

Effects of varying presentation time on long-term recognition memory for scenes: Verbatim and gist representations

Fahad N. Ahmad¹ · Morris Moscovitch² · William E. Hockley¹

Published online: 17 November 2016
© Psychonomic Society, Inc. 2016

Abstract Konkle, Brady, Alvarez and Oliva (*Psychological Science*, 21, 1551–1556, 2010) showed that participants have an exceptional long-term memory (LTM) for photographs of scenes. We examined to what extent participants' exceptional LTM for scenes is determined by presentation time during encoding. In addition, at retrieval, we varied the nature of the lures in a forced-choice recognition task so that they resembled the target in gist (i.e., global or categorical) information, but were distinct in verbatim information (e.g., an “old” beach scene and a similar “new” beach scene; exemplar condition) or vice versa (e.g., a beach scene and a new scene from a novel category; novel condition). In Experiment 1, half of the list of scenes was presented for 1 s, whereas the other half was presented for 4 s. We found lower performance for shorter study presentation time in the exemplar test condition and similar performance for both study presentation times in the novel test condition. In Experiment 2, participants showed similar performance in an exemplar test for which the lure was of a different category but a category that was used at study. In Experiment 3, when presentation time was lowered to 500 ms, recognition accuracy was reduced in both novel and exemplar test conditions. A less detailed memorial representation of the studied scene containing more gist (i.e., meaning) than verbatim (i.e., surface or perceptual details) information is retrieved from LTM after a short compared to a long

study presentation time. We conclude that our findings support fuzzy-trace theory.

Keywords Scene recognition · Global information · Local information · Novel test · Exemplar test · Two-alternative forced-choice recognition · Fuzzy-trace theory

Several researchers have found evidence for young adults showing an exceptional memory for pictures compared to non-pictorial stimuli such as words (Brady, Konkle, Alvarez, and Oliva 2008; Hockley 2008; Hockley and Bancroft 2011; Konkle, Brady, Alvarez, and Oliva 2010a, b; Potter 1976; Standing, Conezio, and Haber 1970; Standing 1973; Wichmann, Sharpe, and Gegenfurtner 2002). In a landmark study by Standing et al. (1970), 2,500 photographs of pictures of human beings, animals, mechanical objects, and city scenes were randomly taken from news and travel magazines and presented to participants for 10 s each. Participants' recognition was tested in a two-alternative forced-choice test in which an old test picture was paired with a new picture selected randomly from the original set of collected pictures. Participants showed exceptional discrimination performance (85 %), even 3 days after study. Moreover, reducing presentation time to 1 s or reversing the orientation of target stimuli did not affect performance significantly. Standing et al. concluded that the long-term memory (LTM) capacity for pictures is exceptionally high relative to non-pictorial stimuli.

As Konkle et al. (2010b) note, however, this immense LTM capacity for pictures found by Standing et al. (1970) could have been due to conceptual distinctiveness of the pictures allowing high performance to be achieved with sparse representations of the images in which only basic-level category information was stored. In addition, Standing et al. (1970) did not examine performance by the type of forced-choice test.

✉ Fahad N. Ahmad
ahma2720@mylaurier.ca

¹ Department of Psychology, Wilfrid Laurier University, 75 University Avenue W, Waterloo, ON N2L 3C5, Canada

² Department of Psychology, University of Toronto, Toronto, Ontario, Canada

Rather than compare forced-choice recognition for random pictures as Standing et al. (1970) did, one can co-vary the nature of the similarity between the test alternatives to gain a better appreciation of the encoded representations of the pictures. Konkle et al. (2010b) varied the number of scene exemplars in order to determine the effect of interference on scene recognition. During the study phase, participants viewed 2,912 images from 128 different scene categories, with, 1, 4, 16, or 64 exemplars presented per category. The photographs of scenes were presented for 3 s each. In the forced-choice recognition test that followed, there were two conditions. In the *exemplar test condition*, a target scene was paired with a lure that was not presented during the study phase, but was from the same category. In the *novel test condition*, the target scene was paired with a new scene from a category not present during the study phase. Konkle et al. found that participants' memory for scenes was very high (i.e., 96 %) in the novel test condition. Importantly, participants' memory for scenes was high (i.e., 76 %) even with 64 scene exemplars per category in the exemplar test condition. Konkle et al. concluded memory representations of scenes are of high fidelity. Participants were able to distinguish between scene exemplars even when the number of exemplars increased, but performance still decreased in the exemplar condition indicating that memory for meaning does play a role and the pictures are not of the highest fidelity.

According to Konkle et al. (2010b), the first glance of a scene provides a statistical summary representation of the global structure of the scene and a basic-level category is formed. The category knowledge then guides the observer to the details of the scene to distinguish it from other scenes. Accumulation of visual details of the scene occurs over time. With enough pre-existing knowledge (i.e., basic-level schemas) and time of presentation, diagnostic visual details are accumulated and stored in long-term memory. Thus, the key finding from Konkle et al.'s (2010b) study was that performance was above chance in the exemplar test condition, even when the number of studied exemplars was increased, indicating that humans have an exceptional LTM capacity for the details of visual scenes. Brady et al. (2008) and Konkle et al. (2010a) demonstrated that this exceptional LTM capacity is also observed in memory for visual objects.

Andermane and Bowers (2015) hypothesized that if the test delay was increased to 1 week, the exceptional LTM capacity for visual details shown by participants in the Brady et al. (2008) study would be significantly reduced. They based their prediction on fuzzy-trace theory (Brainerd and Reyna 2002, 2004). According to fuzzy-trace theory, memory performance is dependent on the retrieval of both verbatim and gist traces. Verbatim traces are episodically instantiated representations of surface forms of experienced items, including contextual cues, whereas gist traces are episodic representations of concepts (meanings, relations, patterns) that have been retrieved as a

result of encoding items' surface forms. Thus, gist representations represent information at the level of general senses and meaning, whereas verbatim representations reflect the specific surface features of the item. Andermane and Bowers defined gist-like visual memories as those that contained information about the basic level category of the studied visual object. Gist-like visual memories would be beneficial for recognition in the novel test but not in the exemplar test. In contrast, they defined detailed visual memories as those that enabled distinguishing a beach scene with a perceptually similar but different beach scene, as in an exemplar test. With a longer retention interval, Andermane and Bowers predicted verbatim information would be lost at a greater rate than gist memory as this would follow with other studies that had reported faster forgetting for detailed visual compared to gist memories (e.g., Lampinen et al. 2001; Tuckey and Brewer 2003). Moreover, according to fuzzy-trace theory, detailed (or verbatim) memories are forgotten more quickly than gist memories. Andermane and Bowers noted that the faster forgetting for detailed information is how fuzzy-trace theory accommodates the increased rate of false memories over time in many contexts.

As expected for the short study-test delay, participants showed exceptional LTM for pictures of objects, replicating the Brady et al. (2008) and Konkle et al. (2010a) studies. However, following a delay of 1 week, recognition accuracy was reduced in both the exemplar and novel test conditions. In fact, there were comparable reductions when the foils were categorically similar and dissimilar. Andermane and Bowers (2015) suggested the reason that they found similar forgetting rates for detailed and gist memories of studied objects after a 1-week delay was that the task instructions emphasized studying each image in detail. That is, the task instructions in their study emphasized studying each image carefully, since highly similar foils would be presented at test.

