Transsaccadic Scene Memory Revisited: A 'Theory of Visual Attention (TVA)' Based Approach to Recognition Memory and Confidence for Objects in Naturalistic Scenes.

Melissa L.-H. Võ[1]

Werner X. Schneider[1,2]   Ellen Matthias[1]

[1]Department of Psychology, Ludwig-Maximilians-University Munich, Germany
[2]Department of Psychology, Bielefeld University, Germany

The study presented here introduces a new approach to the investigation of transsaccadic memory for objects in naturalistic scenes. Participants were tested with a whole-report task from which — based on the theory of visual attention (TVA) — processing efficiency parameters were derived, namely visual short-term memory storage capacity and visual processing speed. By combining these processing efficiency parameters with transsaccadic memory data from a previous study, we were able to take a closer look at the contribution of visual short-term memory capacity and processing speed to the establishment of visual long-term memory representations during scene viewing. Results indicate that especially the VSTM storage capacity plays a major role in the generation of transsaccadic visual representations of naturalistic scenes.

Introduction

The seeming ease and effectiveness with which we orient ourselves in our environment and the ability to select and store relevant information while inspecting visual scenes is based on a complex interplay of cognitive processes. These involve, for example, the selection and uptake of current as well as the storage of previously processed information in order to construct a meaningful visual representation of the world.

The importance of attention-memory interactions becomes evident when we consider the processes involved during scene viewing. Since high quality vision is restricted to a small region at the center of gaze, we constantly have to shift our eyes from one location in space to another at a rate of approximately three to four times per second. Visual information can only be gathered when the eye fixates, while information uptake is disrupted during the milliseconds necessary to make a saccade. This results in a sequence of visual snapshots interspersed with brief blind periods. Thus, the retinal image of the world changes with every saccadic eye movement. Nevertheless, we seem to be able to form a stable representation of the visual world that surrounds us by integrating information from prior fixations with new information from subsequent eye movements. However, exactly what information can be stored transsaccadically and what processes are involved when fleeting retinal images are transformed into more stable visual representations is still in dispute. The study presented here took a closer look at the contributions of two parameters of visual processing efficiency — visual perceptual processing speed and visual-short term memory storage capacity — to the establishment of transsaccadic memory for objects encountered during visual search in naturalistic scenes.

Visual memory can be subdivided into at least three different memory stores: Iconic memory, visual short-term memory (VSTM), and visual long-term memory (VLTM) (see Irwin, 1992b and Hollingworth, 2006 for reviews). Iconic memory is characterized by precise, high-capacity sensory traces generated across the visual field, which are transient and susceptible to new sensory input, e.g., masking. VSTM on the other hand stores sensory information enabling accumulation of visual information across saccades, while its storage seems to be limited to about three to four objects (Irwin, 1992a; Luck & Vogel, 1997). VLTM is assumed to maintain similar visual representations as stored in VSTM, but with a storage capacity and duration that can greatly exceed VSTM (see Hollingworth, 2004, 2005, 2006). Exactly what memory systems are involved when we inspect complex visual scenes or what information survives saccadic eye movements have been in dispute in the domain of transsaccadic memory.

Since our visual system is exposed to much more information than it could possibly process, the first step in generating a meaningful representation is to select relevant information for access to VSTM. How this selection takes place — a process that is labeled 'encoding' by memory theories — is the subject of theories on visual attention (e.g., Awh, Vogel, & Oh, 2006; Bundesen, 1990; Duncan & Humphreys, 1989; Schneider, 1995, 1999). According to these models, two limiting factors determine the efficiency of encoding: the speed of processing perceptual information usually measured as the number of elements that can be processed within a second and the amount of information, which can be stored in VSTM in order to be further processed (e.g., Bundesen, 1990).

We therefore argue that representations in transsaccadic memory heavily depend on the ability to process and store information in VSTM. In order to investigate the influence of visual processing efficiency on transsaccadic object memory, the present study used an integrated theoretical and methodological approach which permits components of visual attention to be assessed independently of each other and of any (potentially confounding) motor components, namely: the Theory of Visual Attention (TVA; Bundesen, 1990, 1998). TVA assumes that a number of latent processes underlie overt performance. These processes are formally described by a coherent, mathematical theory in terms of a set of (mathematically)

independent quantitative parameters (for a detailed mathematical description, see Bundesen, 1990, 1998; Kyllingsbaek, 2006). The TVA model is strongly related to the biased-competition conceptualization of visual attention (Desimone & Duncan, 1995). In this view, visual objects are processed in parallel and compete for selection (i.e., conscious representation). The race among objects can be biased such that some objects are favored for selection based on either automatic, 'bottom-up' (e.g. sensory salience) or intentional, 'top-down' (e.g. task context) factors. In TVA, selection of an object is synonymous with its encoding into limited-capacity VSTM. Objects that are selected and hence may be reported from a briefly exposed visual display are those elements for which the encoding is completed before the sensory representation of the stimulus array has decayed and before VSTM has filled up with other objects. Because VSTM capacity is limited (to around four elements in normal subjects), objects in the stimulus array compete to be encoded into VSTM (especially if their number exceeds the VSTM capacity).

