Published in final edited form as:

Vis cogn. 2016 ; 24(1): 2–14. doi:10.1080/13506285.2016.1170745.

Ultrafast scene detection and recognition with limited visual information

Carl Erick Hagmann and Mary C. Potter

Massachusetts Institute of Technology

Abstract

Humans can detect target color pictures of scenes depicting concepts like *picnic* or *harbor* in sequences of six or twelve pictures presented as briefly as 13 ms, even when the target is named after the sequence (Potter, Wyble, Hagmann, & McCourt, 2014). Such rapid detection suggests that feedforward processing alone enabled detection without recurrent cortical feedback. There is debate about whether coarse, global, low spatial frequencies (LSFs) provide predictive information to high cortical levels through the rapid magnocellular (M) projection of the visual path, enabling top-down prediction of possible object identities. To test the “Fast M” hypothesis, we compared detection of a named target across five stimulus conditions: unaltered color, blurred color, grayscale, thresholded monochrome, and LSF pictures. The pictures were presented for 13–80 ms in six-picture rapid serial visual presentation (RSVP) sequences. Blurred, monochrome, and LSF pictures were detected less accurately than normal color or grayscale pictures. When the target was named *before* the sequence, all picture types except LSF resulted in above-chance detection at all durations. Crucially, when the name was given only *after* the sequence, performance dropped and the monochrome and LSF pictures (but not the blurred pictures) were at or near chance. Thus, without advance information, monochrome and LSF pictures were rarely understood. The results offer only limited support for the Fast M hypothesis, suggesting instead that feedforward processing is able to activate conceptual representations without complementary reentrant processing.

Keywords

object recognition; scene understanding; identification; feedforward; RSVP; magnocellular

The visual system extracts information from each saccadic fixation of the eyes. Previous experiments have shown that conceptual information can be readily derived from sequentially presented pictures with exposure durations shorter than a typical fixation of 300 ms (Intraub, 1981; Meng & Potter, 2008; Potter, 1976). However, the limits and mechanisms of such rapid visual processing are only beginning to be established (Potter, Wyble, Hagmann, & McCourt, 2014; Potter & Hagmann, 2014). Despite the vast increase of

Correspondence concerning this article should be addressed to Carl E. Hagmann, Department of Psychology, Syracuse University, 426 Ostrom Ave, Syracuse, NY 13210. cehagman@syr.edu.

Carl Erick Hagmann and Mary C. Potter, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts.

The authors report no conflict of interest involving this research.

knowledge about the visual system over the last half-century, there is as yet no definitive account of how we are able to understand visual stimuli with such brief exposure durations.

According to one class of models (e.g., Serre, Oliva, & Poggio, 2007), understanding is based on a feedforward process, with bottom-up information passing from the retina through the early visual system to the temporal lobe and the prefrontal cortex. In contrast, other models propose that top-down feedback from higher to lower levels is necessary for understanding in addition to the feedforward information (e.g., Bullier, 2001). In particular, it has been proposed that information in the magnocellular (M) pathway, which is transmitted faster than that in the parvocellular (P) pathway, is fed back from higher to lower levels to provide guidance to the feedforward P pathway. Here we test predictions of that hypothesis for detection and recognition of pictures that are blurred, black-white thresholded, or restricted to low spatial frequencies, compared with normal color or grayscale pictures.

In the primate lateral geniculate nucleus (LGN), four layers of the cortical column contain small ventral-pathway P cells and two layers contain larger, dorsal pathway M cells (Shapley, Kaplan, & Soodak, 1981). M neurons increase their responding at low contrast and spatial frequency relative to P neurons (Kaplan & Shapley, 1986), but are essentially “color-blind,” with illumination response that is either on or off regardless of wavelength (Livingstone & Hubel, 1988). Low spatial frequencies (LSF) represent coarse, global, visual information, while fine details are represented in the high spatial frequency (HSF) ranges. Just as color processing is restricted to the slower P projection (Schiller, Logothetis, & Charles, 1990), low spatial frequency information, as well as motion and flicker, is segregated to the faster M path. Sugase, Yamane, Ueno, and Kawano (1999) found a latency difference of 51 ms between LSF and HSF face information in the inferior temporal cortex (IT). However, Skottun and Skoyles (2007; 2008; Skottun, 2013) have suggested that manipulating spatial frequency is not the best way to separate M and P streams. Lesions to the P path may be the only sure way to isolate its psychophysical contributions (Schiller et al 1990). Further problems arise due to switching and mixing of inputs, requiring interactions between distinct M and P streams to be examined between the LGN and the primary visual cortex (Skottun & Skoyles, 2007).

Bullier (2001) proposed that the M stream performs the initial analysis of global attributes in visual input and suggested that the M analysis is made available on “active blackboards” in V1 and V2 as P information arrives. Building on these ideas, Bar (2003) proposed a recognition facilitation mechanism in which fast M pathways provide top-down facilitation of object recognition (henceforth called the Fast M hypothesis). Specifically, Bar et al. (2006) proposed that low spatial frequency information is projected rapidly to the orbitofrontal cortex (OFC) through the M pathway, where this coarse information “is sufficient for activating a minimal set of the most probable interpretations of the input, which are then integrated with the bottom-up stream of analysis to facilitate recognition” (Fig. 1. caption). They used fMRI and MEG experiments to test this hypothesis and found differential activation unique to object recognition in the OFC earlier than activation in other cortical regions that mediate object recognition, such as the fusiform face area (FFA). The early OFC activity was driven by LSF in the pictures, leading Bar et al. to suggest that the

OFC allowed for top-down feedback at short latencies. Kveraga, Boshyan, and Bar (2007) subsequently found that information biased towards the M path could facilitate object recognition earlier than P-biased information, and the OFC was activated more by M- than P-biased stimuli. They proposed that P-biased stimuli reduced top-down facilitation by the OFC, causing greater engagement of the fusiform cortex for bottom-up object recognition. Others who have suggested similar ideas (e.g., Laycock, Crewther, & Crewther, 2008; Patai, Buckley, & Nobre, 2013) include Tapia and Breitmeyer (2011), who say, “M channels indirectly contribute to conscious object vision via top-down modulation of reentrant activity in the ventral object-recognition stream” (Abstract).

