Massachusetts Institute of Technology

Enhanced recognition of memorable pictures in ultra-fast RSVP

Nico Broers, University of Münster, Germany. n.broers@wwu.de

Mary C. Potter, Massachusetts Institute of Technology, USA. mpotter@mit.edu

Mark R. Nieuwenstein, University of Groningen, The Netherlands.

m.r.nieuwenstein@rug.nl

Word Count:

Abstract: 172 (word limit is 250)

Body of the paper: 3241 words.

Abstract

Long-term recognition memory for some pictures is consistently better than for others (Isola, Xiao, Parikh, Torralba, & Oliva, 2014). Here, we investigated whether pictures found to be memorable in a long-term memory test are also perceived more easily when presented in ultra-rapid RSVP. Participants viewed 6 pictures they had never seen before that were presented for 13 to 360 ms per picture in a rapid serial visual presentation (RSVP) sequence. In half the trials, one of the pictures was a memorable or a non-memorable picture and perception of this picture was probed by a visual recognition test at the end of the sequence. Recognition for pictures from the memorable set was higher than for those from the non-memorable set and this difference increased with increasing duration. Non-memorable picture recognition was low initially, did not increase until 120 ms, and never caught up with memorable picture recognition performance. Thus, the long-term memorability of an image is associated with initial perceptibility: a picture that is hard to grasp quickly is hard to remember later.

Introduction

Our understanding and memory of natural scenes is seemingly effortless and limitless (Brady, Konkle, & Alvarez, 2008; Hollingworth, 2004; Potter, 1976; Standing, 1973; Thorpe, Fize, & Marlot, 1996). Recently however, researchers have shown that scenes vary dramatically in the likelihood that they will be remembered (Isola, Xiao, Parikh, Torralba, & Oliva., 2014; see also, Bainbridge, Isola, & Oliva, 2013; Bylinskii, Isola, Bainbridge, Torralba, & Oliva, 2015; Khosla, Raju, Torralba, & Oliva, 2015). These studies showed that after 2.4 seconds of uninterrupted encoding time, some visual scenes are more likely to be remembered than others, and this difference in memorability was furthermore found to be consistent across observers and retention intervals ranging from 36 seconds to 40 minutes, thus suggesting that long-term memorability is an intrinsic property of a visual scene.

Thus far, research on long-term picture memorability has focused primarily on examining which properties of a picture might be predictive of its long-term memorability. In addressing this matter, it has been found that people's judgments about whether a picture is interesting, memorable, or aesthetic are negatively – not positively – correlated with actual memorability, thus making clear that these intuitive judgments are very poor predictors of memorability (Isola et al., 2014; Khosla et al., 2015). In another approach, researchers have used computer vision algorithms to quantify various low and high-level image statistics and to correlate them with memorability. This computational work has shown that memorability is not associated with low-level properties such as the mean or variance in hue, saturation, or luminance, or the area occupied by objects and the number of distinguishable objects present in a picture (Isola et al., 2014). Instead, high-level properties such as scene category and scene descriptors did combine to produce a rank correlation of predicted

and empirical memorability of .54, and this correlation was based on the combined predictive value of many scene descriptors that produce either positive or negative effects on memorability. Specifically, Isola et al. concluded that "people, interiors, foregrounds, and human-scale objects tend to contribute positively to memorability" whereas "exteriors, wide-angle vistas, backgrounds, and natural scenes tend to contribute negatively to memorability (p. 10)". In a similar vein, Khosla et al. concluded that the predicted memorability of a picture decreases across pictures showing "people, busy images [lots of gradients], specific objects, buildings, and finally open scenes" (legend Figure 5 in Khosla et al.). Taken together, these results of modeling picture memorability make clear that while pictures of people are generally more memorable than pictures of landscapes, there also exists considerable variability in the contents of memorable and non-memorable pictures.

