

Reports of the icon's impending demise are premature

John Jonides

Department of Psychology, University of Michigan, Ann Arbor, Mich. 48104

Look up from this page for a moment, scan around the room, and reflect on your phenomenological experience of the visual scene. While you looked around, were you confused about the relative positions of objects? Did the scene seem blurred? Were you disoriented by each new fixation? Did each fixation seem like a brand-new view of the scene? Probably not. When scanning the visual world, most of us perceive it as coherent, clear, and continuous. How do we account for this experience in view of the fact that our scanning behavior consists of an alternating series of fixations and brief saccades? To put it simply, how do we manage to experience a continuous visual scene given temporally and spatially discontinuous input?

One hypothesis that merits attention rests on the claim that early in the stream of processing, there exists a storage device capable of holding the contents of more than one fixation. The key feature of this device, by hypothesis, is that it stores information from multiple fixations in a code that is tied to spatial, not retinal, coordinates. In this way, then, a composite image of a scene could be constructed, in essence, with individual fixations serving as the building blocks. This composite image could then serve as the basis of our panoramic perceptual experience in that it would contain more information than was included in any one fixation.

What would be the source of inputs for this hypothesized spatiotopic buffer? One possibility is that the data could be drawn from fixations themselves. After all, as Haber notes, each fixation typically lasts for at least a quarter second. But, as some have speculated, perhaps information is not processed during

the entire duration of each fixation. In this case, one might wish to suppose that an earlier buffer outlasts the duration of useful processing for each fixation, storing the contents of that fixation in retinotopic form. This information could then be passed on to the hypothesized storage device that retains environmentally coded information.

Either of these possibilities raises the question of how information in retinotopic coordinates can be converted to environmental coordinates. Again, there are several possibilities. One might suppose, with Helmholtz, that momentary knowledge of eye position and information from a retinal image are somehow combined to yield an image that is coded spatially. Or one might propose that overlap in visual detail in a scene from one fixation to another is used as a basis for spatially reconciling the individual glimpses. There are other possibilities as well. Suffice it to say that the problem of *hypothesizing* a spatially coded memory with well-specified inputs from multiple fixations is not a difficult one. The question is, is this a reasonable hypothesis to test?

I think so. As Haber mentions, however, there is a great deal of skepticism about this view (Hochberg 1968; Neisser 1967; 1976). The principal rationale for skepticism is that a persisting image that outlasted the duration of a fixation and the succeeding saccade would merely hinder the extraction of information from the next fixation because of a sort of double-exposure effect (take the photographic analogy as seriously as you wish). To be sure. But this would only happen if the two images overlapped, as they would if they were both coded in retinal coordinates. But suppose, as we have suggested, that the images had been reconciled prior to storage so that each represented its proper *spatial* position. In this case, not only might the first spatially correct image not hinder processing of the second, it might actually help, even though there might be interference between the first and second retinal images.

The hypothesis we have been discussing would be mere speculation were it not for an accumulating body of evidence that offers some support. The first line of evidence is represented by the studies of Ritter (1976) and Davidson, Fox, and Dick (1973). Ritter (1976) has shown that two light flashes presented at the same spatial locus, but separated by a saccade, are perceived as one if the interval between them is less than 75 msec. Davidson et al. (1973) revealed a similar effect for two figures presented at the same spatial location within 85 msec of each other, but separated by a saccade. Subjects reported that the figures fused together at their proper spatial positions, which were quite different retinally. Although this line of evidence should be viewed cautiously since it is based on introspection only, it does suggest an integration of spatially reconciled information over time.

Wolf, Hauske, and Lupp (1978; 1980) have provided some corroborating evidence from a different paradigm. They found a detection threshold decrease for a postsaccadic test pattern exposed at a saccade goal when an identical presaccadic pattern had been presented to subjects at the same spatial location. This suggests a position-specific integration of information from the periphery of one fixation and the fovea of the next, a suggestion supported by the findings of Jonides, Irwin, and Yantis (1982). Once again, caution is needed in interpreting these results since they are recent and have not yet been carefully replicated. But they do lead one to suppose that a spatially coded visual memory may be a viable construct.

Finally, there is the evidence that derives from viewing figures that are passed behind slits (e.g., Hochberg 1968). There is little doubt that under some conditions, the phenomenal coherence and completeness of a figure in such tasks is due to retinal painting, a result that implicates a retinally coded memory. There is also little doubt, though, that there are other circumstances in which subjects achieve a spatially correct image of a figure for which an account based on retinal coordinates of the image parts is simply inappropriate (R. Fendrich,

personal communication, 1982). In the latter case it seems most appropriate to suppose that successive glimpses from the slit are somehow used to construct a spatially coded image of the entire figure, although this image may not be veridical.

At this point, I am sure that the anxious reader is asking what the relationship is between the retinally or spatially coded memories under discussion here and what has been called iconic memory. The answer is that we don't know. But surely, you might exclaim, after 22 years of research on iconic memory, there must be mounds of evidence that bear on the issue of whether the icon is used in the service of normal perception, specifically to integrate information from successive fixations. As the short reference list in Haber's article attests, however, this is not so. And the reason, as Haber has recognized, is that little thought has been devoted to the probable function of the icon (other than lengthening various resumé's, that is). But as I hope I have made clear above, there is a perfectly plausible function for an early visual memory that codes information by spatial coordinates, and perhaps also an earlier memory that makes use of a retinotopic code. This is a timely hypothesis because it implicates multiple early visual memories, an implication that is also suggested by recent reviews of the iconic literature itself (e.g., Coltheart 1980).

To summarize then, it is not reasonable to call for an end to research on the icon at this time. On the contrary, research concerned with the mechanism of iconic storage should proceed apace, but now guided, as perhaps it should have been all along, by sensible hypotheses concerning its function. It is entirely too premature to assert that evidence about iconic storage is just so much debris in the path to understanding normal perceptual processes.

ACKNOWLEDGMENTS

The preparation of this manuscript was supported by NIMH grant 1 R03 MH36869-01 and by a grant from AFOSR.

Textons, rapid focal attention shifts, and iconic memory

Bela Julesz

Bell Laboratories, Murray Hill, N.J. 07974

I vividly recall the period in the sixties when Ulrich Neisser visited many psychological laboratories, including my own, in search of the enigmatic "temporal moment," the "quantum of psychological time." He thought that the existence of such a basic element might unify psychology, which otherwise seemed rather fragmented. Perhaps it was a disappointment that the "temporal moment" proved to be elusive, each psychological process seeming to have its own processing unit (or time constant). So, when his influential book (Neisser 1967) appeared, I was not surprised that it was not on the temporal moment; on the other hand, I was surprised that "iconic memory" took its place as a psychological universal, at least for vision. For audition he took another universal, "echoic memory," and regarded our finding (Guttman & Julesz 1963; Julesz & Guttman 1963) of 1 sec (the longest perceivable repetition of periodic random noise segments without scrutiny) as the best example for such a memory.

Even in vision the various short-term memories are as plentiful as the temporal moments proved to be. Let me cite from my own work a few such visual short-term memories with greatly varying durations. While the iconic memory of Sperling (1960) is about 250 msec, I have shown (Julesz 1964) that for random-dot stereograms the persistence of an ordered state lasts about 50-70 msec. This time was confirmed by Ross and Hogben (1974) using dynamic random-dot stereograms. So, the "cyclopean icon" has about a quarter of the duration of the classical