Processes that have short latencies might be expected to play a disproportionate role when stimuli are presented very briefly. To test the limits of visual processing ability as exposure duration is shortened, Potter et al. (2014) used sequences of six or 12 color photographs each shown for 13, 27, 53, or 80 ms in a rapid serial visual presentation (RSVP). The pictures were of a large variety of scenes and the target picture was specified by a descriptive name given immediately before or after the sequence. Although detection (d') improved with longer durations, performance was above chance even at 13 ms and even when the name was presented after the sequence. This result is consistent with feedforward models of visual processing in which a single initial feedforward wave of neural activity can be sufficient to allow detection and identification of an object.

Pictures in RSVP sequences are both backward- and forward-masked, limiting the time available for processing at each level in the visual hierarchy to the duration of the picture. This also limits recurrent processing, on the assumption that the feedforward and recurrent processing of a given stimulus must be temporally coincident (e.g., Keyser, Xiao, Földiák, & Perrett, 2001). If recurrent information requires at least two successive synapses (one up, one back), the recurrent information would be delayed by 20 ms or more (given estimates of synaptic transmission time), which would be too late to coincide with a 13 ms and perhaps even a 27 ms stimulus. The Fast M hypothesis (Bar, 2003; Bar et al, 2005; Kveraga et al, 2007), which holds that M information arrives by a separate, faster pathway to the OFC and is fed back to intersect with a slower P pathway, could solve such a coincidence problem, even for stimuli that are seen very briefly. In the present study we investigate whether detection of briefly presented pictures is based on coarse information of the kind carried by the M pathway.

We conducted five experiments with different types of filters and thus different amounts of contrast, color, and spatial frequency information. Experiment 1 tested unfiltered color pictures; Experiment 2, blurred color pictures; Experiment 3, grayscale pictures; Experiment 4, monochrome (black and white) pictures; and Experiment 5, pictures with only low spatial frequencies. As before (Potter & Hagmann, 2014), four different presentation durations were used: 13, 27, 53, or 80 ms per picture. Performance was expected to decrease with shorter durations in all conditions. If the M path can shuttle advance information to visual cortex, what information is it providing that might help observers under extreme conditions? Color was manipulated to see if the slower arrival of color information through the P route affected the benefit of color versus grayscale pictures at different presentation durations. Contrast was manipulated to test if binary luminance information in the M path could preserve

performance in thresholded relative to grayscale pictures. Performances with blurred and LSF pictures were compared to the standard color and grayscale pictures, respectively, to examine if low spatial frequency information facilitated object recognition in accord with the Fast M hypothesis. If low spatial frequencies carry the M information used to comprehend very briefly presented, masked pictures, then performance with the blurred, monochrome, and LSF pictures should be similar to that of unfiltered color and grayscale pictures when the presentation duration is very short – too short for recurrent processing. With longer presentation durations (>50 ms) that allow for recurrent processing in the P system, performance would improve for the unfiltered pictures, relative to the blurred, monochrome, and LSF pictures. We compared giving the target name just *before* with immediately *after* the stimulus sequence. Predictive information from the M pathway should be most useful in the *after* condition because there is no advance information to facilitate target processing.

General Method

The design and procedure, based on the method of Potter et al. (2014), was the same in each group in each experiment, except for the format of the pictures.

Participants

The participants in Experiments 1–4 were volunteers 18–35 years of age who were paid for their participation and signed a consent form approved by the MIT Committee on the Use of Humans as Experimental Subjects. The participants in Experiment 5 were volunteers 18–35 years of age who received course credit for their participation; the experiment took place in the Center for Autism Research in Electrophysiology lab (CARE Lab) at Syracuse University. The Syracuse University Institutional Review Board approved all consent and testing procedures. There were 16 participants in each name position group (Before/After) in each experiment; no one participated in more than one experiment and no participant was replaced. All participants were self-reported native English speakers. See Table 1 for gender and age.

Procedure

Participants viewed a rapid serial visual presentation (RSVP) of six pictures and were asked to detect a picture specified by a written name given just *before* or immediately *after* the sequence, between subjects. The one-to-four-word name reflected the gist of the target picture as judged by the experimenters. Examples are: *swan*, *traffic jam*, *boxes of vegetables*, *children holding hands*, *boat out of water*, *campfire*, *bear catching fish*, and *narrow street*.

For those in the *before* group, each trial began with a fixation cross for 500 ms, followed by the written name of the target for 700 ms and then a blank screen of 200 ms and the sequence of pictures. A blank screen of 200 ms also followed the sequence and then the question “Yes or No?” appeared and remained in view until the participant pressed “Y” or “N” on the keyboard to report whether he or she had seen the target. Those in the *after* group viewed a fixation cross for 500 ms at the beginning of the trial, followed by a blank screen of 200 ms and the sequence. At the end of the sequence another 200 ms blank screen

appeared, and then the name was presented simultaneously with the yes/no question until the participant responded. For all groups, on trials in which the target had been presented, the participant's response was followed by a two-alternative forced choice between two pictures that matched the target name. The participant pressed the "G" or "J" key to indicate whether the left or right picture, respectively, had been presented. In Experiments 1 and 2, the forced-choice pictures were normal color pictures, whether the presented sequence had been normal or blurred. Similarly, in Experiments 3–5 the forced-choice pictures were grayscale whether the sequence had been grayscale, monochrome, or LSF. On no-target trials the words "No target" appeared instead of the pair of pictures. Participants were then prompted to start the next trial by pressing Enter.

Design—Each experiment began with a practice block presented at 133 ms per picture, followed by eight blocks of experimental trials. Across blocks, the durations were 80, 53, 27, and 13 ms per picture, repeated with new pictures in the next four blocks. There were 22 trials per block, including six no-target trials. The target, which was never the first or last picture, appeared in serial position 2, 3, 4, or 5, balanced over target trials within each block so that each serial position was represented four times per block. Within a block the order of trials was randomized. Across every eight participants, the eight blocks of trials were rotated so that the pictures in each block of trials were seen equally often at each duration and in each half of the experiment. We used a between-subjects design so that no participant would see the same image more than once.