The Current Study

While previous studies have thus sought to identify correlates of long-term picture memorability in image statistics and people's intuitive judgments, the aim of the current study was to explore the perceptual correlates of picture memorability. Are memorable pictures more quickly understood than less memorable pictures, or do they actually take longer to understand? Using rapid serial visual presentation (RSVP), Potter (1976; see also Intraub, 1980) found that visual recognition of pictures shown in RSVP increases from near chance at 100 ms to more than 90% accuracy at 1 second per picture, thus suggesting that the difference between memorable and non-memorable pictures might be determined by processing that only begins after the first 100 ms and builds up over the first second. In addition, the results of Potter and colleagues showed that when memory was tested with a name (e.g., "two people eating"), performance after a presentation duration of 167 ms was nearly as good as

when tested with the picture itself (Potter, Staub, & O'Connor, 2004), showing that conceptual information is available from brief presentations of pictures. Perhaps most dramatically, recent studies (e.g., Potter, Wyble, Hagmann, & McCourt, 2014) have shown that conceptual and visual recognition is well above chance after presentations as brief as 13 ms in a short RSVP sequence, thus suggesting that the influence of long-term picture memorability might already be seen after only 13 ms of uninterrupted processing time.

In the current study, we examined whether and how long-term memorability influences the initial encoding of a visual scene. To this end, we compared immediate visual recognition performance for highly memorable and less-memorable scenes (from Isola et al., 2014) that were embedded in an RSVP of six scenes, all new to the viewer on every trial. By varying the presentation duration from only 13 ms to 360 ms, we aimed to establish at which duration an advantage for immediate recognition of memorable pictures would arise, thus shedding light on the rate at which the features underlying image memorability can be extracted. Thus, the current study aimed to determine whether the difference between memorable and less memorable pictures is associated with visual features extracted early in visual processing, or whether the difference only emerges when the pictures can be processed and inspected for several hundred milliseconds (e.g., Isola et al, 2014; Khosla et al., 2015).

Method

Our selection of memorable and non-memorable pictures was drawn from the stimulus set used by Isola et al. (2014), which includes 2222 pictures rank-ordered in terms of their long-term memorability, based on 665 participants. From this set, we selected the 136 most and the 136 least memorable pictures for use as targets and

recognition foils in the RSVP task (see examples in Figure 1). The memorable pictures included many pictures of humans, indoor scenes and human-scale objects, whereas the non-memorable pictures included more pictures of outdoor scenes and large spaced interiors, as noted by Isola et al. (2014) and Khosla et al. (2015). Filler pictures were taken from the remaining pictures of intermediate memorability. Every picture was new to the participants, and no picture was repeated in the experiment.
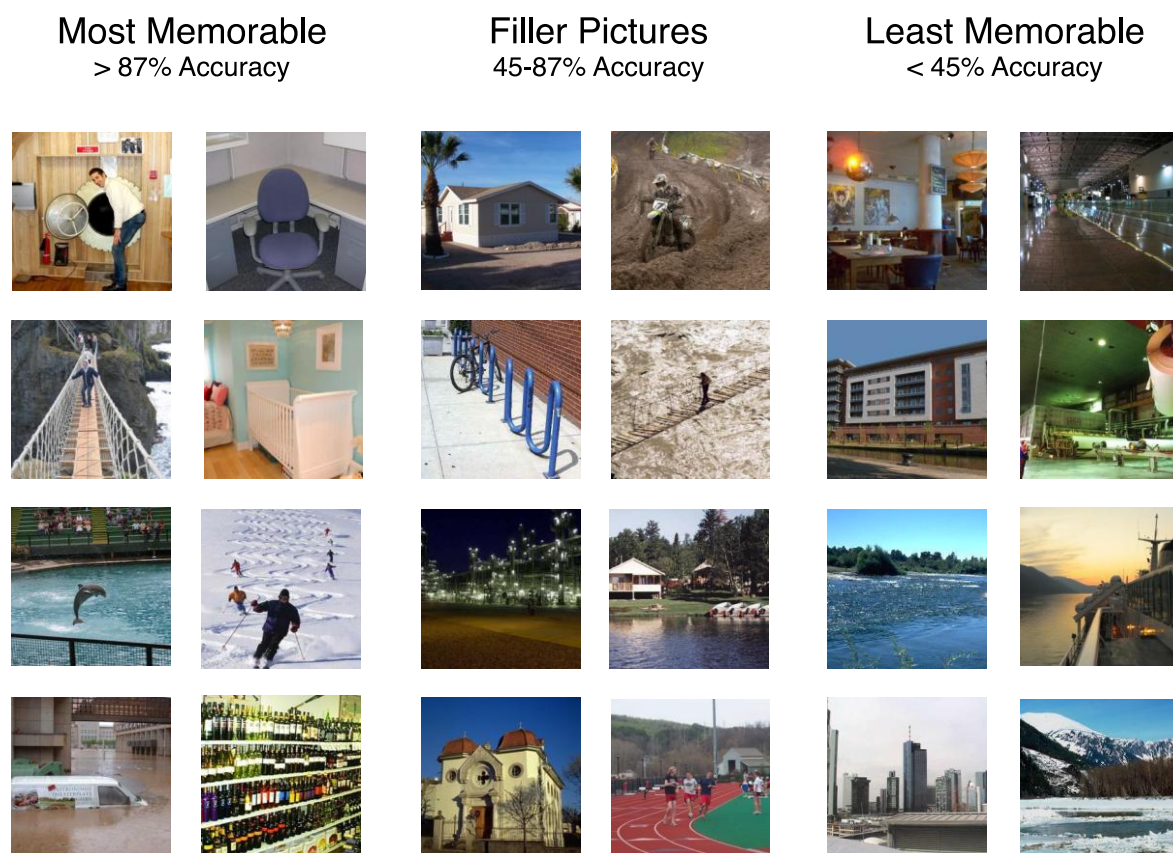
Figure 1. Examples of the pictures used in the current study. The pictures labeled as "most memorable" produced a hit rate higher than 87% in Isola et al.'s study (2014), whereas the pictures labeled "least memorable" are pictures that yielded a hit rate lower than 45%.