In contrast, in past studies encoding of gist information was favored. For example, some past studies assessed memory for less-detailed perceptual information such as verbal material and other past studies that assessed memory for visual images presented in the context of meaningful scenes that could support gist-like inferences and expectations (e.g., Lampinen et al. 2001; Tuckey and Brewer 2003). In addition, the inclusion of the exemplar test favored the retrieval of detailed memories. Previous studies that had compared forgetting rates for gist and detailed visual memories had used different encoding and retrieval contexts such as verbal cued-recall tests (e.g., Tuckey and Brewer 2003), or different views of objects at study and test (e.g., Lampinen et al. 2001). Thus, the reason past studies found that detailed memories decayed faster than gist memories was that they included study and test conditions that favored the role of gist, whereas in Andermane and Bowers' study the test and study conditions favored the role of verbatim information. Andermane and Bowers concluded

that their finding of gist and detailed visual memory decaying at similar rates was inconsistent with the assumption of fuzzy trace theory that detailed memories decay more quickly than gist memories.

Andermane and Bowers (2015) showed lengthening the retention interval reduced the exceptional LTM capacity found by Brady et al. (2008) and Konkle et al. (2010a) for details of studied objects. However, reducing presentation time of scenes could also reduce exceptional LTM for studied scenes. What type of scene information is available at retrieval after a short compared to a long presentation time? An explanation of how presentation time could affect retrieval of the type of scene from LTM can be provided by taking fuzzy-trace theory into consideration. Relevant to the current study, according to the first principle of fuzzy-trace theory stated by Brainerd and Reyna (2002), participants process the surface form (verbatim) and the meaning content of experiences (gist) in parallel, storing dissociated representations of verbatim and gist traces, respectively. However, the effect of presentation time on storage and retrieval of verbatim and gist information from LTM has not been examined.

Although such research has not been undertaken, there is evidence of the detrimental effects of reducing presentation time on short-term memory (STM) for scenes. Researchers have shown that as the duration of presentation decreases recognition performance also decreases (Castelhano et al. 2009; Intraub 1997; Irwin and Zelinsky 2002; Melcher 2001 2006, Potter 1976, 1992, Tatler et al. 2003). In studies testing STM for scenes, one or four scenes would be presented one at a time, followed immediately by a yes/no recognition test. Potter (1976) showed as the duration of each picture in a continuous sequence is reduced from 2 s to 125 ms, STM performance dropped significantly. In addition, most natural images require multiple fixations to encode a complex picture fully due to the small portion of the image that falls on the fovea at any given time (Intraub 1997), and the number of visual fixations is correlated with memory for the visual scenes (Irwin and Zelinsky 2002).

Importantly, researchers have found there is a differential effect of varying presentation time on retrieval of gist and detailed information in a STM task (Melcher 2006; Tatler et al. 2003). In a study conducted by Tatler et al. (2003), presentation time of scenes was varied randomly from 1 s to 10 s and questions on gist and detail were immediately asked after presentation of the scene. Tatler et al. found the proportion correct for questions on gist information was approximately 88 %, but for questions on different types of verbatim information for objects in the scene (i.e., presence, shape, color, position, relative distance) proportion correct was approximately 60 %. Viewing time did not influence performance in gist questions that tested the identity of the scene, but there was a linear increase in performance in verbatim questions with an increase in presentation time. Based on their

findings, Tatler et al. concluded that an early STM representation of a scene consists of mostly gist information and with prolonged viewing a more detailed representation of a scene is formed. Other researchers have provided further support for this finding by showing when only objects are presented from studied scenes at test, recognition increases linearly as viewing time of scenes increases up to 20 s (Castelhano, Mack, and Henderson 2009; Melcher 2006). Surprisingly, no researcher has examined the effect of varying presentation time on long-term recognition memory for scenes.

It is important to examine the effect of varying scene presentation time on long-term recognition for scenes for two main reasons. Firstly, it needs to be determined to what extent presentation time during encoding determines participants' exceptional LTM for scenes. Secondly, examining the effect of varying study presentation time on LTM for scenes provides a further test of fuzzy-trace theory. In other words, how does varying scene presentation time affect retrieval of gist and verbatim information as measured by the forced-choice novel test for the former and the exemplar test for the latter? Although one can infer from the STM studies that verbatim memory will be affected by changes in presentation time, it is not known whether the relative preservation of gist is maintained in LTM or whether gist is also affected. Moreover, we do not know whether the time needed to encode verbatim memory in STM also suffices to encode that memory in LTM.

In the current study, we combined a manipulation of encoding time with two types of forced-choice recognition tasks in order to examine how scenes are encoded and represented in LTM. We manipulated the lures in the forced-choice test in order to determine how degree of processing mediated by study presentation time (i.e., 500 ms, 1 s, 4 s) influenced the type of scene representation retrieved from LTM. In the novel test, the lure is of a category different from the target and, as a result, participants can rely on gist memory of the target to correctly discriminate the lure from the target. In contrast, for the exemplar test, the lure is of the same category as the target and, as consequence, participants must rely on verbatim memory of the target to correctly discriminate the lure from the target. In addition, two additional measures, confidence ratings and response times, were collected to provide support for our predictions. In two separate experiments, we compared the recognition memory of participants in novel and exemplar test for scenes presented for 1 s or 500 ms to scenes presented for 4 s. We were specifically interested in the information retrieved at test by participants for the different scene presentation times.

In Experiment 1, we took 1 s as the short study presentation time since short-term recognition memory for scenes is preserved for scenes presented for 1 s (Castelhano and Henderson 2005; Potter 1976; Potter et al. 2002). In contrast, for testing long-term recognition memory, 3 s or 4 s has been used as a

presentation time for scenes during the encoding phase (e.g., Konkle et al. 2010b; Standing et al. 1970). As a result, we chose 4 s as the long study presentation time. We predicted for the 1-s presentation time participants would still show exceptional recognition performance in the novel test condition, but there would be a significant decline in performance in the exemplar test condition. This prediction is based on the assumption that, although gist and verbatim processing begin in parallel, the verbatim processing of detailed scenes would take longer than the processing of gist. Therefore, less verbatim information would be encoded after 1 s compared to 4 s of presentation time. In contrast, sufficient gist information would be encoded in 1 s to support discrimination in the novel test condition. As a consequence, with a 1-s presentation time, participants would show reduced performance in the exemplar test, but still show exceptional performance in the novel test.

Experiment 1

In Experiment 1, we examined if presenting scenes for a short 1-s study presentation time, as compared to a long 4-s study presentation time, would result in differential effects on long-term recognition memory performance in the exemplar and novel test conditions. We measured accuracy, response time, and confidence of recognition memory for the scenes. We predicted that participants would show lower performance in the exemplar test condition for scenes presented for short compared to long study presentation times. This prediction was based on the assumption that with a short presentation time of 1 s, participants would encode more gist than detailed verbatim information (cf. Castelano et al. 2009; Melcher 2006; Tatler et al. 2003). However, for a longer presentation time of 4 s, there would be enough time to encode sufficient gist and verbatim information of the scene. We predicted similar performance for the short and long presentation time in the novel test condition, because even at the short duration sufficient gist information would be encoded to identify the general category of the scene, and to use that information to distinguish the target from a lure in the novel test condition. We also predicted confidence would be higher for verbatim- than gist-based decisions, because retrieval of perceptual details would elicit more confidence in the accuracy of the memory. That is, confidence in correct responses in the exemplar test condition would be higher than in the novel test condition as correct details of the studied item can be retrieved. Confidence would be lower, however, for incorrect responses in the exemplar test condition compared to the novel test condition as the participant could not retrieve correct details of the studied item. We also predicted faster response time in the novel than exemplar test condition for correct responses, because according to fuzzy trace theory (see Brainerd and Reyna 2002, for

review) gist information is retrieved faster than verbatim information.

