Visual Cognition

ISSN: 1350-6285 (Print) 1464-0716 (Online) Journal homepage: http://www.tandfonline.com/loi/pvis20

RSVP at the movies: dynamic images are remembered better than static images when resources are limited

Ayse Candan, James E. Cutting & Jordan E. DeLong

VISUAL COGNITION, 2016
http://dx.doi.org/10.1080/13506285.2016.1159636

Routledge
Taylor & Francis Group

RSVP at the movies: dynamic images are remembered better than static images when resources are limited

Ayse Candan[a], James E. Cutting[a] and Jordan E. DeLong[b]

[a]Department of Psychology, Cornell University, Ithaca, USA; [b]Department of Psychological and Brain Sciences, Indiana University, Bloomington, USA

ABSTRACT
We examined whether dynamic images benefit memory when visual resources are limited. Almost all previous research in this area has used static photographs to examine viewers' memory for image content, description, or visual attributes. Here, we investigated the short-term retention of brief stimuli using rapid serial visual presentation (RSVP) with short videos and static frames of 80, 160, 200, and 400 ms/item. Memory performance for dynamic images was generally better than for comparable still images of the same duration. There was also a strong recency effect for items briefer than 400 ms, which suggests that an optimal duration of about 400 ms may be necessary for dynamic images to be detected and fully processed. Interestingly, we also found that the presence of motion increased performance while the amount of motion did not.

ARTICLE HISTORY Received 13 August 2015; Accepted 23 February 2016

KEYWORDS Motion; movies; natural scenes; RSVP; visual images; visual memory

Most everyday visual scenes are highly complex. Researchers have long been intrigued by the speed and depth of processing of visual information. Previous research has shown that visual scenes can be processed as holistic units and categorized as fast as individual objects (Intraub, 2012; Greene & Oliva, 2009a, 2009b; Konkle, Brady, Alvarez, & Oliva, 2010; Potter, 1976; Potter, Staub, & O'Connor, 2004; Thorpe, Fize, & Marlot, 1996; for a review see Fabre-Thorpe, 2011). But, what happens when visual scenes are presented in a rapid sequence?

Because the visual system has limited capacity, both simultaneous and sequential information competes for resources (Potter, 1976; Potter et al., 2004). Rapid serial visual presentation (RSVP) is a standard paradigm used to measure the limits of visual short-term memory. A typical RSVP trial shows a series of very brief images presented to the viewer without breaks (Potter &

CONTACT Ayse Candan ac885@cornell.edu Department of Psychology, Cornell University, 109 Tower Road, Uris Hall, Ithaca, NY 14853-7601, USA
Supplemental data for this article can be accessed 10.1080/13506285.2016.1159636.

Levy, 1969). Typical images are indoor and outdoor natural scenes involving people, food, and animals (Intraub, 2012; Potter et al., 2004). Researchers have tested two types of memory for RSVP scenes: identification and recognition. They have found that viewers can well identify the presence of a predefined scene (i.e., "businessman at table", "a road with cars") in a sequence where each item is presented as fast as 125 ms (Meng & Potter, 2008). Moreover, recent research showed that such scenes could be detected even when they were presented for only 13 ms (Potter, Wyble, Hagmann, & McCourt, 2014).

Contrary to this rapid identification, later recognition of visual scenes is quite poor. Potter and Levy (1969; Potter, Staub, Rado, & O'Connor, 2002) demonstrated that, although recognition memory improves with increased duration of exposure, it is relatively poor for briefer presentations (100 to 300 ms). Potter (1976) proposed that brief pictures are held in a conceptual store, called CSTM, which helps encoding and retrieval through activation of related concepts. This type of facilitation appears to rely on top-down influences; a critical interval of about 300 ms may be required to process images at a meaningful level and consolidate them into memory (Potter, 1976). Potter et al. (2004) also showed that subjects falsely recognized conceptually similar items. This again suggests that what gets registered in memory during that brief period might be a general description, or a summary, of an image but not its particular characteristics. Nonetheless, it is not clear how the nature of visual stimuli affects later memory and which stages of processing result in poor recognition.