Each trial consisted of a sequence of six pictures, followed by a test picture. On target-present trials, the test picture matched one of the pictures of the sequence and this picture could be a memorable or a non-memorable picture. On target-absent

trials, all six pictures in the sequence were fillers, and the non-matching test picture was equally likely to be drawn from the memorable or non-memorable set. The duration of each picture in the sequence was 13, 27, 53, or 80 ms in Experiment 1 and 80, 120, 200, or 360 ms in Experiment 2.

Participants

The participants in the study were 32 volunteers from the MIT community who were paid $12 for their participation. All signed a consent form approved by the MIT Committee on the Use of Humans as Experimental Subjects. Sixteen participants (age range 18-29, 11 female) were in Experiment 1, and the other 16 participants (age range 18-26, 10 female) were in Experiment 2. The number of participants was chosen to allow for counterbalancing of which pictures occurred across four durations and four serial positions. The decision on the number of participants was made based on previous research using a similar task (e.g., Potter et al., 2014) showing that 16 was sufficient for robust above-chance recognition of pictures shown in ultra-rapid RSVP. The number of participants in each experiment was thus decided upon in advance.

Materials and apparatus

The presentation of stimuli and the registration of responses were controlled using MATLAB 2012A, the Psychophysics Toolbox Extension (Brainard, 1997). Timing precision was controlled by using Wyble's Stream package for MATLAB. We checked the timing on the RSVP sequence of each trial and excluded trials in which a timing error of + /- 3 ms or greater affected the target picture. The experiment was run on a Mac Mini with a 2.4 GHz Intel Core 2 Duo processor and a 17-inch CRT monitor, with a 1024 x 768 resolution and a 75-Hz refresh rate.

As described above, the stimuli were selected from the 2222 color photographs used by Isola et al (2014) that were rank-ordered by memorability.

Design

Both experiments had a 2 x 2 x 4 within-subjects design, with the factors being *Memorability* (high versus low), *Target* (present versus absent), and *Duration* (13, 27, 53, 80 ms for Experiment 1; 80, 120, 200, and 320 ms for Experiment 2). On target-present trials the target picture was equally often presented at the second, third, fourth or fifth position in the RSVP sequence. The selection of target pictures was controlled in such a way that the same picture was used equally often in each temporal position and in each duration condition across our 16 participants. The different trial types were randomly intermixed across four blocks of 64 trials, with the constraint that every block included 32 target and 32 non-target trials. The number of trials was chosen to keep the total length of the experiment short enough to maintain the attention of the participants (256 trials plus 16 practice trials; data from practice trials were excluded from statistical analysis).

Procedure

As illustrated in Figure 2, each trial started with a fixation cross that was shown for 500 ms. After the fixation cross, there was a 200-ms blank interval before the presentation of the RSVP stream. For Experiment 1, each picture was shown for just 13, 27, 53, or 80 ms, and for Experiment 2, for 80, 120, 200, or 360 ms. After the presentation of the RSVP stream there was a 200-ms blank interval before the appearance of the test image, which was presented for 400 ms followed by a blank. Participants indicated whether the test picture matched a picture in the sequence, using the "Y" and "N" keys of the keyboard.