Method

Participants All participants in each experiment were undergraduate students enrolled in an introductory psychology course at the University of Toronto who participated for course credit. Twenty students participated in Experiment 1. All institutional ethics review board procedures were followed and consent was received from all participants in each experiment.

Materials The experiment was run on two PC-compatible laboratory computers equipped with 17-in. LCD color monitors. Eprime 2.0 software was used to control stimulus presentation and response recording. A total of 320 exemplars from 24 categories were taken from the database of photographs of scenes developed by Xiao, Hays, Ehinger, Oliva, and Torralba (2010). This was the same database Konkle et al. (2010b) used for the images in their study. There were eight exemplars from 16 categories (eight indoor: bedroom, church, classroom, office, dinner room, conference room, hair salon, empty room, and eight outdoor: airport, beach, bridge, tent, castle, cemetery, house, playground) making a total of 128 scenes presented during the study phase. Four exemplars per category were presented in the short presentation time condition and the other four exemplars per category were presented in the long presentation condition.

For both exemplar and novel test conditions, an indoor scene was paired with an indoor scene and the same for outdoor scenes. For the test phase, an additional eight scenes per category from the 16 categories were presented as exemplars in the exemplar test. Finally, a further eight scenes per category from eight additional categories (four indoor: bar, bathroom, gymnasium, lobby, and four outdoor: barn, temple, tennis court, swimming pool) were presented as lures in the novel test condition. As a result, the only difference between the exemplar and novel tests was that in the novel test condition, the lure was a new scene from a novel category.

Four sets of scenes were created for the purpose of counterbalancing. The scenes presented as targets in the exemplar and novel test conditions were counterbalanced across participants. In addition, the scenes in the short and long study presentation blocks were also counterbalanced across participants.

Procedure At the beginning of the study phase, participants were told they would be presented with scenes each of which would be presented for a short or long presentation time, after which there would be a recognition test for the scenes. Shortly

after providing study instructions to the participant, the study phase began. 128 scenes were presented with an inter-stimulus interval of 800 ms. There was a different presentation order of the scenes for each participant. During the study phase, presentation time of scenes was blocked in order to minimize rehearsal borrowing between presentations. That is, half of the scenes were presented for 4 s each, whereas the other half were presented for 1 s each. The presentation order of the short and long presentation time blocks was counterbalanced over participants.

The scenes were presented one at a time at the center of the screen (subtending $7.5^\circ \times 7.5^\circ$ visual angle) and alternated with a fixation cross presented for 800 ms. Immediately after the study phase, the researcher provided the participant with test instructions for 1 m, indicating which keys to press on the key board and to provide a confidence rating (1 – “not sure”, 2 – “sure,” and 3 – “very sure”) shortly after their old response. During the test phase, 128 scene pairs were presented in a random order for each participant. For the short presentation time condition, 32 scene pairs in the novel test condition and 32 scene pairs in the exemplar test condition were presented. This was the same for the long presentation time condition. The target scene was presented with the appropriate type of lure. Both images of the target and lure were of the same size and were presented side by side on the computer screen. For both the exemplar test and the novel test conditions, half of the target images for each category were placed on the left side of the screen, whereas the other half of the target images of the same category was placed on the right side of the screen. Figure 1 displays examples of the stimuli and test conditions for Experiment 1. Accuracy, response time, and confidence rating were recorded.

Results

Accuracy The mean proportions of correct old responses (hits) and incorrect old responses (false alarms) in the exemplar and novel test conditions for short and long presentation times are presented in Table 1. A 2 (Presentation time: 1 s vs. 4 s) \times 2 (Condition: Exemplar vs. Novel) repeated analysis of variance (ANOVA) was conducted on proportion of correct responses. There was a main effect of presentation time, $F(1, 19) = 12.7$, $MSE = .004$, $p < .05$, $\eta^2 = .401$. Higher accuracy was shown for scenes presented for 4 s compared to 1 s. There was a main effect of test condition, $F(1, 19) = 63.7$, $MSE = .006$, $p < .001$, $\eta^2 = .770$. Higher accuracy was found in the novel than in the exemplar test condition. The mean difference scores (i.e., difference between means) showed that accuracy was greater in the novel test compared to the exemplar test at both the short presentation time (.18) and the longer presentation time (.11). Importantly, there was a significant interaction between presentation time and condition, $F(1, 19) = 4.87$,

$MSE = .005$, $p < .05$, $\eta^2 = .204$. Accuracy did not increase much at the longer presentation time on the novel test (.02) but increased much more at the longer presentation time on the exemplar test (.09). Paired-tests showed that in the exemplar test condition, accuracy was higher for scenes presented for a longer presentation time, $t(19) = -3.78$, $p < .05$, but there was no difference between accuracy for scenes presented for short and long study presentation times in the novel test condition, $t(19) = -.942$, $p = .36$.

Confidence The mean confidence ratings for correct and incorrect old responses for each presentation and test condition are shown in Table 2. The same repeated measures analysis was conducted on mean confidence of correct old responses. There was a main effect of presentation time, $F(1, 19) = 11.786$, $MSE = .100$, $p = .003$, $\eta^2 = .383$. Mean confidence was higher for correct responses for scenes presented for a long than a short presentation time. There was no main effect of test condition, $F(1, 19) = .609$, $MSE = .054$, $p = .445$, $\eta^2 = .031$, and no significant interaction between presentation time and condition, $F(1, 19) = .020$, $MSE = .062$, $p = .890$, $\eta^2 = .001$. The corresponding analysis for mean confidence judgments of incorrect old responses revealed no significant effects, all F s (1, 14) < 1 . Degrees of freedom were 14, since five participants did not show false alarms in the novel test condition.

Response time The means of median response times for correct and incorrect old responses for each condition are presented in Table 3. The same ANOVA conducted on mean response time for correct responses revealed a main effect of test condition, $F(1, 19) = 16.4$, $MSE = 180,974$, $p = .001$, $\eta^2 = .464$. Response times were faster in the novel than in the exemplar test conditions. There was no main effect of presentation time, $F(1, 19) = 3.06$, $MSE = 293,655.49$, $p = .098$, $\eta^2 = .139$, and no significant interaction between presentation time and condition, $F(1, 19) = 1.81$, $MSE = 253,218.26$, $p = .568$, $\eta^2 = .017$.

The corresponding analysis of response time for incorrect old responses showed a main effect of presentation time, $F(1, 14) = 6.34$, $MSE = 641442.80$, $p = .025$, $\eta^2 = .312$. Response time was faster in the long presentation time ($M = 3,380.6$, $S.E = 212.2$) than in the short presentation time condition ($M = 3,901.6$, $S.E = 319.3$). There was no main effect of test condition, $F(1, 14) = 1.30$, $MSE = 328,225.42$, $p = .273$, $\eta^2 = .085$, but there was a trend for an interaction between presentation time and condition, $F(1, 14) = 3.94$, $MSE = 721,972.96$, $p = .067$, $\eta^2 = .220$. Response time was faster for incorrect responses in the long than in the short presentation time condition and this difference was larger in the novel than in the exemplar test condition.