The present study compares dynamic and static natural scenes in an RSVP task. Our question is: Are dynamic images remembered better than comparable still images? To answer this query, we presented viewers with brief sequences of short clips and of frames taken from Hollywood movies. In our study, we focused on recognition memory to examine the effects of exposure on later memory as well as the retention interval. Although previous research has indicated that visual scenes are remembered with high accuracy when presented each for a few seconds (Konkle et al., 2010; Potter & Levy, 1969; Standing, 1973), relatively few studies have looked at the retention of dynamic stimuli (see Buratto, Matthews, & Lamberts, 2009; Ferguson, 2014; Matthews, Benjamin, & Osborne, 2007).

For example, Matthews et al. (2007) showed that, when viewers watched either a series of 3-s film clips with no cuts or a series of static frames from those clips, their retention was better for dynamic images as opposed to static images after a week or even a month. In a more recent study, Ferguson (2014) looked at memory for dynamic images using 5-min movie clips. A forced-choice test revealed that viewers were better at identifying the old frames compared to foils from another part of the movie displaying the same characters in the same setting. Intriguingly, a follow-up experiment showed that memory performance was worse for the same test frames if the presentation stimuli were in a static sequence rather than a dynamic one. Better memory was attributed to longer exposure and richer information in the case of dynamic stimuli.

However, these studies used dynamic scenes that lasted several seconds to several minutes. Insofar as we know no previous study has looked at retention of dynamic naturalistic images at brief exposures in an RSVP paradigm. Again, previous RSVP research has almost exclusively used static photographs of visual scenes to examine people's memory (Potter et al., 2004; Intraub, 2012; Potter et al., 2014). As previous studies have attributed the memory boost for dynamic images to the richness of information (Ferguson, 2014; Matthews et al., 2007), dynamic images may be processed differently, and potentially in a more adaptive manner. Since we do not process the world around us as a series of static sequences, dynamic stimuli would be expected to be more complex and conceptually rich. Therefore, in this study, we expected to find better memory performance for dynamic natural scenes compared to similar static ones.

Dynamic images might yield better recognition both by directing attentional resources due to motion onset (Abrams & Christ, 2003) and by activating conceptual memory, which may facilitate encoding and retrieval due to contextual facilitation and binding (Biederman, 1972; Chun & Jiang, 1998). The RSVP paradigm can also help us understand better the time course of processing meaningful dynamic images as well as providing valuable insight into the limits and efficiency of visual information processing.

Method

We employed an RSVP paradigm to present brief sequences presenting either still frames or short dynamic clips. The experiment followed a 4×2 design with four between-subject conditions (presentation length: 400, 800, 1200, and 2000 ms) and two within-subjects conditions (stimulus type: static and dynamic images).

Materials

Sixty-two Hollywood movies were chosen randomly from the sample of 150 films originally analyzed by Cutting, DeLong, and Nothelfer (2010). Those were released between 1935 and 2005 and reflected a variety of genres and aspect ratios, as suggested in Figure 1. A complete list of movie sources used here for stimuli can be found in the Supplemental data. These movies were originally coded frame-by-frame to determine the location and type of each cut, dissolve, or fade. These frame numbers were then used to sample short clips of 2, 4, 5 or 10 consecutive frames from the middle of a random shot within each movie. This procedure avoided the visual disruptions

Figure 1. Sample still frames as stimuli, and their movie sources and aspect ratios (image width/height).

of cuts within a stimulus item. Dynamic stimuli were presented at ~24 frames/s, or 40 ms/frame. Matched static frames were taken from the beginning of each short clip. Movie clips and static frames displayed a wide range of cinematic situations and actions, involving people, animals, and objects.

A MATLAB script using the Psychophysics Toolbox (Brainard, 1997) chose and prepared stimuli and also collected the viewer's responses. Stimuli were displayed on an Apple MacBook Pro laptop running on OS X Mavericks, with 13.3 inch (33.7 cm) diagonal LED-backlit glossy widescreen display, positioned at a distance of approximately 20 inches (50 cm). The screen had the resolution of 1280 × 800 pixels, with a refresh rate of 60 Hz.