Apparatus—The experiments were programmed with Matlab 8.3 and run with the Stream package using the Psychophysics Toolbox extension (Brainard, 1997) version 3. The experiments were run on a Mac mini with 2.4 GHz, Intel Core 2 Duo processor. The Apple 17-in CRT monitor was set to a 1024 × 768 resolution with a 75-Hz refresh rate, gamma-corrected to 1.0. The room was normally illuminated. Timing precision was controlled with the Stream package for MATLAB. We excluded trials in which a timing error of + /− 3 ms or greater affected the target picture. At most, five trials per subject were excluded.

Materials—The 1056 pictures and 132 practice pictures were photographs previously used by Potter et al. (2014), taken from the Web and from other collections of pictures available for research use. They included a wide diversity of subject matter and settings, such as indoor and outdoor, and with and without people. In Experiment 1 participants saw normal color photographs. In Experiment 2 participants saw blurred versions of the color pictures. Blurring consisted of a disk blur with a 5-pixel radius, which affected contrast energy across the whole spectrum of spatial frequencies. In Experiment 3 participants saw grayscale pictures. Grayscale filtering was performed with Matlab's RGB2Gray function, which eliminates hue and saturation information but preserves luminance. In Experiment 4, participants saw monochrome versions of the grayscale pictures. To make the monochrome pictures, the grayscale pictures were altered with a discrete two-dimensional Fourier transform of the picture with its mean brightness subtracted out, followed by application of a Gaussian filter centered at zero frequency and thresholding, producing pictures containing only black and white pixels with a spatial frequency cut-off of 5 cycles/degree. In Experiment 5, participants saw grayscale pictures containing only low spatial frequencies.

Low spatial frequency pictures were created with a 5th order Butterworth filter 3.5% the width of the Fourier transform applied to the grayscale pictures, resulting in a spatial frequency cut-off at 3.8 cycles/degree. Prior to filtering, the grayscale pictures were luminance and contrast-matched. The DC component was restored following filtering to avoid negative luminance values. Figure 2 shows examples of the five types of pictures.

The pictures were sized to 300×200 pixels and were presented in the center of the monitor on a gray background during RSVP. The horizontal visual angle was 10.3° at the normal viewing distance of 50 cm. Each picture was presented only once to a given participant.

Analyses—To correct for extreme hit and false alarm values without introducing bias, we applied a loglinear adjustment to all conditions by adding 0.5 to the count of hits/false alarms and 1 to the number of trials before the d' calculation (Hautus, 1995; Stanislaw & Todorov, 1999). Planned paired t tests at each duration, separately for each group, compared d' with chance (0.0). For each experiment, we conducted repeated measures analyses of variance (ANOVA) on individual participants' d' measures as a function of the between-subjects factor *name position* and the within-subjects factor presentation *duration* (80, 53, 27, or 13 ms per picture). Direct comparisons between experiments include: Color vs. blurred, color vs. grayscale, grayscale vs. monochrome, and grayscale vs. LSF.

For the forced-choice responses on target-present trials, separate ANOVAs were carried out on accuracy across durations, conditional on whether the participant had responded “yes” (a hit) or “no” (a miss). The purpose of this analysis was to discover whether participants had understood more about the picture than just the information in the target name, and whether recognition depended on accurately detecting the target or not. ANOVAs were also carried out to compare forced choice accuracy with different picture types.

Experiment 1: Color Picture

Detection

Target detection results for Experiment 1 based on name position and duration are shown in Fig. 3 (red line). Overall, the results replicate those of Experiment 1 in Potter et al. (2014). Mean d' for each experiment is shown in Table 2. In general, detection was better the longer the duration, and better when the name was presented before the pictures rather than after. A Name position $\times$ Duration ANOVA of d' revealed a significant interaction, $F(3,90)=8.52$, $p=0.001$, $\eta^2=0.160$, with the advantage of the before condition increasing as picture duration increased. In the *before* condition, the target was detected significantly above chance (0.0) at all four durations, based on planned paired t -tests, $t(15)>5.88$, $ps<0.00003$. The *after* condition also produced significantly above chance detection at all durations, $t(15)>3.28$, $ps<0.005$.

Forced Choice

The forced choice results for Exp. 1 after correct yes responses (hits) are shown in Fig. 4 (red line). A name position $\times$ duration $\times$ response (yes or no) ANOVA of forced choice accuracy resulted in main effects of name position, $F(1,30)=24.27$, $p<0.001$, $\eta^2=0.107$, duration, $F(3,90)=7.41$, $p<0.001$, $\eta^2=0.081$, and response, $F(1,30)=97.075$, $p<0.001$,

$\eta^2=0.304$. Accuracy at recognizing the target when presented with two possible choice pictures was better in the *before* ($M=0.69$, $SE=0.015$) than *after* ($M=0.59$, $SE=0.015$) condition. Hits resulted in above chance recognition at all durations with both name positions, $t(7)>3.36$, $p<0.012$. After misses, recognition was above chance only at 13 and 53 ms, $t(7)>3.1$, $p<0.016$. Since misses consistently resulted in chance or near-chance forced choice performance across durations and picture types in subsequent experiments, forced choice performance following misses is not shown.

Experiment 2: Blurred Pictures

Detection

A name position x duration ANOVA of d' revealed main effects of position, $F(1,30)=35.2$, $p<0.001$, $\eta^2=0.183$ and duration, $F(3,90)=24.6$, $p<0.001$, $\eta^2=0.39$, and a significant interaction, $F(3,90)=3.77$, $p=0.013$, $\eta^2=0.092$, with the effect of duration greater in the *before* condition. As observed in Experiment 1, detection improved as duration increased, and detection was better when the name was presented before the pictures (see Fig. 3, blue line). The target was detected significantly above chance (0.0) at all four durations based on planned paired t-tests in both the *before* condition, $t(15)>2.99$, $p<0.009$ and the *after* condition, $t(15)>2.41$, $p<0.029$.