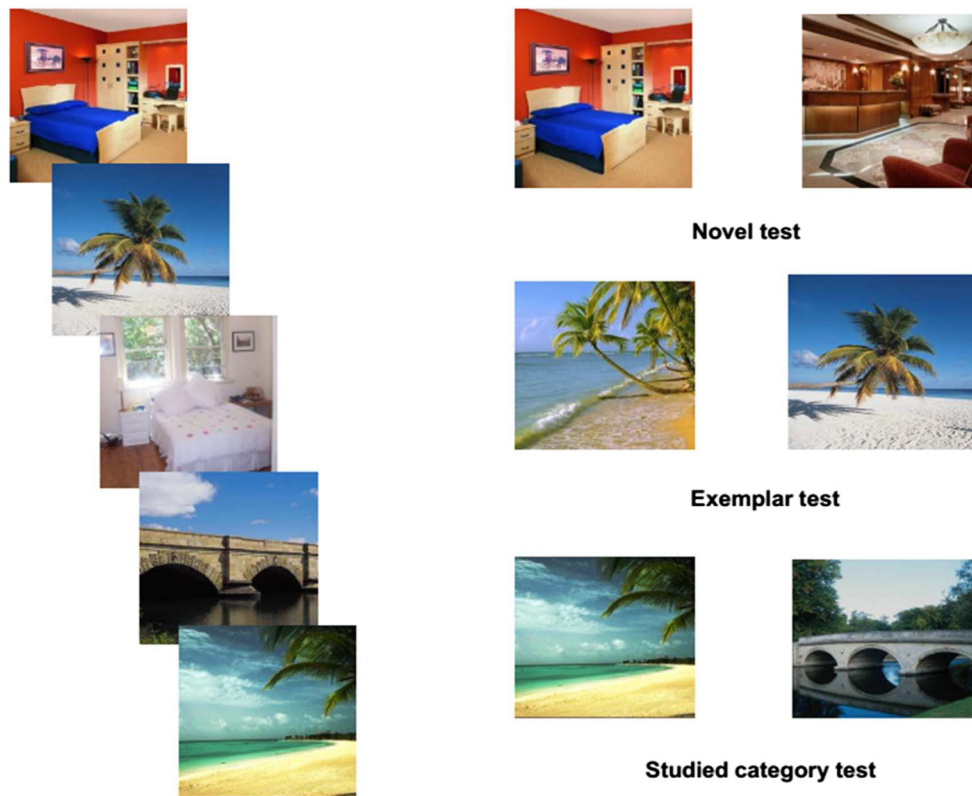


Fig. 1 Displayed is the study phase with a sample of the list of scenes. The test conditions presented in Experiments 1, 3 (Exemplar test, Novel test), and 2 (Exemplar test, Studied Category test) are displayed. Figures of the fixation boxes with an interstimulus interval (ISI) of 800 ms, are not included

An understanding of the effect of presentation time on gist- and verbatim-based retrieval could also be obtained from an analysis of the relationship between accuracy and confidence. In a study by Roediger and DeSoto (2014), the relationship between accuracy and confidence was investigated for yes-no recognition. In their study, a list of words were presented, after

which studied and non-studied words were presented. The non-studied words consisted of unrelated lures, weakly related lures, and strongly related lures. Roediger and DeSoto found a positive correlation between confidence and accuracy for studied items. In contrast, a significant negative correlation was found between confidence and accuracy for strongly

Table 1 Mean proportion correct and incorrect old responses in exemplar and novel test conditions (Exps. 1 and 3) and in exemplar and studied category conditions (Exp. 2) for short and long presentation times at encoding

Presentation time	Short		Long	
Test condition	Novel	Exemplar	Novel	Exemplar
Experiment 1 Correct	0.88 (0.08)	0.70 (0.09)	0.90 (0.08)	0.79 (0.10)
Experiment 1 Incorrect	0.12 (0.08)	0.30 (0.09)	0.10 (0.08)	0.21 (0.10)
Experiment 3 Correct	0.80 (0.12)	0.64 (0.12)	0.88 (0.10)	0.77 (0.10)
Experiment 3 Incorrect	0.20 (0.12)	0.36 (0.09)	0.12 (0.10)	0.23 (0.09)
Test condition	Studied Category	Exemplar	Studied Category	Exemplar
Experiment 2 Correct	0.68 (0.7)	0.67 (0.10)	0.76 (0.10)	0.76 (0.10)
Experiment 2 Incorrect	0.32 (0.7)	0.33 (0.09)	0.24 (0.10)	0.24 (0.11)

The studied category and exemplar test conditions differed in terms of whether the lure was from a different or the same studied category as the lure. The standard deviations of the means are given in parentheses

Table 2 Mean confidence and mean for correct and incorrect old responses to studied items in exemplar and novel test conditions for short and long presentation times at encoding (Exps. 1 and 3)

Presentation time	Short		Long		Mean
	Novel	Exemplar	Novel	Exemplar	
Confidence Correct (Exp 1)	2.04 (0.33)	2.01 (0.28)	2.30 (0.41)	2.24 (0.27)	2.148
Confidence Incorrect (Exp 1)	1.51 (0.43)	1.50 (.36)	1.60 (0.55)	1.50 (0.36)	1.528
Confidence Correct (Exp 3)	2.04 (0.40)	1.88 (0.33)	2.26 (0.43)	2.20 (0.34)	2.095
Confidence Incorrect (Exp 3)	1.65 (0.50)	1.64 (0.38)	1.43 (0.48)	1.70 (0.32)	1.605
Mean	1.810	1.758	1.898	1.910	1.844

The standard deviations of the means are given in parentheses. Marginal means for confidence are also shown

related lures and a significant positive correlation was found between confidence and accuracy for weakly related lures. We wanted to determine whether or not gist-based and verbatim-based retrieval are associated with a positive correlation between confidence and accuracy.

We undertook an analysis of correlation between accuracy (correct responses) and confidence averaged for each participant in order to determine if there was a relationship between confidence and accurate and inaccurate responses. All significant tests were one-tailed. There was one significant positive correlation between accuracy and confidence in the exemplar condition for long presentation time ($r(20) = .502, p = .012$) with the alpha level set at 0.05. To correct for multiple comparisons, we implemented a Bonferroni correction and found this correlation to be significant with an alpha level of 0.125.

Discussion

As expected, accuracy in the exemplar test condition was lower for short (1 s) compared to long (4 s) study presentation times, though it was well above chance in both cases. Accuracy, however, was equivalent at short and long presentation times in the novel test condition. These results confirmed our prediction that more time is required to extract and encode verbatim information needed to distinguish targets from lures in the exemplar test condition. For the novel test condition, accessing gist information is sufficient to differentiate the lure from the target. Because sufficient gist information of a scene is stored in LTM following a short presentation time of 1 s, no difference was observed between performance at the short and long presentation times.

Table 3 Mean of median response times (ms) for correct and incorrect old responses to studied items in exemplar and novel test conditions for short and long presentation times at encoding (Exps. 1 and 3)

Presentation time	Short		Long		Mean
	Novel	Exemplar	Novel	Exemplar	
Response time Correct (Exp 1)	2,869 (788)	3,433 (823)	2,787 (857)	3,061 (452)	3,038
Response time Incorrect (Exp 1)	4,035 (1417)	3,768 (1219)	3,079 (1069)	3,682 (907)	3,641
Response time Correct (Exp 3)	2,471 (830)	3,123 (1090)	2,389 (681)	2,634 (794)	2,653
Response time Incorrect (Exp 3)	3,281 (1408)	2,907 (990)	4,068 (4494)	3,450 (1581)	3,425
Mean	3,164	3,308	3,081	3,206	3,189

The standard deviations of the means are given in parentheses. Marginal means are also shown for response times

It is noteworthy that the confidence of correct responses reflected presentation time whereas the response time of correct responses indexed the effect of test condition. That is, both the results from the analyses of confidence ratings and response times for correct responses supported different aspects of the accuracy results. Mean confidence was higher for correct responses for scenes presented for a long than a short presentation time, whereas mean response time was faster in the novel than in the exemplar test condition. Interestingly, we also found a significant positive correlation between accuracy and confidence in the exemplar condition for the long presentation time. This positive correlation may be indicative of verbatim-based retrieval in this condition.