Subjects and procedure

A total of 79 undergraduate students from Cornell University participated in an IRB-approved set of studies for course credit. Nine other subjects were excluded due to computer error (5), experimenter error (2), or familiarity with some of the stimuli (1).[1] Before starting, the participants were given detailed instruction about what to expect, and told to respond as quickly and accurately as possible to each test item. Viewers were allowed to take breaks between but not within trial blocks.

Viewers took part in one of four experimental conditions (variations in presentation duration), and completed a total of 45 trials. Depending on condition, the trial sequences lasted 400, 800, 1000, or 2000 ms. Each item in the five-item sequence was displayed for 80, 160, 200, or 400 ms (2, 4, 5 and 10 frames for dynamic stimuli), respectively. In this manner, each static frame in a sequence was shown for the same total duration as each video clip for each presentation duration. In a between-subjects design, 16 viewers participated in the 80 ms condition, 19 in the 160 ms condition, 19 in the 200 ms condition, and 16 in the 400 ms condition.

On each trial, participants were presented with a sequence of either five static frames or five short video clips. Each trial was randomly selected to be either a still or a dynamic sequence, with stimulus type varying across successive trials. Each sequence was presented without interruption in black and white and without audio. Trial sequences were assembled from five randomly chosen shots out of a single, randomly chosen movie. Again, the first frame of each video clip was used as the corresponding item in the static sequence. A schematic trial is suggested in Figure 2. Each trial began with a 200-ms fixation cross at mid-screen, followed immediately by the five-item sequence. Static frames and movie clips were displayed in their original aspect ratio, and items within a sequence had no interval between them. Immediately after the fifth static image or last frame of the fifth clip, a white noise mask was displayed to impede visual persistence of the last item.

Each stimulus sequence was then followed by a 10-item test sequence cued by a 1000 ms display of: "Have you seen the following clips [images]?" Viewers then saw a randomly ordered array of five old items and five distracters. While the old items were the same items presented in the initial trial sequence, distractors were chosen randomly from different shots within the same movie. Each test stimulus was presented in the same format (static or dynamic) and for the same duration the original stimulus was presented in the initial trial sequence. For example, if the participant was presented with a dynamic sequence with items of 400 ms duration, each test stimulus (old

[1]One subject recognized a family member, who had been an actor, in the clips and images of one of the movies. Although this viewer's data were no different than that of other viewers, we decided that exclusion was the best policy.

