. L. & Stacey, E. W. J. (1974). On the information extracted from a glance at a scene. Journal of Experimental Psychology, 103, 597-600.

Davis, G., Driver, J., Pavani, F, & Shephard, A. (2000). Reappraising the apparent costs of attending to two separate visual objects. Vision Research, 40, 1323-1332.

Di Lollo, V., Enns, J. T., & Rensink, R. A. (2000). Competition for consciousness among visual events: The psychophysics of reentrant visual processes. Journal of Experimental Psychology: General, 129, 481-507.

Duncan, J. (1984). Selective attention and the organization of visual information. Journal of Experimental Psychology: General, 113, 501-517.

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. Journal of Experimental Psychology: General, 123, 161-177.

Enns, J. T. & Di Lollo, V. (2000). What's new in visual masking? Trends in Cognitive Sciences, 4, 345-352.

Hochberg, J. (1968). In the mind's eye. In R. N. Haber (Ed.), Contemporary theory and research in visual perception (pp. 309-331). NY: Holt Rhinehart and Winston.

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

Hochstein,S., & Ahissar, M., (2002). View from the top: Herarchies and reverse hierarchies in the visual system. Neuron, 36, 791-804

Kawahara, J., Enns, J.T., & Di Lollo, V. (2002). Task switching mediates the attentional blink even without backward masking. Perception & Psychophysics, 65, 339-351.

Lamb, M.R., Pond, H.M., & Zahir, G. (2000). Contributions of automatic and controlled processes to the analysis of hierarchical structure. Journal of Experimental Psychology: Human Perception and Performance, 26, 234-245.

Levin, D.T., & Simons, D.J. (1997). Failure to detect changes to attended objects in motion pictures. Psychonomic Bulletin and Review, 4, 501-506.Navon, 1977

Marr, D. (1982). Vision. New York: Freeman.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. Cognitive Psychology, 9, 353-383.

Peterson, M. A., & Hochberg, J. (1983). Opposed-set measurement procedure: A quantitative analysis of the role of local cues and intention in form perception. Journal of Experimental Psychology: Human Perception and Performance, 9, 183-193.

Peterson, M.A., & Gibson. B.S. (1991). Directing spatial attention within an object: Altering the functional equivalence of shape descriptions. Journal of Experimental Psychology: Human Perception and Performance, 17, 170-182.

Rensink, R.A., O'Regan, J.K., & Clark, J.J. (1997). To see or not to see: The need for attention to perceive changes in scenes. Psychological Science, 8, 368-373.

Rensink, R. A. (2000). Visual search for change: A probe into the nature of attentional processing. Visual Cognition, 7, 345-376

Rensink R. A. (2002). Change detection. Annual Review of Psychology, 53, 245-277.

Scholl, B. J. (2000). Attenuated change blindness for exogenously attended items in a flicker paradigm. Visual Cognition, 7, 377 - 396.

Schyns, P. G., & Gosselin, F. (2003). Diagnostic use of scale information for componential and holistic recognition. In M. A. Peterson & G. Rhodes (Eds.), Perception of faces, objects, and scenes (pp. 120-145). Oxford, UK: Oxford University Press.

Shapiro, K. L., Arnell, K. A., & Raymond, J. E. (1997). The attentional blink: A view on attention and a glimpse on consciousness. Trends in Cognitive Science, 1, 291-296.

Simons, D. J. & Levin, D. T. (1997). Change blindness. Trends in Cognitive Science, 1, 261-267.

Simons, D.J., & Levin, D.T. (1998). Failure to detect changes to people during a real-world interaction. Psychonomic Bulletin and Review, 5, 644-649.

Smilek, D., Eastwood, J. D., & Merikle, P. M. (2000). Does unattended information facilitate change detection? Journal of Experimental Psychology: Human Perception and Performance, 26, 480-487.

Visser, T. A. W., Bischof, W. F., & Di Lollo, V. (1999). Attentional switching in spatial and non-spatial domains: evidence from the attentional blink. Psychological Bulletin, 125, 458-469.

Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. Psychonomic Bulletin & Review, 1, 202-238.

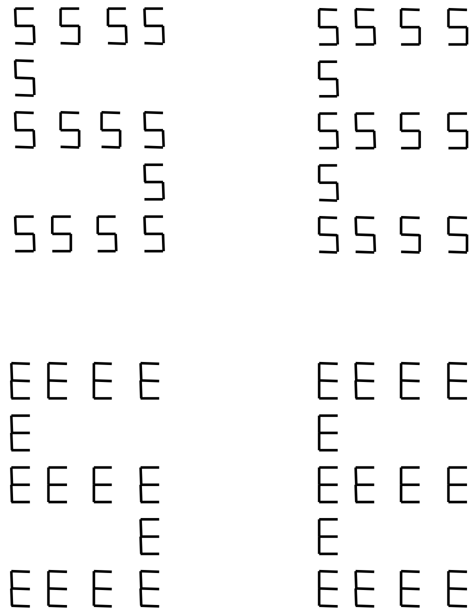
Wolfe, J.M., Klempen, N., & Dahlen, K. (2000) Post-attentive vision. Journal of Experimental Psychology: Human Perception and Performance, 26, 693-716.

Yin, R.K. (1969). Looking at upside-down faces. Journal of Experimental Psychology, 81, 141-145.

Figure Captions

Figure 1. (A) Compound letters used in studies of change detection (Austen & Enns, 2000). Letters on the left form a global S; letters on the right form a global E. Upper letters are S at the local level; lower letters are E at the local level. (B) Faces used in studies of change detection (Austen & Enns, 2003). Faces on the left are both person 1; faces on the right are person 2. Upper faces show 'happy' expression while lower faces show 'sad' expression.

A.



B.