

Visual memory: What do you know about what you saw?

Jeremy M. Wolfe

Recent studies of visual perception are bringing us closer to an understanding of what we remember – and what we forget – when we recall a scene.

Address: Center for Ophthalmic Research, Brigham and Women's Hospital, 221 Longwood Avenue, Boston, Massachusetts 02115, USA.

Current Biology 1998, 8:R303–R304
<http://biomednet.com/elecref/09609822008R0303>

© Current Biology Ltd ISSN 0960-9822

How good is your memory? One line of research starting about 30 years ago shows that your memory for visually presented material is quite remarkably good [1,2]. In a typical picture recognition study, subjects are shown a number of scenes — such as images cut from a glossy travel magazine — each of which is presented for a second or two. In the test phase, subjects are shown a second set of scenes, half of them from the first set and the other half presented for the first time. The task is to identify members of the second set as old or new. Subjects perform very well on such a task, even when thousands of pictures are shown [3,4].

What are the subjects of such a study remembering? Common sense tells us that our memory for a picture is not some sort of highly detailed neural photocopy. Indeed, the details of the image are not well remembered [5]. Imagine trying to distinguish new from old from among a thousand different pictures of the stacks in the university library. A flurry of recent research has shown how bad we are at recognizing differences between similar scenes or changes between two versions of the same scene, a phenomenon that has been referred to as ‘change blindness’ [6,7]. Subjects might be shown a picture of an airplane on a runway, and in the second view the engines would be removed from the plane. The two images, one with and the other without engines, alternate on the screen every few seconds, with a blank screen presented in between to mask luminance transients. It can take a surprisingly long time to notice this change. Subjects also fail to notice changes made during an eye movement [8–10]. This is exploited in movies, where cuts between views render subjects insensitive to changes in clothing, props or even the identity of actors [11].

How can we reconcile excellent performance on picture recognition with dismal performance on change detection? One possibility is that observers do not remember the scene *per se*. Rather, they remember the gist of the scene. Thus, in picture recognition, where all the pictures are quite different, subjects can say to themselves, “Ah yes, I

have seen a picture of a burning house; no, I didn't see a picture of a cat in the bathtub”. By this account, change blindness occurs because the change does not alter the gist. A conversation between two women remains a conversation between two women, even if the clothing or the props change. In support of this idea, there is strong evidence that the meaning of a scene can influence memory for that scene. For instance, Brewer and Treyans [12] had subjects wait in an office, and then questioned them about the contents of the office. Subjects routinely reported books in the office, not because books were present — they were not — but because books are part of the schema for what should be in an office. People routinely remember seeing more of a scene than was presented [13,14]. On a more sinister note, memory for scenes can be colored by the biases of the observer [15].

The difficulty with the appealing idea that we remember the gist of a scene is that there is no consensus about the contents of a ‘gist’. Intuition suggests that an inventory of some of the objects in the scene should be at least a part of the gist. If you asked someone to describe a scene, you would be surprised if the description named no objects but relied only on a description of features, such as color or size. A recent experiment by Luck and Vogel [16] seems to show this coding into memory for objects, rather than simple features. They performed a variation of a change-detection experiment. Two arrays of items were presented to subjects; on half the trials, the second array contained one item that was changed. If one-to-three colored squares were presented, subjects could perfectly detect the change; performance fell off with larger set sizes. These results suggest that subjects can keep track of four colors. Now, suppose that each item on the screen could vary in color, orientation, size and the presence or absence of a gap. Would subjects be able to keep track of just four individual features, or would they be able to keep track of up to four objects with all of their associated features? The answer, in a variety of versions of this experiment, is that subjects kept track of objects. They could detect any single feature change in any of up to four objects, even though that meant keeping track of more than a dozen individual features.

There is a bottleneck between vision and memory. If you close your eyes, you will immediately lose access to many of the details that were obvious a moment ago. The results of Luck and Vogel [16] show that it is objects and not raw features, that move through that bottleneck. The selection of objects is governed by attention. There is copious evidence that it is easier to attend to properties belonging

to one object than to spatially equivalent properties spread over two or more objects [17–19].

Where does this leave us in the search for the gist of a picture? Evidence from visual search experiments suggests that objects can be identified at a rate of at least 20 per second [20]. It is possible that the rate of transfer into any sort of stable memory runs at the slower rates seen in ‘rapid serial visual presentation’ experiments (see [21], for example). In either case, a relatively brief presentation of a scene would allow several objects to be identified and passed to memory. Is that list the gist? One could imagine that a list of *N* objects would be sufficient to categorize a scene, but a series of thought experiments tells us that a gist is more than a list. Some relationships between objects must be coded into the gist [5]. A picture of milk being poured from a carton into a glass is not the same as a picture of milk being poured from a carton into the space next to a glass, even if all of the objects are the same. Moreover, even if all the propositional relationships between objects remain the same, some information about the spatial layout must be incorporated into the gist: consider, for example, the fact that subjects can be quite good at telling if an image has been left–right reversed in the test phase of a picture recognition experiment [22].