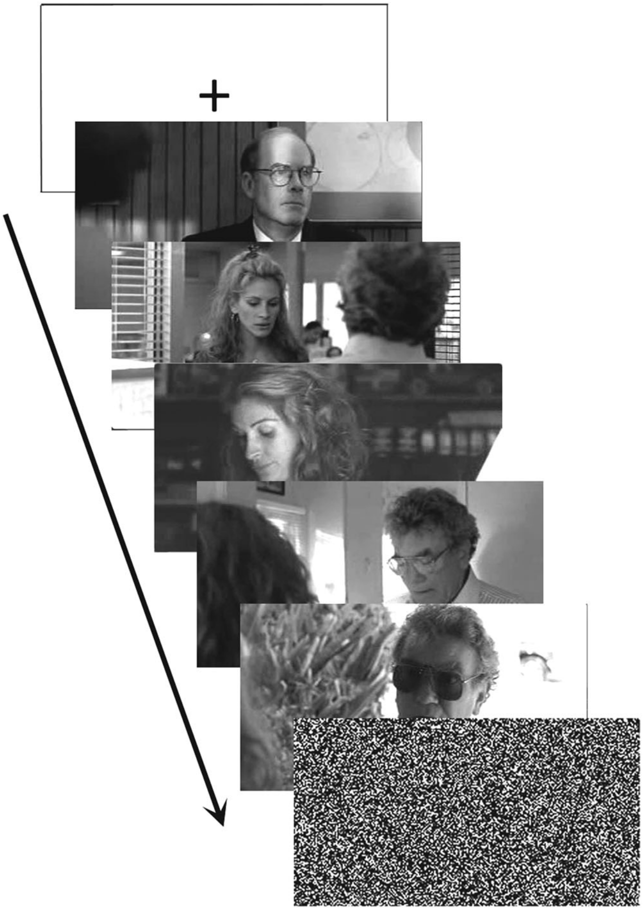

Figure 2. A sample static stimulus sequence with frames from *Erin Brockovich* (2000). Each trial started with a fixation cross in the middle of the screen displayed for 200 ms followed by the five-item stimulus sequence, followed by white noise patch. There were four experimental conditions with trial sequences lasting 400, 800, 1000, or 2000 ms, and each item played or displayed for 80, 160, 200, or 400 ms respectively. Movie sequences were 2, 4, 5, and 10 frames of a film and each sequence came from a randomly chosen shot out of a randomly chosen movie from our sample. Following each stimulus sequence there was a test sequence (not shown). Participants responded yes or no to 10 test items, five that were old and from the stimulus sequence and five that were new and from another part of the same movie.

or new) also lasted for 400 ms and was a video clip. The same applied to still frames. Distracters were chosen randomly from different shots within the same movie. Participants were instructed to respond by pressing the "Y" (for yes) and "N" (for no) keys on the computer keyboard to indicate whether they had seen the item in the stimulus sequence or not. Pressing the key advanced the test to the next item. Once all the test items were presented, a command appeared: “Press the ‘Space’ key to advance to next trial.” The experiment lasted about 30 min.

Data analysis

A MATLAB script recorded the type of stimuli (static or dynamic) presented on each trial as well as the viewer’s responses. Serial positions for stimuli and test items were also recorded and later analyzed. Memory performance was calculated and expressed as d'.[2] Ancillary properties such as whole-frame and whole-clip luminance and visual activity measures were also measured (Cutting, Brunick, DeLong, Iricinschi, & Candan, 2011). The measured luminance was the average 8-bit, non-gamma transformed brightness of all pixels for each item (each pixel measured from 0 to 255). The visual activity index (VAI) was the interframe correlation of pixel luminance values between the first and the last frame of a short clip. This can be taken as a measure of the amount of visual change from combined actor and camera movement. Here, VAI could range from −1.0 to 1.0 with larger and more positive numbers reflecting less motion.

Results

Unsurprisingly, memory performance increased with longer exposure durations ($t(692) = 19.80$, $p < .0001$, $d = 1.5$), as seen in Figure 3. More importantly, overall memory performance for dynamic images was superior to that for static frames presented at the same durations ($t(692) = 3.47$, $p = .0006$, $d = .26$), with reliable results for stimuli at 400 ms ($F(1, 158) = 5.9$, $p = .036$) and 200 ms ($F(1,188) = 4.083$, $p = .045$), although not at 160 ms ($F(1,188) = 0.54$, $p = .66$) and 80 ms ($F(1,158) = 2.9$, $p = .09$).

There was also a reliable effect of serial position among the five-item stimuli, showing a significant recency effect for presentation times shorter than 400 ms/item ($t(692) = 7.37$, $p < .0001$, $d = .56$) but not for those of 400 ms/item, as shown in Figure 4. Consequently, there was a significant interaction of presentation duration and serial position ($t(692) = -5.07$, $p < .0001$, $d = .39$). The order of stimuli in the 10-item test sequence also showed a reliable effect as shown in Figure 5, indicating increased forgetting for later items in both static and dynamic conditions regardless of stimulus duration ($F(9,45170) = 17.0$, $p = .0001$). This result is likely due to the constraints on short-term retention of brief visual stimuli and has been found in previous RSVP research (Potter et al., 2002).

[2]Previous research using RSVP has also used guessing corrections like A′ or other high threshold guessing corrections (Potter et al., 2004) in which they re-expressed the performance in corrected percentages. When we employed the same guessing correction used by Potter et al. (2004) we obtained similar results. For the results here, we will report the d' results.