Experiment 2

In Experiment 2, we wanted to address the question of whether the type of exemplar test determined participants' recognition performance. When the target and the lure are from the same category, participants have more specific retrieval cues that enable retrieval of the target from LTM. That is, they have knowledge of specific verbatim information which is important to discriminate between the two test alternatives.

In fact, Tulving (1981) showed that an increase in degree of similarity between target and distractor improved recognition performance. In the first experiment of his study, participants studied a set of complex pictures, after which they were tested on their recognition of the studied pictures in a series of forced-choice tests, for which the similarity of the distractor to the studied picture was varied. For one forced-choice test condition, the distractor was similar to the target item (i.e., $A' - A'$) in both category and perceptual similarity. In another forced-choice test condition, the distractor was dissimilar to the target but similar to another previously studied picture (i.e., $A' - B'$) in both category and perceptual similarity. One should note that distractors in the $A' - A'$ condition represented the non-studied half of the two-page picture. However, for the $A' - B'$ condition, the distractor represented the other half of a different studied two-page picture. Finally, in the third forced-choice test condition, the distractor was dissimilar from both the target and the other pictures in the study series ($A - X'$). Tulving (1981) found as expected the hit rate was higher for the $A - X'$ condition compared to the other conditions. Interestingly, the hit rate was greater in the $A - A'$ compared to the $A - B'$ forced-choice test condition.

Tulving (1981) suggested the reason for this difference was that in the $A - A'$ condition, both test items are matched to the same memory trace, whereas in the $A - B'$ condition, the two test items are matched with two different memory traces. As a result, there is more efficient access to memory traces for the $A - A'$ forced-choice test condition compared to the $A - B'$ condition and this would account for participants' showing

higher recognition performance in the $A - A'$ compared to the $A - B'$ condition. Based on the results of Tulving's (1981) study, we hypothesized there would be higher recognition when the target and lure are of the same scene category compared to when the lure is from a different studied category.

To test this hypothesis, in Experiment 2, we replaced the novel test condition with a new exemplar test condition (i.e., the studied category test condition), such that the lure was drawn from a category that was studied rather than one that was novel. By presenting the target and the lure from different studied categories, participants do not have access to specific verbatim information that could support retrieval of the target from memory, but still need to base their decision on verbatim and not gist information. The purpose of Experiment 2 was to determine if there was a difference in recognition performance in the exemplar test and studied category test conditions.

Method

Participants Thirty-two students participated in Experiment 2.

Materials The materials were exactly the same as in Experiment 1. However, for the studied category test condition, the scene for the lure was of a different category, but of a category that was present during study. The lures in the studied category test condition were taken from the same categories of which scenes were taken for the exemplar test condition in Experiment 1.

Procedure The procedure was exactly the same as in Experiment 1. However, response time and confidence ratings were not collected because we were interested only if there was any difference in recognition accuracy between the two types of exemplar test conditions. Figure 1 displays examples of the stimuli and test conditions for Experiment 2.

Results and discussion

The mean proportions of correct and incorrect old responses in the exemplar and studied category test conditions for short and long presentation time are presented in Table 1. A 2 (Presentation time: 1 s vs. 4 s) \times 2 (Condition: Exemplar vs. Studied category test) repeated measures analysis showed there was no main effect of condition, $F(1, 31) = 1.92$, $MSE = .004$, $p = .175$, $\eta^2 = .059$. Similar accuracy was observed in the two types of exemplar test conditions. There was a main effect of presentation time, $F(1, 31) = 22.9$, $MSE = .010$, $p < .001$, $\eta^2 = .425$. Higher accuracy was shown for scenes presented for a longer presentation time than those presented for a short presentation time. There was no significant interaction

between presentation time and condition, $F(1, 31) = 1.90$, $MSE = .002$, $p = .178$, $\eta^2 = .058$.

The results of Experiment 2 showed that accuracy is greater for scenes presented for long, compared to short, presentation time in both exemplar test conditions. However, there was no interaction between test condition and presentation time. Thus, the results of Experiment 2 showed there was no difference in accuracy between the exemplar and the studied category test condition. These findings are contrary to the findings of Tulving (1981). One explanation may be that difference in similarity of the lure to the target may have produced different results. For example, in the current study, the target was paired with a lure which shared some perceptual details and could be considered a different scene. In contrast, in the Tulving (1981) study, the lure shared very similar perceptual details with the target scene and could also complete the target scene. That is, it could act as an effective cue for a verbatim trace to compare to that of the target scene. The interpretation of the findings of Experiment 2 will be considered further in the general discussion.

Experiment 3

Experiment 3 was of similar design as Experiment 1, with the only difference being that the short duration was reduced from 1 s to 500 ms. The purpose of Experiment 3 was to determine if reducing study time further might also reduce the encoding of gist information as well as verbatim information. Based on the work of Potter (1976), Potter et al. (2002), Schyns and Oliva (1994), and Greene and Oliva (2009a, b), 500 ms should be enough time to categorize the items and store gist traces of the studied items in STM. It may not, however, be enough time to store and retain information in LTM. We anticipated that there would also be reduced accuracy in the novel test condition for 500 ms compared to the 4-s presentation time, because 500 ms is not only less time to encode verbatim information but also less time to encode sufficient gist information in a study testing long-term recognition memory for scenes.

Method

Participants Twenty students participated in Experiment 3.

Materials The materials were exactly the same as in Experiment 1.

Procedure The procedure was exactly the same as in Experiment 1, except that the short presentation time was reduced to 500 ms.

Results and discussion

Accuracy The mean proportions of correct and incorrect old responses for each presentation and test condition are shown in Table 1. A 2 (Presentation time: 500 ms vs. 4 s) \times 2 (Condition: Exemplar vs. Novel) repeated measures analysis of variance (ANOVA) was conducted on proportion of correct responses. There was a main effect of presentation time, $F(1, 19) = 52.1$, $MSE = .004$, $p < .001$, $\eta^2 = .733$. Higher accuracy was shown for scenes presented for 4 s (long) compared to 500 ms (short). There was a main effect of test condition, $F(1, 19) = 33$, $MSE = .011$, $p < .001$, $\eta^2 = .635$. Higher accuracy was shown in the novel than in the exemplar test condition. That is, as shown by the mean difference scores, accuracy was greater in the novel test compared to the exemplar test at both the short presentation time (.16) and the longer presentation time (.11). There was also a significant interaction between presentation time and condition, $F(1, 19) = 5.67$, $MSE = .005$, $p < .05$, $\eta^2 = .230$. Accuracy increased at the longer presentation time on the novel test (.08) but increased much more at the longer presentation time in the exemplar test (.13). Paired *t*-tests showed, though, that in both the novel test ($t(19) = -4.32$, $p < .001$) and exemplar test conditions ($t(19) = -7.71$, $p < .001$), accuracy was lower at the short presentation time.