Method
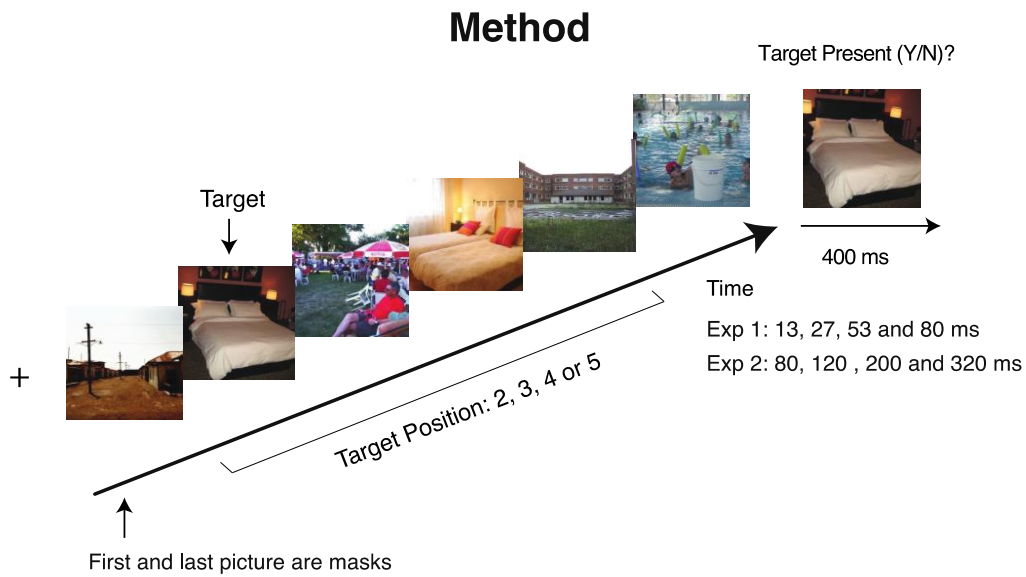
Figure 2. Illustration of the events in a trial. The test picture shown at the end of the sequence matched the target picture in 50% of the trials. On non-matching trials, target-absent trials, the test picture was drawn from the same set (memorable or non-memorable) as the target picture.

Data Analysis

Recognition accuracy was calculated using the measure of *d'* (Stanislaw & Todorov, 1999) which was computed separately for each subject for each duration and memorability condition, based on 16 target-present and 16 target-absent trials in each of the eight conditions. That is, the false-alarm and hit rates were calculated separately for memorable and non-memorable test pictures at each duration, to control for potential differences in response bias to the two sets of pictures. We also carried out separate repeated measures ANOVAs for Experiments 1 and 2 on hit and false alarm rates, with duration and target memorability as within-subject factors: these results are reported in the Appendix. For all results from the RM-ANOVA, we calculated the generalized eta-squared $- \eta_G^2 -$ to provide a measure of effect size (Bakeman, 2005). In interpreting this measure of effect size, the common guidelines are that an $\eta_G^2 = .02$ is considered a small effect, whereas an $\eta_G^2 = .13$ is considered a

medium-sized effect, and $\eta_G^2 = .26$ is considered a large effect (Bakeman, 2005;

Cohen, 1988). For pair-wise *t*-tests, we report the confidence intervals of the mean

difference, and we calculated the effect sizes in Cohen's *d* (Cohen, 1988), taking into

account the correlations between the paired observations (Morris & DeShon, 2002).

We excluded five trials in Experiment 1 and two trials in Experiment 2 due to

timing errors in the presentation of the pictures.

Results

In analyzing the data, we ran two separate repeated measures ANOVAs for

Experiments 1 and 2, using *d'* as the measure of recognition performance, and using

duration (13, 27,53, or 80 ms for Experiment 1, 80, 120, 200, or 360 ms for

Experiment 2) and target memorability as within-subject factors.