In TVA, the general efficiency of the visual processing system is reflected in the parameters visual perceptual processing speed C (number of visual elements processed per second) and visual short-term memory storage capacity K (number of elements maintained in parallel). Both parameters can be assessed using a whole-report task, in which participants are briefly presented with arrays of simple stimuli, e.g., letters, at varying exposure durations and have to identify (name) as many as possible. The probability of identifying a given object x is modeled by an exponential growth function. The slope of this function indicates the total rate of information uptake in objects per second (perceptual processing speed, denoted by C), and its asymptote the maximum number of objects that can be represented at a time in VSTM (VSTM storage capacity, K). A number of behavioral and physiological studies have suggested that the maximum capacity of VSTM ranges between three and four items (e.g., Luck & Vogel, 1997; Todd & Marois, 2004; for a review see Cowan, 2001 and Jonides, Lewis, Nee, Lustig, Bermann, & Moore, 2008) similar to estimates of VSTM capacity limits of three to four objects across saccades (see Irwin, 1992a; Hollingworth, 2006). A key question of this study was therefore whether the TVA parameters C and K derived from a task using simple letters as test material can translate to predict the performance of transsaccadic

LTM for objects embedded in complex, naturalistic scenes.

There have been quite different theoretical positions regarding the nature of transsaccadic memory ranging from theories that suggest that no detailed visual representations accumulate as attention is oriented from object-to-object within a scene (e.g., Becker & Pashler, 2002; Horowitz & Wolfe, 1998; O'Regan, 1992; O'Regan, Rensink, & Clark, 1999; Rensink, 2000, 2002) to theories that propose that indeed very detailed visual information of an object can be stored across saccades (e.g., Hollingworth & Henderson, 2002; Melcher, 2006). A possible reason for such diverging views on transsaccadic memory is the kind of information of an object that is investigated, e.g., its visual form, size, or orientation. Recent studies suggest that different features show different rates of memory accumulation and decay (Melcher & Morrone, 2003; Tatler, Gilchrist, & Rusted, 2003), which might have led to contradictory findings across studies.

In a number of experiments, Hollingworth and colleagues (Hollingworth, 2004; Hollingworth & Henderson, 2002; Hollingworth, Williams, & Henderson, 2001) tested memory for the visual form of objects. They were able to show robust VLTM for visual detail across several eye movements. In a so-called follow-the-dot-paradigm, for example, participants had to fixate a series of objects in a scene, following a dot, which moved from object to object (Hollingworth, 2004). Afterwards, memory was tested showing that object memory was not only superior for recently attended objects, but also for objects attended earlier in the fixation sequence indicating a VLTM component to scene representation. Moreover, change detection experiments revealed that even after a delay of 24 hours change detection performance remained well above chance (Hollingworth, 2005). These findings support the view that visual representations stored in VLTM can be rich in detail and that transsaccadic memory relies on both a VSTM and a VLTM component allowing for the accumulation of information across saccades (e.g., Castelhano & Henderson, 2005; Henderson & Hollingworth, 2003; Hollingworth, Williams et al., 2001; Tatler et al., 2003; Tatler, Gilchrist, & Land, 2005). Interestingly, it seems that also eye movement behavior shows strong similarities during encoding and retrieval of pictorial information regardless of whether retrieval is immediate or delayed by as long as two days (see Humphrey & Underwood, this issue).

In a recent study, we investigated the influence of scene previews on search efficiency as well as incidental memory for the objects encountered during the search (see, Võ & Schneider, submitted). The main features of the paradigm used in the search experiment were that a) during search the visual input was limited to a gaze contingent window of only 2° diameter centered on the participants' fixation, while parafoveal vision of the search scene was impeded by masking the remainder of the search scene (see Castelhano & Henderson, 2007) and b) participants were not told that after completing the search experiment an object recognition memory test would be given (see Castelhano & Henderson, 2005). The study presented here reanalyzed the incidental memory data in order to introduce a new approach to the investigation of transsaccadic scene memory. We used the TVA paradigm to collect parameters regarding the VSTM capacity K on the one hand and processing speed C on the other for those participants who had previously taken part in the visual search experiment outlined above. This allowed us to investigate the contribution of visual processing efficiency to transsaccadic memory performance for objects encountered in naturalistic scenes. Thus, finding differential effects of K or C on transsaccadic memory would lend support for a number of assumptions.

First of all, this would imply that differences found in either of the parameters K or C — collected using a whole-report task with strings of letters as test items — would generalize to explain differences in transsaccadic object memory performance using complex 3D-rendered images of naturalistic scenes. While encoding and memorizing letters does not usually entail visually detailed information, remembering objects placed in naturalistic scenes involves visual representations that are detailed enough to distinguish between objects of the same category, e.g., distinguishing between different types of toasters according to their respective visual shape or color.