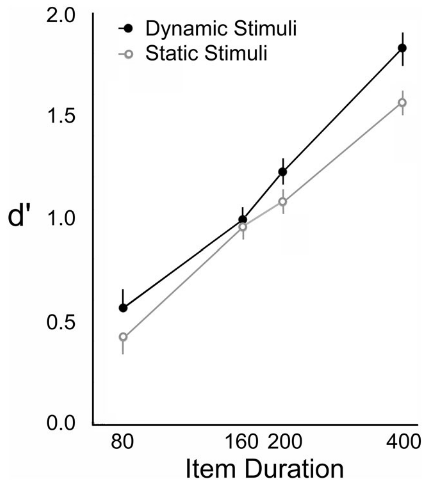

Figure 3. Overall performance (*d′*) for all presentation lengths (80 ms, 160 ms, 200 ms, and 400 ms/stimulus) with respect to stimulus types (static vs. dynamic). Whiskers indicate 95% confidence intervals.

Among factors that may have contributed to these results, we examined first the amount of visual activity in dynamic images (Cutting et al., 2011). Interestingly, VAI did not predict memory performance across durations ($t(345) = 0.31$, $p = 0.76$, $d = .03$). Luminance, on the other hand, showed a modest effect on memory for both static and dynamic images ($t(665) = 2.33$, $p = 0.02$, $d = .18$), with brighter stimuli garnering better performance.

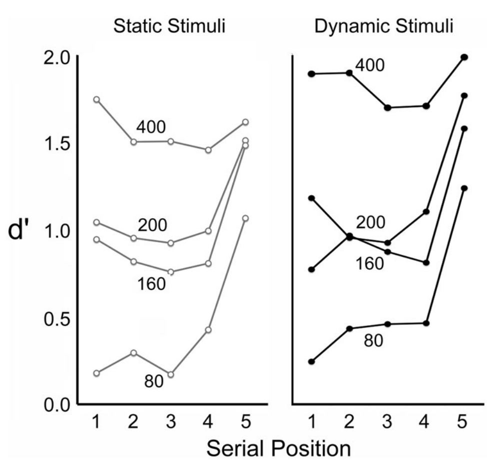

Figure 4. Memory performance for stimuli with respect to serial position in the presentation sequence in static (left panel) and dynamic (right panel) stimuli. The central whisker indicates an average 95% confidence interval for all data points.

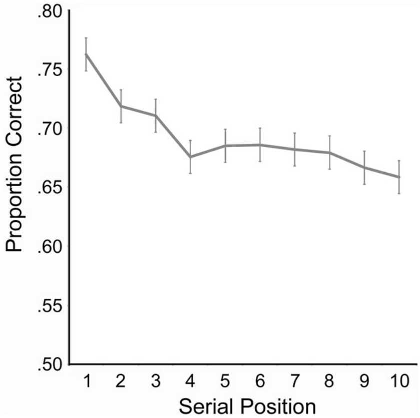

Figure 5. Memory performance across the test sequence for all duration conditions and both static and dynamic stimuli. Whiskers indicate 95% confidence intervals.

Finally, we found no significant effect on memory for either the aspect ratio or whether the original film was in black and white or in colour.

Discussion

Results of the present study show that memory performance for dynamic visual scenes is better, but not much better, than matched static scenes presented for the same duration. This result also suggests that motion is a reasonably salient cue for memory when one has limited exposure to a visual stimulus. This result has a number of possible explanations. For example, motion may increase attentional focus therefore enhance encoding. Abrams and Christ (2003), for example, have shown that motion onset is a strong cue in mobilizing attentional resources even when the motion is not informative. Moreover, Franconeri and Simons (2003, 2005) found that certain dynamic events capture attention even when the onset is absent, suggesting that motion itself may be enough to attract attention. In the present study, the observed memory boost may be the result of heightened attention in the case of dynamic images. Also, these dynamic scenes may be more naturalistic hence providing richer information in conceptually relevant contexts. As one can argue that dynamic images may provide a burden on memory when resources are limited, the nature of the information may be more valuable, which could compensate and even favour the dynamic stimuli.

