

Recognizing Briefly Presented Pictures: Feedforward Processing?

Mary C. Potter

Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology

A typical eye fixation lasts for 250 ms: how much can we understand with such a brief glimpse? When viewing novel pictures in sequence using RSVP (rapid serial visual presentation), viewers can accurately detect named targets such as *picnic* or *harbor with boats* shown as briefly as 100 ms, although reliable memory for pictures requires at least 500 ms (Potter, 1976; Intraub, 1981). The speed of visual comprehension raises the question of whether feedforward processing--the initial wave of information from the retina through the cortical hierarchy--can be sufficient for momentary comprehension.

In feedforward models of the visual system (e.g., Serre, Oliva, & Poggio, 2007) units that process the stimulus are hierarchically arranged, with small receptive fields in the retina converging to represent larger and larger receptive fields and increasingly abstract information along a series of pathways to inferotemporal cortex (IT) and higher. A lifetime of visual experience tunes this hierarchical structure, which acts as a filter that permits understanding in a single forward pass. Evidence for feedforward processing has come from studies of response time to make a yes response or forced choice to the picture of an animal, vehicle, face, or the like, with minimal times as short as 150 or even 100 ms (Fabre-Thorpe, 2011), arguably too short for feedback.

Under normal viewing conditions perception is thought to result from a combination of feedforward and feedback connections, with feedforward activation generating possible interpretations that are fed back and compared with lower levels of processing for confirmation, establishing reentrant loops. It has been proposed that we become consciously aware of what we are seeing only when such reentrant loops have been established (e.g., Hochstein & Ahissar, 2002; Lamme, 2006; DiLollo, 2012). A related suggestion is that consciousness arises from recurrent long-distance interactions (e.g., Dehaene, Kergsberg, & Changeux, 1998). Feedback has been estimated to require a minimum of 50 ms, so that presentation durations shorter than that would not become conscious in a serial presentation in which each stimulus was immediately followed by another stimulus, producing masking. It is possible, however, that a top-down selective set or expectation could modify or resonate with the feedforward process, obviating the need for reentrant processing and enabling conscious perception with presentation durations shorter than 50 ms.

In a recent study (Potter, Wyble, Haggmann, & McCourt, in prep.) participants attempted to detect a named target (e.g., *couple smiling*) among 6 or 12 unfamiliar pictures shown for 80, 53, 27, or 13 ms. Crucially, the target was named either before the sequence (providing a specific attentional set), or just after the sequence (with no advance attentional set). Performance declined with decreasing duration but was above chance at all durations, whether the target was named in advance or only after the sequence. The results suggest that conceptual understanding of a new picture can be obtained in a single forward pass without feedback, even without advance knowledge of what to look for. The implications for theories of perceptual learning will be considered.