Second, according to the TVA, the parameters K and C refer to VSTM while the objects in the recognition memory test had to be stored transsaccadically in a VLTM storage in order to survive the great number of fixated objects that inescapably exceed VSTM storage capacity. Therefore, an effect of either K or C on object memory accuracy would add further support to the claim that visually detailed short-term object file representations can indeed be consolidated into similarly detailed

long-term representations (Hollingworth & Henderson, 2002). Since participants were not told that a memory task would follow the primary search task, strategic verbal encoding of encountered distractor objects was unnecessary. However, even though participants were not asked to memorize distractor objects, storing visual and spatial information of encountered objects in VSTM would prevent unnecessary reinspections of locations already searched and objects viewed. Thus, the efficiency of the participants' VSTM systems should also affect the accuracy of long-term memory object representations.

Accordingly, we expected that a generally greater VSTM capacity should result in greater memory accuracy for distractor objects encountered during target search, i.e., a greater VSTM capacity should allow more information to remain activated for a longer amount of time before it is replaced by new incoming information increasing the probability for information stored in VSTM to consolidate into VLTM and elevating memory accuracy. Processing speed, on the other hand, should not show significant influence on transsaccadic memory for objects encountered during visual search due to the use of a gaze contingent window, which did not allow more than one object to be processed at a time. Therefore, even participants with a low processing speed (e.g., 10 elements per second) should be able to process sufficient information of the one visible object during its fixation which usually lasts at least 200 ms (see Shibuya & Bundesen, 1988).

Third, due to the participants' task to search for verbally predefined target objects, these should be superiorly processed and stored as compared to distractor objects, which should be rejected as non-target objects as soon as they are recognized weakening their memory traces. Therefore, we expected higher recognition memory for targets as compared to distractors.

Finally, in addition to testing recognition memory performance we collected confidence ratings for every object displayed in the test scenes. This provided us with a more finely graded and subjectively mediated measure of transsaccadic memory in addition to the accuracy of memory test performance. Thus, by collecting confidence ratings for recognition memory judgments, we were able to test whether VSTM capacity and processing speed similarly influence memory accuracy and confidence or whether both parameters have differential influence on objective and subjective measures of object recognition memory.

The present study is the first to investigate the contribution of two visual processing efficiency parameters, namely processing speed and VSTM storage capacity, to transsaccadic object memory by relating the outcome of a visual object recognition task to TVA parameters derived from a whole-report task.

Methods

Participants

Twenty-five students (18 female) from the LMU Munich ranging in age between 19 and 28 (M = 23.24, SD = 2.55) participated in the study for course credit or for 8€/hour. All participants reported normal or corrected-to-normal vision and were unfamiliar with the stimulus material. All 25 participants had first taken part in a visual search and object recognition experiment and were later tested with a whole-report task in order to assess the TVA parameters VSTM capacity and processing speed. The TVA parameters were calculated in order to subsequently split the participants into groups of high and low VSTM capacity and high and low processing speed groups, respectively. These TVA based groups were then compared with regard to their transsaccadic object memory performance in the search experiment.

Stimulus Material

The search scenes in the *study phase* consisted of 20 3D-rendered images of real-world scenes. The scenes were displayed on a 19-inch computer screen (resolution 1024 x 768 pixel, 100 Hz) subtending visual angles of 28.98 (horizontal) and 27.65 (vertical) at a viewing distance of 70 cm. The default background color was gray (RGB: 51, 51, 51). Each search scene was preceded by a Control preview, which was created from scrambled quadratic sections (8 x 8 pixel) taken from all search scenes and also served as a mask. Thus the Control was meaningless, but contained colors, orientations, and contours as is the case in unscrambled scenes. Each participant saw each search scene only once.

The stimuli presented in the *test phase* consisted of 20 3D-rendered naturalistic scenes. These scenes mainly resembled the search scenes of the study phase, except for the fact that for each scene about half of the objects (depending on the scene three to five objects) were replaced by different objects which were similar in size as well as in their scene and location plausibility.

For the *whole-report task*, five red target letters (each 0.5° high x 0.4° wide) were presented in a vertical column, 2.5° of visual angle either to the left or to the right of a fixation cross, on a black screen. Stimuli for a given trial were randomly chosen from a pre-specified set of letters (ABEFHJKLMNPRSTWXYZ), with the same letter appearing only once per trial. In some trials letter displays were masked. Masks consisted of letter-sized squares (of 0.5°) filled with a '+' and an 'x'.

Apparatus

Target Search and Object Recognition. Eye movements were recorded with an EyeLink1000 tower system (SR Research, Canada) which tracks with a resolution of .01° visual angle at a sampling rate of 1000 Hz. The position of the right eye was tracked while viewing was binocular. Experimental sessions were carried out on an IBM compatible display computer running on OS Windows XP. Stimulus presentation and reaction recording was controlled by Experimental Builder (SR, Research, Canada). The eye tracker was hosted by another IBM compatible computer running on DOS, which recorded all eye movement data. Both study and test phase were conducted at the same display computer. However, no eye movement data was collected during the test phase.

Whole-report task. The TVA experiment was conducted in a dimly lit, sound-proof cubicle. Stimuli were presented on a 17" monitor (1024x768 pixel screen resolution, 70 Hz refresh rate). Subjects viewed the monitor from a distance of 50 cm, controlled by the aid of a head- and chinrest.