Confidence The mean confidence ratings for correct and incorrect old responses in each condition are shown in Table 2. The same repeated measures analysis was conducted on mean confidence of correct responses. There was a main effect of presentation time, $F(1, 19) = 25.9$, $MSE = .059$, $p < .001$, $\eta^2 = .577$. Mean confidence of correct responses was higher for scenes presented for a long than a short presentation time. There was a strong trend for a main effect of test condition, $F(1, 19) = 4.25$, $MSE = .058$, $p = .053$, $\eta^2 = .183$, as mean confidence judgments were higher in the novel than in the exemplar test condition. There was no significant interaction between presentation time and test condition, $F(1, 19) = 1.44$, $MSE = .037$, $p = .245$, $\eta^2 = .07$.

The corresponding analysis of confidence for incorrect old responses showed there were no main effects of presentation time, $F(1, 15) = 1.48$, $MSE = .113$, $p = .243$, $\eta^2 = .090$, or test condition, $F(1, 15) = 2.33$, $MSE = .113$, $p = .147$, $\eta^2 = .135$. There was, however, a significant interaction between these variables, $F(1, 15) = 7.40$, $MSE = .040$, $p = .016$, $\eta^2 = .331$. For the novel test condition, confidence of incorrect responses was higher in the short than the long presentation time condition, whereas in the exemplar test condition confidence was lower in the short than in the long presentation time.

As was done in Experiment 1, we conducted an analysis of correlations between accuracy and confidence. All significant tests were one-tailed. In contrast to Experiment 1, we did not

find any significant correlations, with our alpha level set to 0.0125 by a Bonferroni correction.

Response time The means of median response times for correct and incorrect old responses for each condition are presented in Table 3. The same analysis conducted on mean response times for correct responses revealed a main effect of presentation time, $F(1, 19) = 6.69$, $MSE = 244,553.63$, $p < .05$, $\eta^2 = .260$. Response time for accurate responses was faster at long than short presentation times. There was a main effect of test condition, $F(1, 19) = 19.8$, $MSE = 203016.82$, $p < .05$, $\eta^2 = .510$. Response times were faster in the novel than in the exemplar test conditions. There was a significant interaction between presentation time and condition, $F(1, 19) = 6.52$, $MSE = 127057.15$, $p < .04$, $\eta^2 = .256$; the increase in response time was greater between the novel and exemplar test conditions for short compared to long presentation times

The same analysis based on response times for incorrect old responses revealed no significant effects. There was no main effect of presentation time, $F(1, 15) = 1.31$, $MSE = .5374567.71$, $p = .269$, $\eta^2 = .081$, or test condition, $F(1, 15) = 1.039$, $MSE = .3781366.67$, $p = .324$, $\eta^2 = .065$, and these variables did not interact, $F(1, 15) = .755$, $MSE = .2374432.47$, $p = .755$, $\eta^2 = .007$.

The results of Experiment 3 showed that when presentation time is reduced to 500 ms, long-term recognition memory performance declined in both the novel and the exemplar test conditions, but the decline was greater in the exemplar test condition. The results suggest that 500 ms may not be sufficient time to encode even gist information fully, resulting in more sparse visual representations of the scenes stored in LTM and in lower accuracy compared to the longer presentation time in both exemplar and novel test conditions.

Encoding time and test condition also affected both confidence judgments and response times, but in different ways. There was a higher confidence for correct responses for scenes presented for a long than short presentation time, and a strong trend for confidence to be higher in the novel compared to the exemplar test. Correspondingly, response times were faster for longer presentation time and in the novel than in the exemplar test conditions. In addition, the faster response time in the novel compared to the exemplar test condition was larger for the short study presentation time.

Finally, in contrast to Experiment 1, we did not find any correlation between accuracy and confidence. The fact that we did not find the pattern of positive and negative correlations observed by Roediger and DeSoto (2014) may be due to the difference in recognition test procedure. Roediger and DeSoto (2014) examined yes/no recognition whereas we used a forced-choice test. Indeed, some researchers have suggested that there is a greater use of familiarity in the forced-choice test compared to the yes-no test (Ahmad and Hockley 2014; Ahmad, Fernandes, and Hockley 2015). Moreover, for the

forced-choice test, the recognition decision may depend on detecting only one element that distinguishes the remembered item from the lure. As a result, the differential use of familiarity and recollection depending on type of recognition test could affect the correlation between accuracy and confidence.

General discussion

The novel goal of the current study was to examine the influence of study presentation time on long-term recognition memory for scenes. We were interested in the type of scene representation that was retrieved after short and long study presentation times. Based on fuzzy-trace theory, we predicted that at short study presentation times more gist than verbatim information would be retrieved from LTM, as there would be less time for participants to encode detailed verbatim information. As evidence, participants would show higher recognition performance in the novel test condition compared to the exemplar test condition for both short and long presentation times.

As predicted, we found accuracy was influenced by whether performance at test was dependent on retrieval of gist or verbatim information. Specifically, accuracy was worse at short (1 s) compared to long (4 s) study presentation times only in the exemplar test condition in Experiment 1, which required retrieval of detailed verbatim information to distinguish targets from lures, but not in the novel test condition where retrieval of gist information sufficed. Interestingly, confidence reflected the effect of presentation time whereas response time was longer for recognition decisions based on retrieval of verbatim compared to gist information. In Experiment 2, we examined whether the type of exemplar test affected recognition performance. Contrary to what was found by Tulving (1981), there was no difference in recognition performance between the standard same exemplar test and a studied category test. Finally, for Experiment 3, when study presentation time was reduced even further to 500 ms, performance based on gist retrieval was also reduced, although not to the same extent as for verbatim retrieval.

Overall performance in our experiments for a study presentation time of 4 s was comparable to that reported by Konkle and colleagues (Konkle et al. 2010b). The novel contribution of our study was the finding of a differential effect of varying encoding time on retrieval of gist and detailed verbatim information of scenes from LTM. The findings from our study supported fuzzy-trace theory.

In Experiment 1, when presentation time was reduced to 1 s, participants showed an average of 88 % accuracy in the novel compared to 70 % accuracy in the exemplar test condition. In addition, accuracy for the novel test condition was similar for 1-s and 4-s presentation time. We interpret this finding as supporting the fuzzy-trace theory. Even though there is parallel

storage of gist and verbatim representation of a scene for both short and long study presentation times, encoding of verbatim information in a scene takes longer than that of gist information. As a consequence, higher accuracy is shown in the novel compared to the exemplar test for both short and long presentation times. Even with a short study presentation time of 1 s of a scene, participants' accuracy in the novel test was the same as for 4 s, because sufficient gist information could be encoded in LTM for a 1-s presentation time.

One of the main principles of fuzzy-trace theory is that processing of gist information is faster than processing of verbatim processing. Several studies using verbal stimuli have shown that not only is the meaning of target items encoded independently, but the surface form of the same studied items is encoded later than their meaning (Moravcski and Healy 1995; Wallace, Stewart, Schaffer, and Barry 1998; see Brainerd and Reyna 2004, 2005 for review). Our results show that the faster processing of gist compared to verbatim information extends to visual scenes.