As illustrated in Figure 3, the results for Experiment 1 showed significant

main effects of memorability, $F(1, 15) = 123.27$, $p < .001$, $\eta_G^2 = .38$, and duration,

$F(3, 45) = 4.06$, $p = .012$, $\eta_G^2 = .09$, as well as a significant interaction of these

factors, $F(3, 45) = 6.13$, $p = .001$, $\eta_G^2 = .09$. The effect of duration was significant for

memorable pictures, $F(3, 45) = 8.01$, $p < .001$, $\eta_G^2 = .24$, and best characterized as a

linear relationship, $F(1, 15) = 14.89$, $p = .002$, $\eta_G^2 = .30$. For the non-memorable

pictures, the effect of duration was non-significant, $F(3, 45) = 0.12$, $p = .95$, $\eta_G^2 = $

.005. To examine at which durations the difference between memorable and non-

memorable pictures was significant, we conducted pairwise comparisons with an

alpha set at .05/4 = .0125. These tests showed that memorable pictures were

recognized significantly better than non-memorable pictures at all durations, with

$t(15) = 3.08$, $p = .008$, *95% CI$_{MeanDifference}$* = [.13, .73], *Cohen's d = .77*, $t(15) = 5.00$, *p*

< .001, *95% CI$_{MeanDifference}$* = [.35, .86], *Cohen's d = 1.25*, $t(15) = 6.02$, *p < .001*, *95%*

CI$_{MeanDifference}$ = [.67, 1.41], *Cohen's d = 1.56*, and $t(15) = 7.10$, *p < .001*, *95%*

$CI_{MeanDifference}$ = [.89, 1.65], *Cohen's d* = 1.88 for the durations of 13, 27, 53, and 80 ms, respectively.
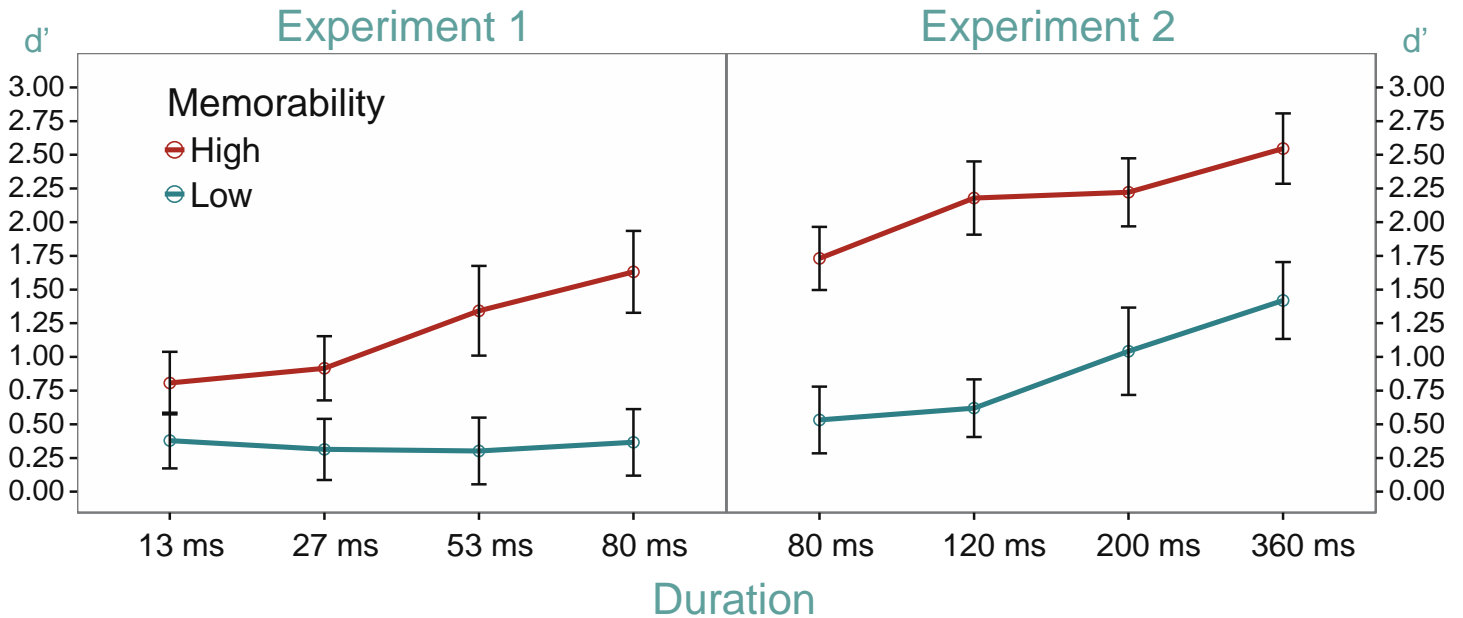
Figure 3. Results Experiments 1 and 2. Recognition accuracy (*d'*) is shown as a function of picture duration and target memorability. Error bars denote 95% confidence intervals around the mean (Morey, 2008).