Procedure

Study Phase. The procedure of the study phase closely followed the procedure of the "Flash Preview Moving Window" paradigm used in the experiments of Castelhano and Henderson (2007). Experimental sessions were conducted in a moderately lit room (background luminance about 500 lx), in which the illumination was held constant. Each participant received written instructions before being seated in front of the presentation screen. Participants were informed that they would be presented with a series of scenes in which they had to search for a target as fast as possible. They were also informed that short previews of the scene would precede the display of the search scene and that they should attend to these previews since they could provide additional information. At the beginning of the experiment, the eye tracker was calibrated for each participant. Therefore, the participants' viewing position was fixed with a chin and forehead rest, followed by a 9-point calibration and validation.
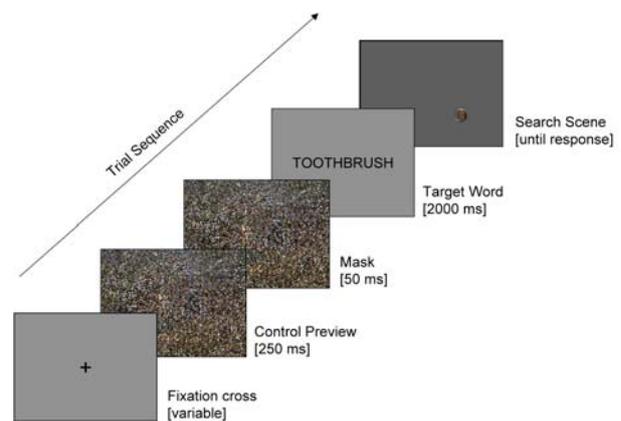

Figure 1. Trial sequence of the target search.

As can be seen in Figure 1, each trial sequence was preceded by a fixation check, i.e., in order to initiate the next trial, the participants had to fixate a cross centered on the screen for 200 ms. When the fixation check was deemed successful, the fixation cross was replaced by the presentation of the scene's preview for 250 ms. However, for this study only those trials were of interest that were not preceded by scene previews, but by a 250 ms mask. Subsequently a 50 ms mask followed, which was identical to the mask used in the Control condition. Then a black target word was displayed at the center of the gray screen for 2000 ms, which indicated the identity of the target object. Afterwards the search scene was shown through a 2° diameter circular window moving contingent on the participants' fixation location. The rest of the display screen was masked in gray. Thus, no peripheral

vision was possible throughout the entire visual search. Participants had to search the scene for the target object and indicate the detection of the target object by holding fixation on the object and pressing a response button. The search scene was displayed for 15 s or until button press. Three practice trials at the beginning of the experiment allowed participants to get accustomed to the experimental setup and the restricted vision during search due to the gaze contingent window. The study phase lasted for about 20 minutes. However, participants did not know that there would be a recognition task after they had completed the visual search experiment.

Test Phase. After participants had completed the visual search, they were again seated at the display computer with their heads fixed by the chin and forehead rest. Participants were informed that they would be presented with the same scenes as they had encountered during search in which some "old" objects had been replaced with "new" objects. All objects of interest were marked with a surrounding rectangle. Participants were to indicate for each object whether it was "old" or "new". In order to do so, participants used a mouse to click onto each of the objects within the scene marked by a rectangle, which activated a response screen on which they were asked to give a confidence rating on a 6-point scale from "very sure old", "sure old", "not sure old" to "not sure new", "sure new", "very sure new". The three ratings indicating "old" decisions were presented on the left half of the screen, the three ratings indicating "new" decisions were presented on the right half. This response setup was counterbalanced across participants. The participants were asked to progress from left to right by starting with the object on the far left and ending with the object on the far right. The scenes were presented in the same randomized order as they had been presented during the visual search task. Thus, the time lag between study and test phase was constant for each scene. The test phase lasted about 35 minutes.
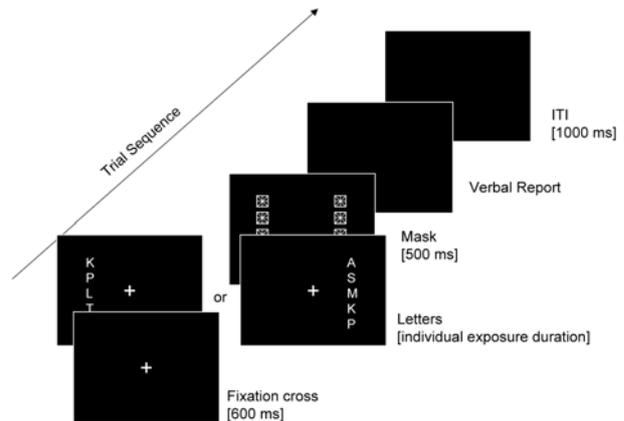

Figure 2. Sequence of frames on a given trial of the TVA based whole-report task.