Beyond object relations and spatial layout, the gist seems to contain information about the presence of as yet unidentified objects. Imagine a scene of a toy drawer jumbled with toys. Only a few toys might be identified, but the gist would surely include the fact that there were a lot of other objects that could be identified, given time. Finally, at the most basic level, the gist would seem to include impression of the low-level visual features that fill the scene. Imagine the milk, the carton and the glass in their proper spatial relationships. Even if those are the only identified objects, it will make a difference to the gist if the space around the objects is empty or filled with this ‘visual stuff’. In this view, the gist of a scene would have, as its foundation, visual stuff spread out over some representation of surfaces and objects in three-dimensional space. Added to that base would be information about the identity and relationships of a limited number of the objects in the scene.

This definition of gist is only a proposal at this point. There is, however, evidence that each of its components are available in a brief look at a scene. Information about basic features [23], the existence of surfaces [24] and objects [25], and their three-dimensional disposition [26] is all available ‘preattentively’. Add that material to the list of objects passed by attention through the bottleneck to memory, and you might just have the gist. Even if gist is represented in this way, it will not be easy to explain how a brain can remember thousands of gists in a picture recognition experiment. Another program of research would be

needed to show that changes in the gist were necessary and sufficient for efficient change detection. Given the vitality of this area of research, it seems probable that we will have a clearer picture of scene recognition in the next few years.

References

1. Shepard RN: Recognition memory for words, sentences, and pictures. *J Verb Learn Verb Behav* 1967, 6:156-163.
2. Potter MC, Levy EI: Recognition memory for a rapid sequence of pictures. *J Exp Psychol* 1969, 81:10-15.
3. Standing L, Conezio J, Haber RN: Perception and memory for pictures: single trial learning of 2500 visual stimuli. *Psychon Sci* 1970, 19:73-74.
4. Standing L: Learning 10,000 pictures. *Q J Exp Psychol* 1973, 25:207-222.
5. Mandler J, Ritchey GH: Long-term memory for pictures. *J Exp Psychol [Hum Learn]* 1977, 3:386-396.
6. Rensink R, O'Regan JK, Clark JJ: To see or not to see: the need for attention to perceive changes in scenes. *Psychol Sci* 1996, 8:368-373.
7. Simons DJ, Levin DT: Change blindness. *Trends Cog Sci* 1997, 1:261-267.
8. Grimes J: On the failure to detect changes in scenes across saccades. In *Perception*. Edited by Akins K. New York: Oxford University Press; 1996:89-110.
9. Irwin DE: Integrating information across saccadic eye movements. *Curr Dir Psychol Sci* 1996, 5:94-100.
10. Blackmore SJ, Brelstaff G, Nelson K, Troscianko T: Is the richness of our visual world an illusion? Transsaccadic memory for complex scenes. *Perception* 1995, 24:1075-1081.
11. Levin DT, Simons DJ: Failure to detect changes to attended objects in motion pictures. *Psychol Bull* 1997, 4:501-506.
12. Brewer WF, Treysans JC: The role of schemata in memory for places. *Cognit Psychol* 1981, 13:207-230.
13. Intraub H, Bender RS, Mangels JA: Looking at pictures but remembering scenes. *J Exp Psychol Learn Mem Cogn* 1992, 18:180-191.
14. Intraub H, Richardson M: Wide-angle memories of close-up scenes. *J Exp Psychol Learn Mem Cogn* 1989, 15:179-187.
15. Loftus EF, Palmer JC: Reconstruction of automobile destruction: An example of the interaction of language and memory. *J Verb Learn Verb Behav* 1974, 13:585-589.
16. Luck SJ, Vogel EK: The capacity of visual working memory for features and conjunctions. *Nature* 1997, 390:279-281.
17. Duncan J: Selective attention and the organization of visual information. *J Exp Psychol Gen* 1984, 113:501-517.
18. Duncan J: Coordination of what and where systems in the visual control of behaviour. *Perception* 1993, 22:1261-1270.
19. Baylis GC, Driver J: Visual attention and objects: evidence for hierarchical coding of location. *J Exp Psychol Hum Percept Perform* 1993, 19:451-470.
20. Wolfe JM: Visual search. In *Attention*. Edited by Pashle H. London: UCL Press; 1998.
21. Potter MC: Short-term conceptual memory for pictures. *J Exp Psychol [Hum Learn]* 1976, 2:509-522.
22. Intraub H: Presentation rate and the representation of briefly glimpsed pictures in memory. *J Exp Psychol [Hum Learn]* 1980, 6:1-12.
23. Treisman A: Properties, parts, and objects. In *Handbook of Human Perception and Performance*. Edited by Boff KR, Kaufmann L, Thomas JP. New York: Wiley and Sons; 1986:35.1-35.70.
24. He JJ, Nakayama K: Surfaces versus features in visual search. *Nature* 1992, 359:231-233.
25. Wolfe JM, Bennett SC: Preattentive object files: shapeless bundles of basic features. *Vision Res* 1997, 37:25-44.
26. Enns JT, Rensink RA: Scene based properties influence visual search. *Science* 1990, 247:721-723.