The analysis of the results for Experiment 2 showed significant main effects of target memorability, $F(1, 15) = 178.64$, $p < .001$, $\eta_G^2 = .57$, and duration, $F(3, 45) = 14.27$, $p < .001$, $\eta_G^2 = .25$, but no interaction, $F(3, 45) = 1.28$, $p > .25$, $\eta_G^2 = .02$. Both memorable and non-memorable pictures showed a significant effect of duration, $F(3, 45) = 8.18$, $p < .001$, $\eta_G^2 = .21$, and $F(3, 45) = 10.57$, $p < .001$, $\eta_G^2 = .31$, respectively, and this effect was best fit by a linear contrast for both memorable and non-memorable pictures, $F(1, 15) = 21.92$, $p < .001$, $\eta_G^2 = .27$, and $F(1, 15) = 25.41$, $p < .001$, $\eta_G^2 = .41$, respectively. Comparisons of performance for memorable and non-memorable pictures showed that the difference was significant at all durations, with $t(15) = 11.27$, $p < .001$, *95% $CI_{MeanDifference}$* = [.97, 1.43], *Cohen's d* = 2.85, $t(15) = 8.58$, $p < .001$, *95% $CI_{MeanDifference}$* = [1.17, 1.95], *Cohen's d* = 2.19, $t(15) = 5.63$, $p <$

.001, *95% CI$_{MeanDifference}$* = [.73, 1.63], *Cohen's d* = 1.41, and *t*(15) = 5.71, *p* < .001,

95% CI$_{MeanDifference}$ = [.71, 1.55], *Cohen's d* = 1.44 for the durations of 80, 120, 200,

and 360 ms, respectively.

Discussion

In the current study, we explored whether pictures that show better long-

term retention after 2.4 seconds of encoding time are also perceived more readily

when shown in an RSVP sequence for very brief durations. Replicating the results of

earlier work (Potter et al., 2002a; Potter et al., 2014), our findings show that both

memorable and non-memorable pictures could be recognized at above-chance levels

even after being shown for only 13 ms. Crucially, pictures found to be memorable in

the earlier long-term memory study by Isola and colleagues (2014) were perceived

more readily in RSVP at durations of 13 ms and higher. Performance for memorable

pictures improved markedly with increasing duration, whereas performance for the

non-memorable pictures only started to improve beyond durations above 120 ms and

never caught up with performance for the memorable pictures.

An important implication of the current findings stems from the fact that an

extremely brief, pre- and post-masked exposure in ultra-rapid RSVP precludes the

opportunity for deep encoding and reflection on the contents of a picture. Specifically,

at rates as rapid as 13 or 27 ms per picture, the only information available for

processing is the information that can be extracted from feedforward activation, as

these rapid rates do not allow for enough time for feedback signals to produce

reentrant processing at such short intervals (e.g., Di Lollo, Enns, & Rensink, 2000;

Enns & Di Lollo, 2000; Lamme, 2007). By implication, our finding that memorable

pictures were more easily recognized than non-memorable pictures at a duration of 13

ms per picture can be taken to suggest that picture memorability is associated with

features that are extracted very early in visual processing. Since Potter et al. (2014) have shown that conceptual and visual recognition is approximately equally good with presentations as brief as 13 ms in RSVP, our findings can be said to corroborate feedforward models of scene recognition (Fukushima, 1980; Serre, Oliva, & Poggio, 2007) which assume that a conceptual understanding of a visual scene can be achieved without feedback. According to this view, the memorability of a picture may be determined by or at least correlated with the level and depth of understanding reached in the initial sweep. We hypothesize that if the features of the picture are not adequate to activate a coherent and rich conceptual interpretation in the first pass (before the masking effect of the following stimulus), processing may be stalled at a low level. With longer presentation times, however, feedback loops (e.g., Lamme, 2007) will eventually lead to some level of encoding, even of the least memorable pictures, and will refine understanding of the more memorable pictures.

Our results suggest that the features underlying long-term memorability of pictures are already extracted during the first 13 to 120 ms of processing. What does this imply about longer-term memory for pictures? We know from other work (Potter, 1976; Potter, Staub, Rado, & O'Connor, 2002b) that pictures will not be retained in longer term memory unless they have an average of about 500 ms of uninterrupted processing time before the next picture. Intraub (1980) showed, however, that the picture does not need to remain in view during that time: performance is almost as good when the picture is only shown for about 100 ms, followed by a blank interval of 400 ms. Apparently, long-term memory is preferably based on first-pass understanding of the picture, without requiring eye movements to take in further information (contrary to what some have suggested, for example, Konkle et al (2010) p. 1555). Once the picture is understood, longer-term memory seems only to require a

time-dependent process of consolidation of what has already been understood and evaluated in the first glimpse.