Whole-report. Figure 2 shows the trial sequence of the whole-report task. Participants were first instructed to fixate a white cross (0.3° x 0.3°) presented for 600 ms in the centre of the screen on a black background. Then five red target letters were presented in a vertical column either to the left or to the right of the fixation cross. The participants had to report as many letters as possible. The experiment comprised two phases: In phase 1 (pre-test), three exposure durations of the target letters were determined for phase 2 (main test), in which the data were collected. The pre-test comprised 24 masked trials with an exposure duration of 86 ms. It was assessed whether the subject could, on average, report one letter (20 %) per trial correctly. If this was achieved, exposure durations of 43 ms, 86 ms, and 157 ms were used in the main test. Otherwise, longer exposure durations of 86 ms, 157 ms, and 300 ms were used. Here, letter displays were presented either masked or unmasked. The masks were presented for 500 ms at each letter location. Due to 'iconic-memory' buffering, the effective exposure durations are usually prolonged by several hundred milliseconds in unmasked as compared to masked conditions (Sperling, 1960). Thus, by factorially combining the three exposure durations with the two masking conditions, six different 'effective' exposure durations were produced. These were expected to generate a broad range of performance, tracking the early and the late parts of the functions relating response accuracy to effective exposure duration. In several previous studies that used a similar paradigm (e.g., Finke, Bublak, Krummenacher, Kyllingsbaek, Müller, & Schneider, 2005), highly reliable estimates of the parameters C and K were obtained on the basis of 16 trials per target condition. On this basis, each subject

completed 288 trials (2 hemi-fields x 2 masking conditions x 3 exposure durations x 16 trials per target condition) in the present experiment. Before each phase, subjects were given written and verbal instructions.

Data reduction and statistical analysis

For the present study we reanalyzed a subset of the original target search and object recognition experiment data, i.e., data was solely taken from those 25 participants that had subsequently also taken part in the TVA experiment. Further, we only included trials which were preceded by Control previews, i.e., non-informative masks, since the focus of the study reported here was not the investigation of the influence of the previews. The complete data of all participants and all preview conditions is reported elsewhere (see, Võ & Schneider, submitted).

Dependent variables of interest in this study comprised both Object Recognition Accuracy defined as the percentage of objects correctly judged as "old" or "new" and Confidence operationalized as the mean confidence rating given for the old/new response. These dependent variables were separately analyzed for target and distractor objects. For the analyses of all dependent variables only correct searches were included, i.e., when the participant pressed the response button while fixating the target object. Additionally, we excluded trials with a trial fixation number greater than 50 mostly caused by unstable calibration of the gaze contingent window [8.15 %].

The experimental results of the whole report task are described by the TVA parameter estimates for 'visual perceptual processing speed' and 'VSTM storage capacity'. These parameters were estimated using the standard procedure introduced by Duncan, Bundesen, Olson, Humphreys, Chavda, & Shibuya (1999) and used in several other recent studies (e.g., Bublak, Finke, Krummenacher, Preger, Kyllingsbaek, Müller, & Schneider, 2005; Finke, Bublak, Dose, Müller, & Schneider, 2006; Habekost & Rostrup, 2007; Hung, Driver, & Walsh, 2005).

For both targets and distractors, independent T-tests were calculated for Object Memory Accuracy and Confidence with processing speed C ("high speed" vs. "low speed") and memory storage capacity K ("high capacity" vs. "low capacity") as between-subject factors. Since accuracy and confidence ratings for targets were subject to ceiling effects, we additionally calculated the non-parametric Wilcoxon Rank Sum Test, which yielded identical p-values.

Results

As mentioned in the introduction, in TVA, the efficiency of processing is defined by two parameters: visual perceptual processing speed C and VSTM storage capacity K (Bundesen, 1990, 1998; Bundesen, Habekost, & Kyllingsbaek, 2005). Parameter *C* was estimated (by TVA model fitting) as the average of the summed processing rate values v for the objects presented to the left and the right of fixation, respectively. *C* is defined as a measure of the perceptual processing speed in elements/second. And parameter *K* reflects, the number of letters that can be simultaneously maintained in VSTM (Bundesen, 1990; Duncan, Bundesen, Olson, Humphreys, Chavda, & Shibuya, 1999; Kyllingsbaek, 2006).

Table 1: Summary of mean values Accuracy and Confidence as a function of participant groups ("low-K-value group" vs. high-K-value group", "low-C-value group" vs. "high-C-value group") split for targets and distractors.

Planned Contrasts	Targets	p-value	Distractors	p-value
K groups				
Mean Accuracy in %				
high K values	100.00 [.02]	.21	56.69 [.02]	.04
low K values	98.33 [.01]		50.07 [.03]	
Mean Confidence				
high K values	3.00 [.08]	.12	1.53 [.07]	.04
low K values	2.87 [.07]		1.34 [.08]	
C groups				
Mean Accuracy in %				
high C values	97.73 [.01]	.13	55.66 [.02]	.17
low C values	100.00 [.01]		52.00 [.03]	
Mean Confidence				
high C values	2.93 [.08]	.43	1.54 [.07]	.04
low C values	2.91 [.09]		1.35 [.08]	

VSTM Capacity K: K across all participants ranged from 2.38 to 4.00 (M = 3.31, SD = .57). According to K values we divided participants into two groups: The low-K-value group consisted of ten participants that showed a K value less than 3 (M = 2.67, SD = .19), while the high-K-value group consisted of 15 participants showing a K value greater than 3 (M = 3.75, SD = .21).