Forced Choice

The forced choice results for hits are shown in Fig. 4 (blue line). A name position x duration x response (yes or no) ANOVA of forced choice accuracy resulted in main effects of name position, $F(1,30)=31.76$, $p<0.001$, $\eta^2=0.096$, duration, $F(3,90)=8.15$, $p<0.001$, $\eta^2=0.104$, and response (yes-no), $F(1,30)=118.46$, $p<0.001$, $\eta^2=0.301$. There was also a name position x response interaction, $F(1,30)=4.306$, $p=0.046$, $\eta^2=0.015$. Following a hit, forced choice accuracy was above chance at all durations, $t(7)>4.02$, $p<0.005$, but was at chance at all durations following a miss. Forced choice accuracy was higher in the *before* ($M=0.67$, $SE=0.009$) than the *after* condition ($M=0.58$, $SE=0.011$).

Comparisons – Color v Blurred

Color and blur d' results were compared in a three-way ANOVA of name position, picture type, and duration. We found main effects of all three variables: picture type, $F(1,60)=73.41$, $p<0.001$, $\eta^2=0.24$, with the normal color pictures more readily detected than the blurred pictures; name position, $F(1,60)=94.61$, $p<0.001$, $\eta^2=0.29$, with detection better in the *before* than the *after* condition; and duration, $F(3,180)=71.27$, $p<0.001$, $\eta^2=0.47$, with better performance for the longer durations. There were significant interactions between picture type and name position, $F(1,60)=5.21$, $p=0.026$, $\eta^2=0.02$, and between name position and duration, $F(3,180)=10.84$, $p<0.001$, $\eta^2=0.12$.

In separate analyses of the two name positions, the *before* group had main effects of picture type, $F(1,30)=84.0$, $p<0.001$, $\eta^2=0.33$, and duration, $F(3,90)=63.0$, $p<0.001$, $\eta^2=0.63$. The *after* group also had main effects of picture type, $F(1,30)=15.2$, $p<0.001$, $\eta^2=0.14$, and duration, $F(3,90)=14.7$, $p<0.001$, $\eta^2=0.25$, but both effects were smaller than in the *before* condition. Tukey HSD tests confirmed that detection was significantly better for color

relative to blurred pictures for *before* at 13, 53, and 80 ms and *after* at 80 ms ($p < .05$), but not significant for *before* at 27 ms ($p = 0.12$) and *after* at 13, 27, and 53 ms ($ps > 0.085$).

With regard to the predictions of the Fast M hypothesis, the absolute difference between normal color and blurred color target detection was smaller with shorter durations, consistent with the hypothesis that the M channel conveys much of the earliest high-level information about the scene. The hypothesis that the advantage of the normal over the blurred picture would be further reduced in the *after* condition was also supported, since color pictures significantly outperformed blurred pictures only with 80 ms exposure time.

Forced choice test

On target-present trials, a two-alternative forced choice followed, whether the response had been a correct yes (a hit) or a false no (a miss). Note that the test pictures were normal color pictures in both the blur and normal conditions. We analyzed the proportion correct as a function of whether the response was a hit or miss, with picture type (blurred, normal), name position (before, after), and duration as variables. For the target trials to which the participant had said yes, the results are shown in Fig. 4. The main effect of picture type, with lower performance for the blurred pictures, was only marginally significant, $F(1,60) = 2.78$, $p = 0.10$, $\eta^2 = 0.015$. There were substantial main effects of name position, $F(1,60) = 53.33$, $p < 0.001$, $\eta^2 = 0.226$, and duration, $F(3,180) = 22.57$, $p < 0.001$, $\eta^2 = 0.202$. There were no interactions. This result indicates that, given a correct yes response, the same level of information about the picture was retained whether or not the picture had been blurred, whereas, even when a response was a hit, it was better remembered in the *before* than in the *after* condition.

Experiment 3: Grayscale Pictures

Detection

Detection results for grayscale pictures are shown in Fig. 3 (black line). A name position x duration ANOVA of d' showed that detection was better with slower rates, $F(3,90) = 36.45$, $p < 0.001$, $\eta^2 = 0.46$, and seeing the name before the pictures improved target detection, $F(1,30) = 38.44$, $p < 0.001$, $\eta^2 = 0.27$. There was a significant interaction, $F(3,90) = 2.85$, $p = 0.042$, $\eta^2 = 0.063$, with a larger effect of duration in the *before* condition. Based on planned paired t-tests, detection was above chance in the *before*, $t(15) > 3.12$, $ps < 0.007$ and *after* conditions, $t(15) > .54$, $ps < 0.003$.

Forced Choice

Grayscale forced choice results are shown in Figures 4 and 6 (black line) for hits only. A name position x duration x response (yes-no) ANOVA of forced choice accuracy resulted in main effects of duration, $F(3,90) = 5.89$, $p < 0.001$, $\eta^2 = 0.075$, and response, $F(1,30) = 123.85$, $p < 0.001$, $\eta^2 = 0.297$, as well as interactions between name position x response, $F(1,30) = 15.02$, $p < 0.001$, $\eta^2 = 0.048$, and duration x response, $F(3,90) = 3.49$, $p = 0.018$, $\eta^2 = 0.035$. Accuracy was very similar in the *before* ($M = 0.60$, $SE = 0.015$) and *after* conditions ($M = 0.59$, $SE = 0.008$). Correct yes responses with grayscale pictures resulted in

above chance forced choice recognition at all durations and both name positions, $ts(15) > 3.73$, $ps < 0.002$. Forced choice accuracy was not better than chance after misses.

Color vs. Grayscale Pictures

Many participants in Potter et al.'s (2014) study reported reliance on color in detecting targets. A color-based detection strategy makes sense if the target is *tomato* or *pumpkin*, objects for which participants could selectively attend to red- and orange-colored pictures, respectively. This strategy is less useful if the named concept does not have a typical color (e.g., *boat out of water*; *woman with umbrella*), or if the name is given after the sequence and color has not been encoded. Color is predominantly represented in the P pathway (Schiller et al., 1990): it is processed in areas V4/V8, prior to object related areas, but subsequent to early processing in areas V1 and V2 (Grill-Spector & Malach, 2004). Previous research by Meng and Potter (2008) found no difference between grayscale and colored pictures at durations of 53 ms and longer in both detection and recognition RSVP experiments. If detection of very brief pictures depends in part on the fast M pathway, color should have little or no benefit at short durations.