Our suggestion that processing of gist information takes less time than verbatim information for scenes is also supported by studies examining short-term recognition memory for scenes (Castelano et al. 2009; Melcher 2006; Potter 2002; Tatler, et al. 2003). Tatler and colleagues (2003) showed that for a short presentation time of 1 s, complete gist information contained in pictures of real-world natural scenes is retained. However, retention of more detailed information in the scene such as presence of an object, object shape, color of objects, object location, and relative distances of objects required a longer presentation time. More generally, the finding that processing of gist is faster than verbatim information is consistent with the top-down guidance mechanisms that have been shown to influence perceptual processing, learning, and conscious perception (Ahissar and Hochstein 2004).

Interestingly, when examining the influence of presentation time on proportion correct for each type of information question there were some notable differences. For gist questions, performance was not influenced by viewing time. However, for the other information questions that tested retrieval of verbatim information, such as the presence of an object in the scene, there was a linear increase in performance with presentation time. Indeed, this continued accumulation of detailed information with an increase in presentation time is consistent with Melcher's (2006) finding of a linear increase in performance in answering questions about details of scenes (e.g., location of an object in the scene) as a function of total time the scene was viewed. In Melcher's study, participants were tested on their STM for details contained in scenes. The scene was presented on the screen for a period of 5, 10, or 20 s, after which four multiple questions were presented on the screen. Two questions probed the color of the object and location of the object in the scene. The third question tested knowledge of the view of the scene background. The last question tested

forced-choice recognition of shape of the object that was present in the scene. As expected, Melcher found recognition performance for all question types increased as a function of viewing time. He concluded that his findings supported previous studies (Melcher 2001; Melcher and Kowler 2001; Tatler et al. 2003) that have shown that scene representation accumulates over time, with gist and general layout represented first and details about specific objects added on as a result of subsequent fixations.

Thus, we extended the findings of Tatler et al. (2003) and Melcher (2006) to the effect of varying presentation time on long-term recognition memory for scenes. Whereas participants' long-term recognition memory of gist information was not affected by whether the scene was presented for 1 s or 4 s as shown by performance in the novel test condition, we showed reduced retrieval of verbatim information for short compared to longer presentation time in the exemplar test condition.

In Experiment 2, we examined if using a different type of exemplar test affected long-term recognition performance for scenes. Contrary to Tulving (1981), we did not find higher recognition performance in the exemplar test, where the lure was of the same category as the target compared to the studied category test, in which the lure was of a different studied category than the target. Thus, similarity of the lure did not improve recognition performance in the exemplar test. Importantly, the differential effect of study presentation time on retrieval in the exemplar and novel test could not have occurred due to the type of lure present in the exemplar test. An explanation for why our findings from Experiment 2 were not concordant with that of Tulving (1981) can be derived from the identity-similarity distinction of fuzzy-trace theory. As stated by Brainerd et al. (1999), according to the identity-similarity distinction, different comparisons lead to different patterns of accept-reject responses. Acceptance of targets and rejection of related distractors will occur even when the overlap between related distractor and target is great, because the related distractor's verbatim representation will be non-identical to the target's representation. Various manipulations can encourage verbatim retrieval in the face of gist retrieval from semantically-related distractors, when verbatim traces are highly accessible. False alarm rates will be reduced because, according to the identity-similarity distinction, retrieval can produce feelings of contrast about distractor probes. One such manipulation is varying the delay between presentations of targets and related distractors. Because verbatim traces are highly accessible after presentation, related distractors presented immediately after targets should be more easily rejected compared to related distractors presented after unrelated distractors which have followed the targets. The former condition would increase the chance that related distractors will cue retrieval of those traces that produce non-identity judgments (Brainerd et al. 1999). Thus, the comparison of test

item's verbatim representation with retrieved verbatim memory leads to correct rejection of related distractors, because there is a contrast effect setup between targets and related distractors but not between targets and unrelated distractors. This contrast effect was referred to as the verbatim priming effect. Brainerd et al.'s (1999) verbatim effect in the conjoint recognition paradigm was elaborated on by Stahl and Klauer (2008) for associates, category exemplars, category names, and rhymes.

Thus, to explain Tulving's (1981) study results, we can look to the verbatim priming effect. The lure in the A – A' condition in Tulving's study represented the non-studied half of the studied picture, while in our A – A' condition, the distractor was simply a scene of a similar category to the target. Moreover, for the A – B' condition or studied category test condition in our study, the distractor was similar to the studied scene of a different category, whereas in Tulving's study the distractor represented the other half of a different studied two-page picture. We argue that presentation of a target with a lure that shared very similar perceptual details and that could complete the target picture aided retrieval of the target (A – A'). In a sense, the distractor in Tulving's study acted as a prime of the verbatim trace, because there was a verbatim trace stored for the distractor, which aids verbatim retrieval. As a result, the false alarm rates for related distractors were below those of unrelated distractors. We could also argue that in contrast to Tulving's A – A' condition, for the comparable condition in our study, there is a reduced verbatim priming effect. The distractor item does not serve as a valid cue for the verbatim trace, and as a result the contrast effect is not set up between targets and related distractors (exemplars). This important difference would account for similar accuracy in A – A' and A – B' test conditions in our study. The results of Experiment 2 showed that the differential effect of study presentation time on retrieval in the exemplar and the novel test was not due to the type of lure present in the exemplar test that we implemented in our study.

Finally, in Experiment 3, we found when presentation time was reduced to 500 ms recognition performance was also reduced in the novel test compared to 1-s study presentation time. We interpret this finding as suggesting that for a long-term recognition memory task, a 500-ms presentation time may represent a study duration for which there is insufficient gist information stored and later retrieved to support the use of gist in rejecting novel distractors. It is probable that with even shorter presentation times, participants would have shown further decline in performance in the novel test.

The recognition advantage for gist over verbatim information in the novel test compared to the exemplar test as shown in both higher accuracy and faster response time can be related to the extralist-feature effect (ELF) demonstrated by Mewhort and Johns (2000). They found that the rejection of lures can be faster than endorsing targets when there is an extralist feature.

An extralist feature is a feature of a lure that was not present in the any of the study items. One can suggest that the "extra list feature" does not have to be a feature per se, but could also be gist-like information such as a category, that was not studied. In Mewhort and Johns (2000; Experiment 2), two shapes of the same color or of the same shape were presented during the study phase. At test, three types of probes were presented. One type of probe had an extralist feature for color and shape, indicating no overlap in features with study items. The other type of probe had one extralist feature and shared one feature with the probe. The last type of probe shared both features with studied items. Mewhort and Johns (2000) found there was higher accuracy and faster response times for test probes with extralist features compared to probes that shared features with studied items. The researchers concluded that the probe's features are compared with a global representation of a study set's features. Similarly, the novel category of the lure is an extralist feature. There is less similarity or overlap in features of the lure in the novel test compared to lures in the exemplar test. In addition, it is likely that even at the feature level there is little overlap between items belonging to one category (e.g., a beach) and another category (e.g., an office). As a result, there was greater accuracy and overall faster response times in the novel test condition compared to the exemplar test condition for correct responses.

In summary, gist and verbatim information are differentially available depending on encoding time, and for visual scenes, the encoding of verbatim information takes longer than gist information. To our knowledge, this study is the first to expand the fuzzy-trace principal findings of parallel processing and faster gist than verbatim processing to scene processing and examine its effect on long-term recognition memory for scenes. We showed in two experiments that when only encoding time is reduced and retention interval is kept constant, there is less verbatim than gist information retrieved from LTM, simply because there is significantly less verbatim compared to gist information encoded for 1-s and 500-ms study presentation times. Therefore, it takes longer for processing the verbatim than gist information in the scene and this reduces long-term recognition memory for verbatim details of a scene when the scene is presented for a short presentation time.