Processing Speed C: C across all participants ranged from 6.22 to 33.46 (M = 17.06, SD = 7.10). According to C values we divided participants into two groups: The low-C-value group consisted of 11 participants that showed a *C* value less than 15 (M = 11.29, SD = 2.56), while the high-C-value group consisted of 14 participants showing a *C* value greater than 15 (M = 21.60, SD = 6.15).

K groups: As can be seen in Table 1, mean accuracy and mean confidence for distractor objects were greater for the high-K-value group than for the low-K-value group, t(24) = 3.35, p < .05 and t(24) = 3.47, p < .05, respectively. For targets, the neither accuracy nor rated confidence for targets differed between K groups, t(24) = .81, p > .05 and t(24) = 1.19, p > .05, respectively. Non-parametric Wilcoxon Rank Sum Tests yielded identical p-values. Thus, a high VSTM capacity leads to greater LTM accuracy accompanied by higher degrees of confidence for distractor, but nor target objects.

C groups: As can be seen in Table 1, participants with high C values showed greater confidence for distractors than participants with low *C* values, t(24) = 3.59, p < .05. Both *C* value groups did not differ in the accuracy for distractor object recognition, t(24) = .99, p > .05.

For targets, the two *C* value groups differed neither in the mean accuracy nor in the confidence for target objects, all t(24) < 1. Again, Wilcoxon Rank Sum Tests resembled the outcome of the t-tests.

Contrary to VSTM capacity, higher visual perceptual processing speed could not account for LTM accuracy. However, processing speed seems to have differential effects regarding distractor and target objects: while participants with higher processing speeds show greater con-

fidence in their recognition judgments for distractors, this was not the case when judging targets.

Correlation of Processing Speed C and VSTM Capacity K: There was a significant correlation between the two TVA parameters *C* and *K*, r = .45, p < .05. However, given the specific set-up of the search experiment, e.g., the gaze-contingent window, the VSTM capacity *K* seems to be the parameter that determines transsaccadic object memory performance to a greater degree than processing speed *C*.

Discussion

The departure point of the study was to investigate the contributions of visual-short term memory storage capacity on the one hand and visual perceptual processing speed on the other to the establishment of transsaccadic memory for objects encountered during visual search in naturalistic scenes. We therefore re-tested a group of participants — that had taken part in a visual search experiment — with a TVA whole report experiment, which provided us with information regarding the efficiency of the individual's processing system. This additional information allowed us to reanalyze data on transsaccadic object memory according to either high or low VSTM capacity and high or low processing speed, respectively.

We found that the participants greatly differed in their speed of processing shortly flashed information as well as in their ability to store this information in VSTM for later report. According to the TVA parameters collected from the whole-report task, we were able to split participants into two groups, which differed significantly in their VSTM capacity: While the low-K-value group showed a mean VSTM capacity of about three objects, participants with high K values were able to store about four items on average. Even more distinct were the differences between participants regarding high and low processing speed: The low-C-value group was almost half as fast in processing information as the high-C-value group, i.e., while the former group processed about 11 items per second on average, the latter group could process up to 22 items per second. Thus, the participants demonstrated significant group differences, which allowed us to further investigate the relation of interindividual differences in TVA pa-

rameters to the respective object memory performance. Despite a significant correlation between C and K we found that both parameters had differential effects on recognition memory performance and confidence for objects encountered during search.

The first issue to address is the finding of great differences in encoding and storage of visual information regarding target and distractor objects. Even though participants were not told to memorize targets for later test, target objects were primed by the target word and were — contrary to distractor objects — essentially task relevant. Consequently, the task relevance of the target object as compared to the distractors might have led to an increased amount of attention deployed to the target object upon detection further benefiting its consolidation into VLTM. Therefore, target objects showed remarkable recognition memory accuracy amounting to nearly 100%. With such nearly perfect memory performance a neither VSTM capacity not processing speed could further modulate recognition memory accuracy. Therefore, we concentrated on recognition memory performance for non-target objects and their susceptibility to the influence of VSTM capacity and processing speed.

In line with our hypotheses, we found that object recognition memory for distractors significantly varied as a function of VSTM storage capacity. Recognition accuracy was higher for participants of the high-K-group than for participants of the low-K-group. Thus, we were first of all able to show that the TVA parameter K could be generalized to explain effects in transsaccadic memory for objects displayed in naturalistic scenes. It seems that the capacity to store information in VSTM, which is involved in the encoding and short-term retention of simple stimuli is also a determining factor during the inspection of more complex stimulus material. This is in line with recent findings suggesting that VSTM is limited in terms of items regardless of their complexity (Awh, Barton & Vogel, 2007).

A second conclusion can be drawn from these results regarding the involvement of both visual short-term and visual long-term components in transsaccadic object memory during scene viewing. According to TVA, the parameters C and K determine the efficiency of selective encoding of visual information into VSTM (see Bundesen, 1990, 1998; Kyllingsbaek, 2006). However, we found K-group differences in recognition memory for objects that had been encoded with a delay that clearly

exceeds VSTM retention intervals. How can an increased VSTM storage capacity lead to better VLTM performance? The moving window technique of the visual search task imposed a serial one-object-at-a-time encoding strategy. VSTM is loaded with objects until its capacity limitation is reached. A larger VSTM capacity implies that the first encoded object remains within VSTM for a longer amount of time before it is replaced. Consequently, objects encoded into a VSTM store with a greater capacity also have a greater likelihood of being consolidated into VLTM representations.