We compared the color condition in Exp. 1 to the grayscale condition in Exp. 3, with a picture type $\times$ name position $\times$ duration ANOVA of d' . We obtained a significant main effect of picture type, $F(1,60)=18.37$, $p<0.001$, $\eta^2=0.09$, with detection better for color than grayscale pictures ($d'=1.32$ and 1.02 , respectively). There were also significant effects of name position, $F(1,60)=96.53$, $p<0.001$, $\eta^2=0.33$, with the name *before* better than *after*, and of duration, $F(3,180)=86.21$, $p<0.001$, $\eta^2=0.50$, as seen in Fig. 3. There was a significant interaction between name position and duration, $F(3,180)=10.24$, $p<0.001$, $\eta^2=0.10$, with larger effects of duration in the *before* condition. No other interactions were significant: in particular, there were no interactions with picture type, $F_s < 1.0$. Despite the significant main effect of color versus grayscale picture types, Tukey HSD tests comparing picture type at each name position and duration were only significant in the *before* condition at 80 ms ($p=0.016$) and none were significant in the *after* condition ($ps > 0.59$). In the analysis of the forced choice results for hits, overall performance was not significantly better for color than for grayscale pictures (Fig. 4).

The Fast M hypothesis – that the M pathway facilitates detection by feeding back guesses that interact with the slower feedforward P pathway – implies that the guesses would not include color information. If the M pathway were responsible for the first stages of picture identification, then with very short durations one would expect less or no benefit of color, particularly in the condition in which the name comes after the sequence. There was no indication, however, that the significant overall color advantage was smaller with shorter exposure durations in the *after* condition.

Experiment 4: Monochrome Pictures

Detection

The detection results for thresholded monochrome pictures are shown in Fig. 3 (orange line). A name position $\times$ duration ANOVA of d' gave a main effect of name position,

$F(1,30)=21.98$, $p<0.001$, $\eta^2=0.184$, and an effect of duration, $F(3,90)=6.21$, $p<0.001$, $\eta^2=0.125$, with no interaction. However, inspection shows that in the *before* condition detection was better with slow rates than fast rates, but in the *after* condition longer viewing times did not significantly improve detection from 13 ms ($M=0.16$) to 80 ms ($M=0.37$). Monochrome pictures in the *after* condition produced low sensitivity across the durations and performance was significantly above chance only at 27 ms ($p=0.016$) and 80 ms ($p=0.002$).

Forced Choice

An ANOVA of forced choice accuracy produced effects of name position, $F(1,30)=8.04$, $p=0.008$, $\eta^2=0.02$, and response (yes/no), $F(1,30)=6.04$, $p=0.019$, $\eta^2=0.016$, as well as a name position x response interaction, $F(1,30)=5.64$, $p=0.024$, $\eta^2=0.015$. Overall, forced choice accuracy following correct detection of the target in monochrome picture sequences was close to chance in the *before* ($M=0.57$, $SE=0.015$) and *after* conditions ($M=0.52$, $SE=0.008$). There was above chance recognition on *before* trials, at 13, 53, and 80 ms, $t(15)>2.74$, $p<0.015$, but chance performance on *after* trials, consistent with the near-chance detection observed in that condition (see Fig. 4, orange line).

Comparisons - Monochrome vs. Grayscale

Detection—Overall, d' was markedly lower for monochrome pictures ($M=0.45$) than for grayscale pictures ($M=1.03$). In a three-way ANOVA of picture type, name position, and duration, we obtained main effects of all three variables: picture type, $F(1,60)=75.76$, $p<0.001$, $\eta^2=0.27$, with grayscale pictures more readily detected than monochrome pictures; name position, $F(1,60)=59.83$, $p<0.001$, $\eta^2=0.23$, with detection better in the *before* than the *after* condition; and duration, $F(3,180)=37.34$, $p<0.001$, $\eta^2=0.30$, with better performance for longer durations. There was a significant interaction between picture type and duration, $F(3,180)=8.82$, $p<0.001$, $\eta^2=0.09$, with accuracy increasing more sharply in the grayscale condition than in the monochrome condition as duration increased, and an interaction between name position and duration, $F(3,180)=4.85$, $p<0.01$, $\eta^2=0.05$, with larger duration effects when the name came before the sequence. The three-way interaction was not significant, $F<1.0$.

Tukey HSD tests of whether grayscale sensitivity was superior to monochrome at each name position and duration were significant ($p<0.05$), except for *before* at 13 ms ($p=0.97$) and *after* at 13 and 27 ms ($p>0.9$). Although detection was reduced when participants viewed monochrome pictures, targets on *before* trials were detected significantly above chance at all four durations with both picture types.

Forced Choice—On target-present trials, a two-alternative forced choice test, using grayscale pictures for both picture types, followed the yes/no task whether the response had been correct (a hit) or a false no (a miss). We analyzed the proportion correct as a function of whether the response was a hit or miss, with picture type (grayscale or monochrome), name position (before or after), and duration as factors.

Figure 4 shows the forced choice results following correct detection of the target. Overall, we found a large main effect of picture type, with lower performance for the monochrome pictures, $F(1,60) = 56.37$, $p < 0.001$, $\eta^2 = 0.211$. While grayscale pictures produced above chance recognition in the before and after condition, monochrome pictures only produced above chance recognition in the before condition. There were also main effects of name position, $F(1,60) = 22.65$, $p < 0.001$, $\eta^2 = 0.097$, and duration, $F(3,180) = 8.24$, $p < 0.001$, $\eta^2 = 0.089$. There was an interaction between picture type and duration, $F(3,180) = 5.26$, $p < 0.01$, $\eta^2 = 0.059$, with strong effects of duration only for grayscale pictures, not monochrome pictures.