        In short, the present results suggest that there is a strong link between the speed of understanding a picture and the likelihood of remembering it. What you are more likely to remember, you may also be more likely to see.

Author's Note

References

Bainbridge, W. A., Isola, P., & Oliva, A. (2013). The intrinsic memorability of face photographs. *Journal of Experimental Psychology: General*, *142*(4), 1323–1334. http://doi.org/10.1037/a0033872

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, *37*, 379-384. http://dx.doi.org/10.3758/BF03192707

Brady, T. F., Konkle, T., & Alvarez, G. A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, *105*(38), 14325–14329. http://doi.org/10.1073/pnas.0803390105

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. http://doi.org/10.1163/156856897x00357

Bylinskii, Z., Isola, P., Bainbridge, C., Torralba, A., & Oliva, A. (2015). Intrinsic and extrinsic effects on image memorability. *Vision Research*. http://doi.org/10.1016/j.visres.2015.03.005

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Di Lollo, V., Enns, J.T., & Rensink, R.A. (2000). Competition for consciousness among visual events: The psychophysics of reentrant visual processes. *Journal of Experimental Psychology: General*, *129*, 481-507.

Enns, J.T., & Di Lollo, V. (2000). What's new in visual masking? *Trends in Cognitive Sciences*, *4*, 345-352.

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, *36*(4), 193–202. http://doi.org/10.1007/BF00344251

Hollingworth, A. (2004). Constructing Visual Representations of Natural Scenes: The Roles of

Short- and Long-Term Visual Memory. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(3), 519–537. http://doi.org/10.1037/0096-1523.30.3.519

Isola, P., Xiao, J., Parikh, D., Torralba, A., & Oliva, A. (2014). What makes a photograph memorable? *Pattern Analysis and Machine Intelligence (PAMI)*.

Khosla, A., Raju, A. S.,Torralba, A., & Oliva, A. (2015). Understanding and predicting image memorability at a large scale.

*https://people.csail.mit.edu/**khosla**/papers/iccv**2015_khosla**.pdf*

Lamme, V. A. F. (2006). Towards a true neural stance on consciousness. *Trends in Cognitive Sciences*, *10*(11), 494–501. http://doi.org/10.1016/j.tics.2006.09.001

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology*, *4*(2), 61–64.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, *7*, 105-125.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning & Memory*, *2*(5), 509–522. http://doi.org/10.1037/0278-7393.2.5.509

Potter, M. C., Staub, A., Rado, J., & O'Connor, D. H. (2002a). Recognition memory for briefly presented pictures: The time course of rapid forgetting. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(5), 1163–1175. http://doi.org/10.1037//0096-1523.28.5.1163

Potter, M. C., Staub, A., Rado, J., & O'Connor, D. H. (2002b). Recognition memory for briefly presented pictures: The time course of rapid forgetting. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(5), 1163–1175. http://doi.org/10.1037//0096-1523.28.5.1163

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. http://doi.org/10.3758/s13414-013-0605-z

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences*, *104*(15), 6424–6429. http://doi.org/10.1073/pnas.0700622104

Standing, L. (1973). Learning 10000 pictures. *Quarterly Journal of Experimental Psychology*, *25*(2), 207–222. http://doi.org/10.1080/14640747308400340

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers : a Journal of the Psychonomic Society, Inc*, *31*(1), 137–149.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6582), 520–522. http://doi.org/10.1038/381520a0