Castelhano and Henderson (2005) investigated whether the intention to store visual information is required for successful encoding into transsaccadic memory. Findings of more detailed visual representations stored in memory might be due to the task, which explicitly asks participants to memorize objects of a scene in preparation for a subsequent memory task. In their study, Castelhano and Henderson therefore investigated whether the visual properties of objects are stored incidentally in LTM as a natural consequence of scene viewing, or whether such representations are only stored when participants are intentionally memorizing a scene. Therefore, memory performance following a memorization task, where participants were explicitly asked to intentionally encode and remember object details while viewing a scene, was contrasted with memory performance following a visual search task, in which participants were unaware that a memory test would follow. During memory test, previously viewed objects and foil objects differing only in visual detail were presented in a difficult two-alternative forced choice-memory test. Regardless of whether participants had intentionally memorized objects while viewing a scene or whether they had incidentally scanned objects during visual search, performance was above chance. Thus, their results argue for the involvement of transsaccadic memory even without intentional memorization of scene details.

Our results lend support to the findings of Castelhano and Henderson (2005). Even without the instruction to memorize distractor objects during target search, VSTM capacity showed its effect on memory accuracy in a subsequently administered test. Thus, transsaccadic storage of visual information determines incidental recognition memory accuracy. However, only the high-K-group showed above chance memory performance for distractor objects. This could be due to the additional impediment

of the limited visual field during search with a moving window. Zelinsky and Loschky (2005) found that memory performance decreased when objects were presented serially and therefore without parafoveal vision of other objects on the display arguing for parafoveal benefits when memorizing objects scenes. This would explain the generally lower recognition memory performance in our study as compared to the study by Castelhano and Henderson where the visual field was unrestricted.

Contrary to VSTM storage capacity K, the visual processing speed C did not modulate object recognition accuracy for distractor objects. According to TVA, C mainly shows its effects when presentation time is insufficient to encode multiple items into VSTM or when presentation for a single item is below 200 ms (see Shibuya & Bundesen, 1988). However, participants in our study were only able to search the scenes through a 2° diameter sized window that moved around with the participant's gaze. Thus, only one object at a time was visible during search. With an average fixation time of about 200 ms even participants with a low processing speed should have been able to process enough object information during fixation so that higher processing speeds would not come into effect.

However, participants with high processing speed showed increased confidence in their recognition memory judgments. Even though a high processing speed did not improve recognition memory accuracy for sampled objects, it seems that it increased confidence for recognition judgments regarding those objects. Possibly, with the lack of more objects to sample during one fixation a high C value leads to repeated sampling of the same object. This could in turn lead to a slight strengthening of memory traces for objects sampled at a higher rate thus increasing confidence while leaving accuracy of the recognition memory judgments unaffected. Alternatively, fast encoding due to a high C value may allow active maintenance processes to start earlier. Consequently, a stronger memory trace could have been generated.

The finding of increased confidence, but unaffected recognition memory accuracy for high processing speed might lend support for a view of transsaccadic memory which goes beyond the dichotomous VSTM/VLTM model (e.g., Hollingworth, 2004; 2005). In several studies, Melcher and colleagues have found no evidence for a direct transition from VSTM to VLTM (Melcher, 2001; Melcher & Kowler, 2001) arguing for a "proto-LTM" or

"medium-term memory" in which detailed information can be available for a period of time exceeding VSTM, but then fails to be consolidated into VLTM due to, for example, task irrelevance (Melcher, 2006). From this perspective, information stored in a medium-term memory would leave memory performance unaffected, but could influence a more subtle feeling of confidence. Thus, processing speed might specially influence this transition stage of transsaccadic memory. The present data will not be able to fully resolve this issue, but might provide new tools to further investigate the built-up of scene memory across saccades and time.

Conclusion

In sum, the findings of the study presented here give further evidence to the claim that detailed visual information can be stored across saccades and that transsaccadic memory consists of at least VSTM and VLTM components (see Hollingworth, 2004; Hollingworth & Henderson, 2002; Hollingworth, Williams, & Henderson, 2001). By combining TVA measures of visual processing efficiency such as VSTM storage capacity and visual perceptual processing speed with transsaccadic memory data for objects in complex scenes, we were able to give further evidence that especially the VSTM capacity plays a major role in the generation of transsaccadic visual representations of naturalistic scenes. The main aim of the study presented here was to introduce a new approach to the investigation of transsaccadic scene memory, which might be able to shed new light on the processes involved when constructing a meaningful representation of the world. A next step would be to examine whether cognitive processes described by TVA parameters could also predict eye movement control mechanisms during real-world scene viewing. In this issue, Tatler and Vincent, for example, show that eye movement control in scene viewing can be better explained by including systematic tendencies of occulomotor control during the inspection of scenes. It would be interesting to see whether such hard-wired occulomotor tendencies are modulated by individual differences in processing efficiency. We believe that the combination of the TVA and eye movement research will provide interesting new insights into the study of transsaccadic scene memory.