Altogether, despite evidence that monochrome pictures could be identified better than chance when the target name was given in advance of the pictures, performance was rarely above chance when the name was given after the participant had seen the sequence, even when pictures had been viewed for 80 ms each. Though the monochrome pictures contained high frequency visual information, the lack of gradient may have prohibited the rapid detection that we had observed in the three previous experiments.

Experiment 5: Low Spatial Frequency Pictures

Detection

Performance with the LSF pictures was very poor, whether the name came before or after the picture sequence (see Fig. 3, light blue line). A name position x duration ANOVA of accuracy produced no significant main effect of name position, $F(1,30) = 0.066$, $p = 0.79$, $\eta^2 = 0.001$, but did produce a significant effect of duration, $F(3,90) = 3.37$, $p = 0.02$, $\eta^2 = 0.04$. There was no interaction, $F(3,90) = 0.74$, $p = 0.53$, $\eta^2 = 0.01$.

Not all conditions were at chance, however. One sample t-tests comparing d' to chance (0.0) showed that in the *before* condition, there was significant sensitivity at 80 ms, $t(15) = 3.32$, $p = 0.005$, and near significance at 53 ms, $t(15) = 1.88$, $p = 0.08$. In the *after* condition, performance was not significantly above chance at any duration, $t(15) < 1.55$, $p > .14$.

Forced Choice

A name position x duration x yes-no response ANOVA of forced choice accuracy produced a significant effect of duration, $F(1,30) = 4.97$, $p = 0.003$, $\eta^2 = 0.05$, and a name position x yes-no response interaction, $F(3,90) = 3.074$, $p = 0.03$, $\eta^2 = 0.044$. For correct yes responses (hits), forced choice accuracy was very similar in the *before* ($M = 0.51$, $SE = 0.011$) and *after* conditions ($M = 0.52$, $SE = 0.012$) (see Fig. 4, light blue line). One sample t-tests comparing correct yes responses to chance (0.5) only produced a significant difference at 80 ms in the *before* condition, $t(15) = 2.57$, $p = 0.021$.

Comparisons - LSF vs. Grayscale

An ANOVA compared LSF and grayscale d' , with main effects of name position, $F(1,60) = 22.32$, $p < 0.001$, $\eta^2 = 0.107$, picture type, $F(1,60) = 183.4$, $p < 0.001$, $\eta^2 = 0.496$, and duration, $F(3,180) = 32.19$, $p < 0.001$, $\eta^2 = 0.267$. There were also significant interactions

between name position and picture type, $F(1,60)=16.73$, $p<0.001$, $\eta G^2=0.082$ and picture type and duration, $F(3,180)=12.71$, $p<0.001$, $\eta G^2=0.126$.

Consistent with the Fast M hypothesis, the difference between grayscale and LSF pictures was reduced at the two shortest durations, suggesting that the initial processing of pictures may be based on coarse magnocellular information. However, detection in the LSF condition was not above chance in the *after* condition. A key claim of the Fast M hypothesis is that the M channel provides preliminary guesses about the stimuli that are fed back through recurrent cortical paths to guide P information. Thus, the inability of LSF images to provide such guidance without advance conceptual information about the target is evidence against the Fast M hypothesis.

Item analysis

In an analysis of which target pictures were detected most frequently across picture types, there was a high degree of correlation between performance on the 128 target pictures in the color and blur conditions ($R^2=0.71$) and in the color and grayscale conditions ($R^2=0.75$). However, the correlation was lower when performance on monochrome pictures was compared with that for blurred pictures ($R^2=0.12$). Performance with the LSF pictures correlated minimally with performance on the grayscale pictures ($R^2=0.03$) compared to the monochrome picture correlation with grayscale ($R^2=0.22$). Thus, not only was performance worse in the monochrome and (even more so) in the LSF condition, but the pictures that elicited accurate detection did not match those in the other conditions very well, perhaps reflecting the unnatural visual characteristics of the monochrome and LSF pictures. Table 3 lists the names of the ten most accurately and least accurately detected for each picture type.

General Discussion

We asked whether modified pictures providing different amounts of color, contrast, and low spatial frequency information would be equivalent to unmodified pictures in early stages of processing, as suggested by the Fast M-pathway hypothesis of Bar (2003) and others (Laycock, Crewther, & Crewther, 2008; Patai, Buckley, & Nobre, 2013; Tapia and Breitmeyer, 2011). To limit processing to early stages, we presented targets in RSVP sequences at durations between 13 and 80 ms per picture. We compared color pictures with blurred versions of the pictures, grayscale pictures with thresholded black-white monochrome pictures, and grayscale pictures with LSF pictures.

The results with normal color pictures in Experiment 1 replicate the findings of Potter et al. (2014): Viewers can detect a target picture among six rapidly presented pictures when cued with a target name either before or after the sequence, and performance is above chance even at the shortest duration of 13 ms. We again found that seeing the name before the sequence led to better detection than seeing it after, and slower presentation rates allowed for better performance. The same pattern was observed for grayscale pictures in Experiment 3, although performance was significantly lower than in the normal color condition. Performance was otherwise highly similar in the two conditions. In particular, there was no suggestion that the color advantage was reduced at short durations.

Bar (2003) hypothesized that magnocellular projections, representing low spatial frequencies, carry visual information very rapidly to the prefrontal cortex, enabling fast top-down cortical feedback and accounting for very rapid processing. The results comparing blurred with color pictures and grayscale with monochrome and LSF pictures give mixed support for this hypothesis. The blurred results are consistent with the Fast M hypothesis in that detection of blurred pictures is not much worse than that of normal color pictures at very brief durations. However, the results for the monochrome and LSF pictures undercut the hypothesis that limited M information can provide guesses about the stimuli that aid analysis of the P information. In the absence of advance conceptual information given by the name, the monochrome and LSF pictures were rarely identified at any duration.

Bar (2009) refers to “the link of a novel input to a similar representation existing in memory” as an analogical process (p. 1236). He suggests that coarse information rapidly activates analogical context in the form of a prediction, followed by confirmation by P information. However, the present results suggest that understanding coarse information, particularly in the monochrome condition, depends heavily on a top-down attentional set, as provided by the target name when given before the picture sequence.