The results also suggest that there may be a critical interval (300–400 ms) needed to process and encode motion. This observation is in line with the proposed CSTM store proposed by Potter (1976), which requires around 300 ms for a visual image to be consolidated into memory. If true, this argues for an interplay between the demands of conceptual memory and constraints of dynamic stimulus processing. Also, while the overall results indicate a significant effect for motion on memory, detailed analyses for each presentation length showed a significant effect for stimuli with longer durations (400 and 200 ms) and not for stimuli with shorter durations (160 and 80 ms). This result may not be surprising since by default shorter stimuli have less motion (fewer frames).

From our perspective, the novel result of the study was that the mere presence of motion, but not its amount, was the driving factor behind the slight memory boost for the dynamic images. This surprised us, and suggests that the processing of dynamic scenes is not simply the sum of motion information. Instead, dynamic stimuli offer a qualitatively different experience that differentiates itself from that of its static counterparts. As dynamic stimuli are more complex yet more naturalistic to our everyday visual processing, studying how we process these stimuli under limited resources can give us a better understanding of the limits of our visual system.

The present study also expands on the previous research by focusing on memory for dynamic images under time constraints. While limited research exists that showed that dynamic scenes were remembered better than static ones, even after an interval (Ferguson, 2014; Matthews et al., 2007), those studies only used longer presentation durations, ranging from a few seconds to several minutes. The present study uniquely contributes to the literature by showing that this memory boost effect extends to very brief presentation durations (200 to 400 ms). This indicates that we can efficiently encode and retain dynamic information even when we have limited access to it. As dynamic scenes provide richer and more naturalistic information compared to static scenes, they also appear to be more resilient to limitations due exposure duration. This suggests that we may be better equipped to process dynamic visual scenes, possibly due to the nature of information provided by these stimuli.

Moreover, while the results of the present study suggest an optimal interval of at least 400 ms for a dynamic image to be retained successfully in short-term memory, the mechanisms requiring this interval are unclear. One approach to disentangle presentation duration from consolidation time could be to present non-meaningful visual masks (Potter, 1976) between items of shorter duration, decorrelating stimulus duration and inter-item onsets. Another potential question pertains to the nature of information encoded from dynamic images. As this study only focused on recognition memory but not identification memory (Meng & Potter, 2008; Potter et al., 2014), we don't know whether dynamic images would provide a similar memory boost for detecting dynamic images in RSVP or whether we would observe a stronger effect. Understanding these would clarify further to the mechanisms behind the encoding dynamic images.

Overall, the present study provides insight into our processing of dynamic visual images at brief presentation rates. Studying how people remember such images can provide further information about the limits and efficiency of visual information processing and reveal the dynamics of the encoding and retrieval of complex naturalistic visual stimuli. Insofar as we are aware, no previous research has investigated how we encode and remember dynamic visual scenes presented for such brief durations. Not irrelevant here is the overwhelming increasing presence and increasing pace (Cutting & Candan, 2015; Cutting et al., 2010) of film in visual media. Indeed, it seems possible that our ability to extract visual information from moving displays could improve as societal exposure increases for members of our culture.

Acknowledgments

Ayse Candan and James Cutting are members of the Psychology Department at Cornell University. Jordan DeLong is a Visiting Assistant Professor in the Department of Psychological and Brain Sciences at Indiana University. We thank David James Field, Barbara Finlay, Khena Swallow, Catalina Iricinschi, Kaitlin Brunick, James Golden, Kedar Vilankar and Erin Isbilen for valuable comments. We also thank research assistants Alison Gusick, Cynthia Vella, and Maamie Asamoah-Mensah for their contribution in running participants.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Abrams, R. A., & Christ, S. E. (2003). Motion onset captures attention. *Psychological Science*, *14*(5), 427–432. doi: 10.1111/1467-9280.01458

Biederman, I. (1972, 7 July). Perceiving real-world scenes. *Science*, *177*, 77–80. Retrieved from http://www.jstor.org/stable/1733939