A potential direction for future research is to examine how the processes of familiarity and recollection present in long-term recognition memory for scenes are affected by short and long study presentation times. Remember and know judgments could be included with confidence judgments to provide measures of familiarity and recollection. According to Brainerd and Reyna (2002), retrieval of gist information relies on familiarity, whereas retrieval of detailed verbatim information relies on recollection. It would be interesting to examine how varying study presentation time affects recollection of different types of detailed verbatim information from scenes stored in LTM. Finally, recently age differences in

hippocampal activation during gist-based false recognition have been identified (Paige, Cassidy, Schacter, and Gutchess 2016). Paige et al. (2016) found that for large image set sizes at encoding, older adults showed higher gist-based false alarms than younger adults, and also showed reduced connectivity of hippocampus and posterior temporal regions compared to younger adults. It would be interesting to examine how study presentation time mediates the degree of age-related differences in gist-based false alarms in a study of similar design to the current study.

In conclusion, the findings from our study are consistent with the basic assumptions of fuzzy-trace theory. We show that gist and verbatim information are differentially available depending on encoding time, and that, in terms of visual scenes, the encoding of verbatim information takes longer than gist information. The retrieval of gist and verbatim information to distinguish a studied scene from a lure at test is dependent on study presentation time and the lures against which the stored memory trace must be compared.

Acknowledgments This research was supported by a Discovery Grant from the National Science and Engineering Research Council of Canada awarded to Morris Moscovitch. We thank Marilyn Ziegler for technical assistance. Portions of this work were presented at the 21st Annual Meeting of Cognitive Neuroscience Society, 5–9 April, 2014, Boston, Massachusetts and at the 25th Annual Meeting of the Canadian Society for Brain, Behavior and Cognitive Science (CSBBCS), 5–7 June, 2015, Carleton University, Ottawa.

Author An abstract needs to be able to stand alone, thus references are usually not cited in an abstract. If it is absolutely necessary to cite a reference, full details of the reference need to be given so that the reader can consult the publication – the reference list will not be available on search engines that publish only the abstract. I trust this is acceptable. Please confirm that BOTH Konkle et al. 2010 publications are being referred to in the abstract.

On page 11, please indicate which Konkle et al. 2010 study is being referred to.

Thank you

References

- Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, 8, 457–464.
- Ahmad, F. N., & Hockley, W. E. (2014). The role of familiarity in associative recognition of unitized compound word pairs. *The Quarterly Journal of Experimental Psychology*, 67, 2301–2324.
- Ahmad, F. N., Fernandes, M., & Hockley, W. E. (2015). Improving associative memory in older adults with unitization. *Aging, Neuropsychology, and Cognition*, 22, 452–472.
- Andermane, N., & Bowers, J. S. (2015). Detailed and gist-like visual memories are forgotten at similar rates over the course of a week. *Psychonomic Bulletin & Review*, 22, 1358–1363.
- Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, 105, 14325–14329.
- Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science*, 11, 164–169.
- Brainerd, C. J., & Reyna, V. F. (2004). Fuzzy-trace theory and memory development. *Developmental Review*, 24, 396–439.
- Brainerd, C. J., & Reyna, V. F. (2005). *The science of false memory*. Oxford: Oxford University Press.
- Brainerd, C. J., Reyna, V. F., & Mojardin, A. H. (1999). Conjoint recognition. *Psychological Review*, 106, 160–179.
- Castelhano, M., & Henderson, J. (2005). Incidental visual memory for objects in scenes. *Visual Cognition*, 12, 1017–1040.
- Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision*, 9, 1–15.
- Greene, M. R., & Oliva, A. (2009a). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology*, 58, 137–76.
- Greene, M. R., & Oliva, A. (2009b). The briefest of glances: The time course of natural scene understanding. *Psychological Science*, 20, 464–72.
- Hockley, W. E. (2008). The picture superiority effect in associative recognition. *Memory & Cognition*, 36, 1351–1359.
- Hockley, W. E., & Bancroft, T. (2011). Extensions of the picture superiority effect in associative recognition. *Canadian Journal of Experimental Psychology*, 65, 236–244.
- Intraub, H. (1997). The representation of visual scenes. *Trends in Cognitive Sciences*, 1, 217–222.
- Irwin, D. E., & Zelinsky, G. J. (2002). Eye movements and scene perception: Memory for things observed. *Perception & Psychophysics*, 64, 882–95.
- Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010a). Conceptual distinctiveness supports detailed visual long-term memory for real world objects. *Journal of Experimental Psychology: General*, 139, 558–578.
- Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010b). Scene memory is more detailed than you think: The role of categories in visual long-term memory. *Psychological Science*, 21, 1551–1556.
- Lampinen, J. M., Copeland, S. M., & Neuschatz, J. S. (2001). Recollections of things schematic: Room schemas revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 1211–1222.
- Melcher, D. (2001). Persistence of visual memory for scenes: A medium-term memory may help us to keep track of objects during visual tasks. *Nature*, 412, 401.
- Melcher, D. (2006). Accumulation and persistence of memory for natural scenes. *Journal of Vision*, 6, 8–17.
- Melcher, D., & Kowler, E. (2001). Visual scene memory and the guidance of saccadic eye movements. *Vision Research*, 41, 3597–3611.
- Mewhort, D. J. K., & Johns, E. E. (2000). The extralist-feature effect: Evidence against item matching in short-term recognition memory. *Journal of Experimental Psychology*, 129, 262–284.
- Moravcski, J. E., & Healy, A. F. (1995). Effect of meaning on letter detection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 82–95.
- Paige, L. E., Cassidy, B. S., Schacter, D. L., & Gutchess, A. H. (2016). Age differences in hippocampal activation during gist-based false recognition. *Neurobiology of Aging*, 46, 76–83.
- Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 509–22.
- Potter, M. C., Staub, A., Rado, J., & O'Connor, D. H. (2002). Recognition memory for briefly presented pictures: The time course of rapid forgetting. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 1163–1175.

- Roediger, H. L., & DeSoto, K. A. (2014). Confidence and memory: Assessing positive and negative correlations. *Memory*, *22*, 76–91.
- Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time-and spatial-scaledependent scene recognition. *Psychological science*, *5*(4), 195–200.
- Stahl, C., & Klauer, K. C. (2008). A simplified conjoint recognition paradigm for the measurement of gist and verbatim memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 570–86.
- Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, *25*, 207–222.
- Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science*, *19*, 73–74.
- Tatler, B. W., Gilchrist, I. D., & Rusted, J. (2003). The time course of abstract visual representation. *Perception*, *32*, 579–592.
- Tuckey, M. R., & Brewer, N. (2003). The influence of schemas, stimulus ambiguity, and interview schedule on eyewitness memory over time. *Journal of Experimental Psychology: Applied*, *9*, 101–118. doi:10.1037/1076-898X.9.2.101
- Tulving, E. (1981). Similarity relations in recognition. *Journal of Verbal Learning and Verbal Behavior*, *20*, 479–496.
- Wallace, W. P., Stewart, M. T., Schaffer, T. R., & Barry, C. R. (1998). Are false recognitions influenced by precognition processing? *Journal of Psychology: Learning, Memory, and Cognition*, *24*, 299–315.
- Wichmann, F. A., Sharpe, L. T., & Gegenfurtner, K. R. (2002). The contributions of color to recognition memory for natural scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 509–520.
- Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (2010). SUN database: Large-scale scene recognition from abbey to zoo. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010*, 3485–3492.