Appendix
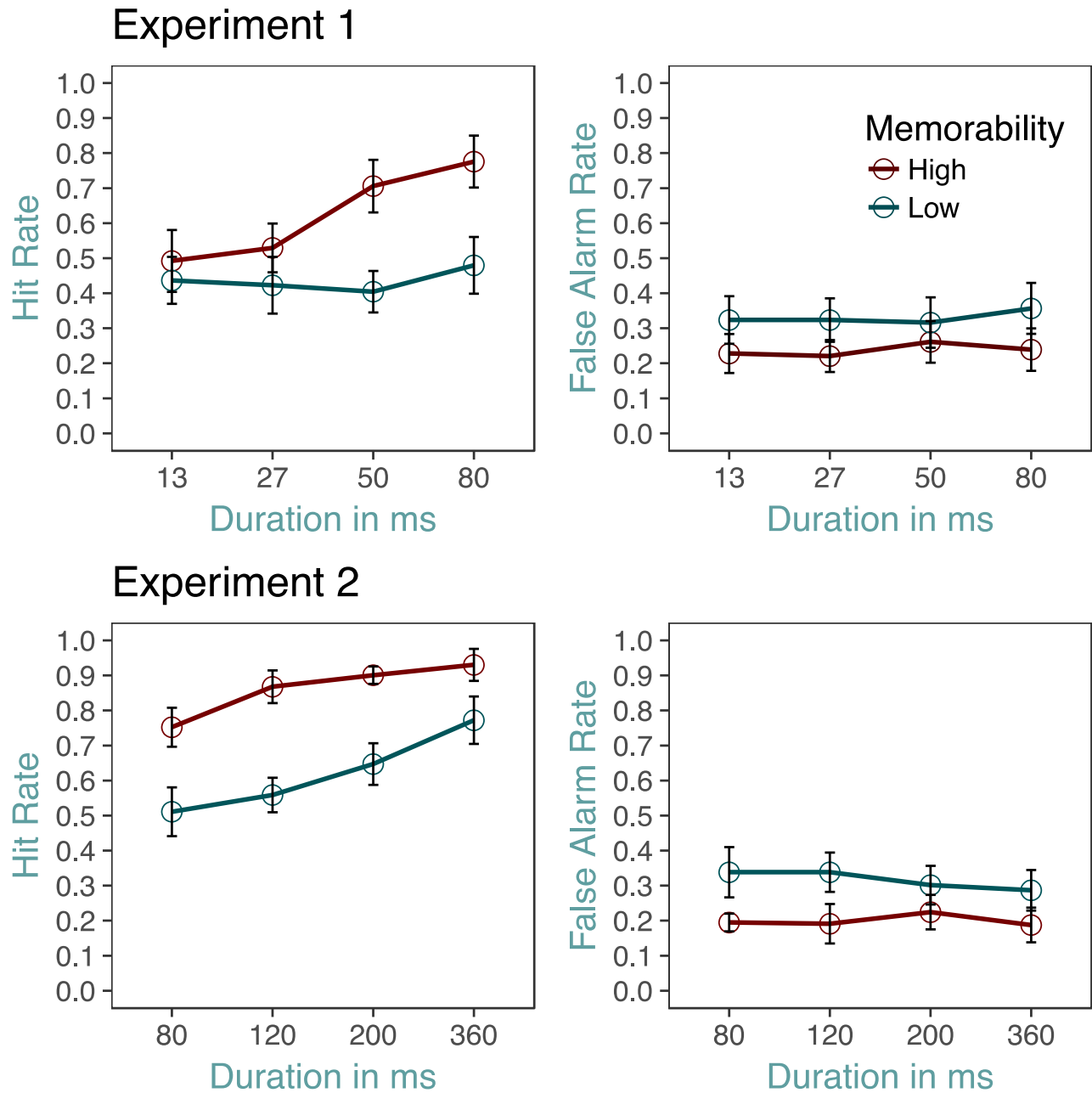
Figure 4. Proportion of Hits and False Alarms in Experiments 1 and 2, as a function of picture duration and target

memorability. Error bars denote 95% confidence intervals around the mean (Morey, 2008). The interaction

between duration and memorability are found only in changes in the hit rate (Experiment 1: $F(1, 45) = 7.67$ $p <$

.001, $\eta_G^2 = .11$, Experiment 2: $F(1, 45) = 7.67$ $p < .001$, $\eta_G^2 = .11$) not the false alarm rate (Experiment 1: $F(1, 45)$

19

= 7.67 p = .61, η_G^2 = .01; Experiment 2: F(1, 45) = .86 p = .47, η_G^2 = .01). The effects of memorability are present both in the hit rate (Experiment 1: F(1, 15) = 36.42 p < .001, η_G^2 = .27, Experiment 2: F(1, 15) = 113.63, p < .001, η_G^2 = .49), and (in reverse) in the false alarm rate (Experiment 1: F(1, 15) = 11.18 p = .004, η_G^2 = .08, Experiment 2: F(1, 15) = 30.43 p < .001, η_G^2 = .16).