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436. doi: 10.1163/156856897X00357

Buratto, L. G., Matthews, W. J., & Lamberts, K. (2009). When are moving images remembered better? Study–test congruence and the dynamic superiority effect. *The Quarterly Journal of Experimental Psychology*, *62*(10), 1896–1903. doi: 10.1080/17470210902883263

Chun, M. M. & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*, 28–71. doi:10.1006/cogp.1998.0681

Cutting, J. E., Brunick, K. L, DeLong, J. E., Iricinschi, C., & Candan, A. (2011). Quicker, faster, darker: Changes in Hollywood film over 75 years. *i-Perception*, *2*, 569–576. doi:10.1068/i0441aap

Cutting, J. E. & Candan, A. (2015). Shot durations, shot classes, and the increased pace of popular movies. *Projections: The Journal for Movies and Mind*, *9*(2), 40–62(23). Retrieved from http://dx.doi.org/10.3167/proj.2015.090204

Cutting, J. E., DeLong, J. E., & Nothelfer, C. E. (2010). Attention and the evolution of Hollywood film. *Psychological Science*, *21*, 432–439. doi: 10.1177/0956797610361679
Fabre-Thorpe, M. (2011). The characteristics and limits of rapid visual categorization. *Frontiers in Psychology*, *2*, 1–12. doi: 10.3389/fpsyg.2011.00243
Ferguson, R. (2014). *Visual recognition for dynamic scenes* (Ph.D. Dissertation)., Arizona State University. Retrieved from http://repository.asu.edu/attachments/135015/content/Ferguson_asu_0010E_13954.pdf
Franconeri, S.L., & Simons, D.J. (2003). Moving and looming stimuli capture attention. *Perception & Psychophysics*, *65*, 999–1010. doi: 10.3758/BF03194829
Franconeri, S.L., & Simons, D.J. (2005). The dynamic events that capture visual attention: A reply to Abrams and Christ (2005). *Perception & Psychophysics*, *67*(6), 962–966. doi: 10.3758/BF03193623
Greene, M. R. & Oliva, A. (2009a). The briefest of glances: the time course of natural scene understanding. *Psychological Science*, *20*(4), 464–472. doi: 10.1111/j.1467-9280.2009.02316.x
Greene, M. R. & Oliva, A. (2009b). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology*, *58*(2), 137–176. doi: 10.1016/j.cogpsych.2008.06.001
Intraub, H. (2012). Rethinking visual scene perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, *3*, 117–127. doi: 10.1002/wcs.149
Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Scene memory is more detailed thank you think: The role of categories in visual long-term memory. *Psychological Science*, *21*(11), 1551–1556. doi: 10.1177/0956797610385359
Matthews, W. J., Benjamin, C., & Osborne, C. (2007). Memory for moving and static images. *Psychonomic Bulletin & Review*, *14*, 989–993. doi: 10.3758/BF03194133
Meng, M. & Potter, M. C. (2008). Detecting and remembering pictures with and without visual noise. *Journal of Vision*, *8*(9):7, 1–10. doi: 10.1167/8.9.7
Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 509–522. doi: 10.1037/0278-7393.2.5.509
Potter, M. C., & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology*, *81*, 10–15. doi: 10.1037/h0027470
Potter, M. C., Staub, A., & O'Connor, D. H. (2004). Pictorial and conceptual representation of glimpsed pictures. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 478–489. doi: 10.1037/0096-1523.30.3.478
Potter, M. C., Staub, A., Rado, J. & O'Connor, D. H. (2002). The time course of competition for attention: Attention is initially labile. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(5), 1149–1162. doi: 10.1037/0096-1523.30.3.478
Potter, M.C., Wyble, B., Hagmann, C.E., & McCourt, E.S. (2014). Detecting meaning in RSVP at 13 ms per picture. *Attention, Perception, & Psychophysics*, *76*(2), 270–279.
Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, *25*, 207–222. doi:10.1080/14640747308400340
Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*, 520–522. doi: 10.1038/381520a0