The organization and transfer of information in the M and P pathways may affect a number of neurological disorders that involve vision. For instance, schizophrenic patients have shown they have a visual bias for LSF pictures (Calderone et al., 2013; Laprevote et al., 2013). There may also be M-processing deficits in dyslexia and other neurological conditions (Schiller, Logothetis, & Elliot, 1990; Stein & Walsh, 1997). However, there continues to be controversy over the roles of the M and P paths, whether they can be isolated psychophysically, and whether they interact in the manner that the Fast M hypothesis suggests. For example, the contradictory findings on the function and psychophysical limits of the M path led Goodbourn, Bosten, Hogg, Bargary, Lawrence-Owen, & Mollon (2012) to compare, within participants (N=1060), four different tests that supposedly assess M sensitivity. Their results suggest the different tests produced greater within-subject variance than would be expected if the tasks were acting on the same M neural substrate in a given individual. They recommend abandoning the search for a general M function, particularly in clinical groups. Moreover, the stimuli Goodbourn et al. used were simple sinusoidal gratings and coherent motion patterns. The psychophysics of processing visual scenes is necessarily more involved, making it even more difficult to determine which components of an image, if any, are isolated to the M and P paths.

The present experiments show that reducing visual information by blurring, thresholding, or filtering spatial frequency information causes a considerable reduction in understanding of scenes across presentation durations, especially when viewers have no advance conceptual information about what they may see. Now that we have established baseline sensitivity and recognition abilities across a range of filtered visual conditions, it will be important to test additional conditions, such as a range of low to high spatial frequencies to see if there is a particular range of spatial frequencies that best facilitates performance. In conclusion, while hypotheses that the M-path provides advance information are intriguing, they are only partially supported by the present research. Instead, when no advance information is provided about a target, perception and immediate memory for pictures presented in a rapid

sequence seems to depend on feedforward processing of the full range of spatial frequencies, without feedback from higher to lower levels in the visual system.

Acknowledgments

This research was supported by a National Institutes of Health Grant MH47432.

References

- Bar M. A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*. 2003; 15:600–9. DOI: 10.1162/089892903321662976 [PubMed: 12803970]
- Bar M. The proactive brain: memory for predictions. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*. 2009; 364(1521):1235–1243. [PubMed: 19528004]
- Bar M, Aminoff E, Boshyan J, Fenske M, Gronau N, Kassam K. The contribution of context to visual object recognition. *Journal of Vision*. 2005; 5(8):88–88.
- Bar M, Kassam KS, Ghuman AS, Boshyan J, Schmid AM, Dale AM, et al. Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America*. 2006; 103:449–54. DOI: 10.1073/pnas.0507062103 [PubMed: 16407167]
- Brainard DH. The Psychophysics Toolbox. *Spatial Vision*. 1997; 10:433–436. DOI: 10.1163/156856897X00357 [PubMed: 9176952]
- Bullier J. Integrated model of visual processing. *Brain Research Brain Research Reviews*. 2001; 36(2–3):96–107. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11690606>. [PubMed: 11690606]
- Calderone DJ, Hoptman MJ, Martínez A, Nair-Collins S, Mauro CJ, Bar M, ... Butler PD. Contributions of low and high spatial frequency processing to impaired object recognition circuitry in schizophrenia. *Cerebral Cortex (New York, NY: 1991)*. 2013; 23:1849–58. DOI: 10.1093/cercor/bhs169
- Goodbourn PT, Bosten JM, Hogg RE, Bargary G, Lawrance-Owen AJ, Mollon JD. Do different “magnocellular tasks” probe the same neural substrate? *Proceedings of the Royal Society B: Biological Sciences*. 2012; 279:4263–4271. DOI: 10.1098/rspb.2012.1430 [PubMed: 22896642]
- Grill-Spector K, Malach R. The human visual cortex. *Annual Review of Neuroscience*. 2004; 27:649–77. DOI: 10.1146/annurev.neuro.27.070203.144220
- Hautus MJ. Corrections for extreme proportions and their biasing effects on estimated values of d' . *Behavior Research Methods, Instruments, & Computers*. 1995; 27:46–51. DOI: 10.3758/BF03203619
- Intraub H. Rapid conceptual identification of sequentially presented pictures. *Journal of Experimental Psychology: Human Perception and Performance*. 1981; doi: 10.1037/0096-1523.7.3.604
- Kaplan E, Shapley RM. The primate retina contains two types of ganglion cells, with high and low contrast sensitivity. *Proceedings of the National Academy of Sciences of the United States of America*. 1986; 83(8):2755–7. [PubMed: 3458235]
- Keysers C, Xiao DK, Földiák P, Perrett DI. The speed of sight. *Journal of Cognitive Neuroscience*. 2001; 13(1):90–101. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11224911>. [PubMed: 11224911]
- Kveraga K, Boshyan J, Bar M. Magnocellular projections as the trigger of top-down facilitation in recognition. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*. 2007; 27:13232–40. DOI: 10.1523/JNEUROSCI.3481-07.2007 [PubMed: 18045917]
- Laprevote V, Oliva A, Ternois AS, Schwan R, Thomas P, Boucart M. Low Spatial Frequency Bias in Schizophrenia is Not Face Specific: When the Integration of Coarse and Fine Information Fails. *Frontiers in Psychology*. 2013; 4(May):248.doi: 10.3389/fpsyg.2013.00248 [PubMed: 23653616]
- Laycock R, Crewther DP, Crewther SG. The advantage in being magnocellular: A few more remarks on attention and the magnocellular system. *Neuroscience and Biobehavioral Reviews*. 2008; doi: 10.1016/j.neubiorev.2008.04.008

- Livingstone M, Hubel D. Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science*. 1988; 240(4853):740–749. [PubMed: 3283936]
- Meng M, Potter MC. Detecting and remembering pictures with and without visual noise. *Journal of Vision*. 2008; 8:7.1–10. DOI: 10.1167/8.9.7
- Patai EZ, Buckley A, Nobre AC. Is attention based on spatial contextual memory preferentially guided by low spatial frequency signals? *PloS One*. 2013; 8:e65601.doi: 10.1371/journal.pone.0065601 [PubMed: 23776509]
- Potter MC. Short-term conceptual memory for pictures. *Journal of Experimental Psychology Human Learning and Memory*. 1976; 2:509–22. [PubMed: 1003124]
- Potter MC, Hagmann CE. Banana or fruit? Detection and recognition across categorical levels in RSVP. *Psychonomic Bulletin & Review*. 2014; 22(2):578–585. DOI: 10.3758/s13423-014-0692-4
- Potter MC, Wyble B, Hagmann CE, McCourt E. Detecting meaning in RSVP at 13 ms per picture. *Attention, Perception & Psychophysics*. 2014; 76:270–279.
- Schiller PH, Logothetis NK, Charles ER. Role of the color-opponent and broadband channels in vision. *Visual Neuroscience*. 1990; 5:321–346. DOI: 10.1017/S0952523800000420 [PubMed: 2265148]
- Serre T, Oliva A, Poggio T. A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences of the United States of America*. 2007; 104:6424–9. DOI: 10.1073/pnas.0700622104 [PubMed: 17404214]
- Shapley R, Kaplan E, Soodak R. Spatial summation and contrast sensitivity of X and Y cells in the lateral geniculate nucleus of the macaque. *Nature*. 1981; 292:543–545. DOI: 10.1038/292543a0 [PubMed: 7254350]
- Skottun BC. On using very high temporal frequencies to isolate magnocellular contributions to psychophysical tasks. *Neuropsychologia*. 2013; 51:1556–1560. [PubMed: 23702475]
- Skottun BC, Skoyles JR. Contrast sensitivity and magnocellular functioning in schizophrenia. *Vision research*. 2007; 47:2923–2933. [PubMed: 17825350]
- Skottun BC, Skoyles JR. Coherent motion, magnocellular sensitivity and the causation of dyslexia. *International Journal of Neuroscience*. 2008; 118:185–190. [PubMed: 18041615]
- Stanislaw H, Todorov N. Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc.* 1999; 31:137–49.
- Stein J, Walsh V. To see but not to read; the magnocellular theory of dyslexia. *Trends in Neurosciences*. 1997; 20:147–152. DOI: 10.1016/S0166-2236(96)01005-3 [PubMed: 9106353]
- Sugase Y, Yamane S, Ueno S, Kawano K. Global and fine information coded by single neurons in the temporal visual cortex. *Nature*. 1999; 400(6747):869–873. [PubMed: 10476965]
- Tapia E, Breitmeyer BG. Visual consciousness revisited: magnocellular and parvocellular contributions to conscious and nonconscious vision. *Psychological Science*. 2011; 22:934–42. DOI: 10.1177/0956797611413471 [PubMed: 21697525]

Example target name: "Covered Wagon"
presented *before* or *after* pictures
Time per picture: 80, 53, 27, or 13 ms

Figure 1.
General method of experiments. The dashed border that denotes the target is for illustrative purposes; participants did not see it.

Figure 2.
Examples of target names and their associated pictures across the five picture conditions.

Figure 3. Target detection ability (d') in the five picture conditions with the target name presented *before* and *after* the RSVP sequence. Chance is at 0.0.

Figure 4. Proportion correct in a two-alternative forced choice between two potential target pictures for hits (correct yeses) in the five picture conditions, *before* and *after* the RSVP sequence. Chance is at 0.5.

Table 1

Participant demographics.

Experiment	Before	Mean age	After	Mean age
1: Color	11 F, 5 M	22.1	10 F, 6 M	26.4
2: Blurred	10 F, 6 M	23.1	8 F, 8 M	23.9
3: Grayscale	9 F, 7 M	23.6	9 F, 7 M	20.5
4: Monochrome	12 F, 4 M	23.8	11 F, 5 M	21.1
5: LSF	4 F, 12 M	18.5	7 F, 9 M	19.9

Table 2

Results of target detection across experiments and durations.

	Before		After	
Experiment	<i>Mean d' (SD)</i>	<i>Above chance?</i>	<i>Mean d' (SD)</i>	<i>Above chance?</i>
1: Color	1.686 (0.208)	All durations	0.949 (0.321)	All durations
2: Blurred	1.020 (0.203)	All durations	0.564 (0.231)	All durations
3: Grayscale	1.327 (0.243)	All durations	0.722 (0.306)	All durations
4: Monochrome	0.658 (0.313)	All durations	0.233 (0.182)	27, 80 ms
5: LSF	0.117 (0.247)	80 ms	0.073 (0.296)	None

Table 3

Best and worst 10 target pictures in the five picture conditions. Pictures that were on the list in more than one condition are in bold.

<i>Top 10 (descending)</i>				
<u>Color</u>	<u>Blur</u>	<u>Grayscale</u>	<u>Monochrome</u>	<u>LSF</u>
woman with children	cactus	cheetah	brick arch	metal sculpture
zebras	brick arch	zebras	row of flags	brick arch
closeup of flower	woman with children	children holding hands	bear catching fish	man and mirror
arched passageway	smiling couple	closeup of flower	closeup of flower	arched passageway
brick arch	boxes of vegetables	person holding infant	woman with children	picnic
row of flags	cut up fruit	arched passageway	zebras	bench by water
smiling couple	sliced tomato	flowers on branch	boats in harbor	corner of house
swan	baseball game	brick arch	parked motorcycle	people walking in city
boats in harbor	boats in harbor	picnic	swan	brick building
mansion	arched passageway	smiling couple	mansion	family and dog
<i>Bottom 10 (ascending)</i>				
hummingbird	cat in box	hummingbird	baseball game	tired man
bedroom	bedroom	avocado	turtle	avocado
tired man	chef	meadow with cabin	hummingbird	elephant
cat in box	turtle	boy at computer	man on phone	foreign money
diner countertop	tired man	cat in box	windsurfers	chef
chef	hummingbird	diner countertop	ducks	couple at table
classroom	man on phone	tired man	spices	narrow street
boy at computer	koala	bedroom	hikers	bird flying over water
baboons	hut on beach	spices	sliced tomato	children on steps
man on phone	marsh	classroom	marsh	turtle