APPENDIX A: Sample Scenes

kitchen scene

office scene

References

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual Working Memory Represents a Fixed Number of Items Regardless of Comlexity. *Psychological Science*, *18(7)*, 622-628.

Awh, E., Vogel, E. K., & Oh, S.-H. (2006). Interactions between attention and working memory. *Neuroscience*, *139*, 201-208.

Becker, M. W., & Pashler, H. (2002). Volatile visual representations: Failing to detect changes in recently processed information. *Psychonomic Bulletin & Review*, *9(4)*, 744-750.

Bublak, P., Finke, K., Krummenacher, J., Preger, R., Kyllingsbaek, S., Müller, H. J., & Schneider, W. X. (2005). Usability of a theory of visual attention (TVA) for parameter-based measurement of attention II: evidence from two patients with frontal or parietal damage. *Journal of the International Neuropsychological Society*, *11*, 843-854.

Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, *97*, 523-547.

Bundesen, C. (1998). A computational theory of visual attention. *Philosophical Transactions of the Royal Society of London Series B - Biological Sciences*, *353*, 1271-1281.

Bundesen, C., Habekost, T., & Kyllingsbaek, S. (2005). A neural theory of visual attention: bridging cognition and neurophysiology. *Psychological Review*, *112*, 291-328.

Castelhano, M., & Henderson, J. M. (2005). Incidental visual memory for objects in scenes. *Visual Cognition*, *12(6)*, 1017-1040.

Castelhano, M., & Henderson, J. M. (2007). Initial Scene Representations Facilitate Eye Movement Guidance in Visual Search. *Journal of Experimental Psychology: Human Perception and Performance*, *33(4)*, 753-763.

Cowan, N. (1995). *Attention and Memory: An Integrated Framework*. New York: Oxford University Press.

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioral Brain Sciences*, *24*, 87-114.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193-222.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*, 433-458.

Duncan, J., Bundesen, C., Olson, A., Humphreys, G., Chavda, S., & Shibuya, H. (1999). Systematic analysis of deficits in visual attention. *Journal of Experimental Psychology: General*, *128(4)*, 450-478.

Finke, K., Bublak, P., Dose, M., Müller, H. J., & Schneider, W. X. (2006). Parameter-based assessment of spatial and non-spatial attentional deficits in Huntington's disease. *Brain*, *129*, 1137-1151.

Habekost, T., & Rostrup, E. (2007). Visual attention capacity after right hemisphere lesions. *Neuropsychologia*, *45(7)*, 1474-1488.

Henderson, J. M, & Hollingworth, A. (2003). Eye movements, visual memory, and scene representation. In M. A. Peterson & G. Rhodes (Eds.), Analytic and holistic processes in the perception of faces, objects, and scenes (pp. 356-383). New York: Oxford University Press.

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.

Hollingworth, A. (2005). The relationship between online visual representation of a scene and long-term scene memory. *Journal of Experimental Psychologe: Learning, Memory, & Cognition*, *31(3)*, 396-411.

Hollingworth. A. (2006). Visual memory for natural scenes: Evidence from change detection and visual search. *Visual Cognition*, *14*, 781-807.

Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 113-136.

Hollingworth, A., Williams, C. C., & Henderson, J. M. (2001). To see and remember: Visually specific information is retained in memory from previously attended objects in natural scenes. *Psychonomic Bulletin & Review*, *8(4)*, 761-768.

Horowitz, T. S., & Wolfe, J. M. (1998). Visual Search has no memory. *Nature*, *394*, 575-577.

Humphrey, K. & Underwood, G. (2008). Fixation sequences in imagery and in recognition during the processing of pictures of real-world scenes. *Journal of Eye Movement Research, 2*(2): 3, 1-15.

Hung, J., Driver, J., & Walsh, V. (2005). Visual selection and posterior parietal cortex: effects of repetitive transcranial magnetic stimulation on partial report analyzed by Bundesen's theory of visual attention. *Journal of Neuroscience*, 25(42), 9602-9612.

Irwin D., E. (1992a). Memory for position and identity across eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 307-317.

Irwin D., E. (1992b). Visual memory within and across fixations. In K. Rayner (Ed.), Eye movements and visual cognition: Scene perception and reading (pp. 146-165). New York: Springer-Verlag.

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Bermann, M. G., & Moore, K. S. (2008). The Mind and Brain of Short-Term Memory. *Annual Review of Psychology*, *59*, 193-224.

Kyllingsbaek, S. (2006). Modeling visual attention. *Behavior Research Methods*, *38(1)*, 123-133.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279-281.

Melcher, D. (2001). Persistence of visual memory for scenes. *Nature*, *412*, 401.

Melcher, D. (2006). Accumulation and persistence of memory for natural scenes. *Journal of Vision*, *6(1)*, 8-17.

Melcher, D., & Kowler, E. (2001). Visual scene memory and the guidance of saccadic eye movements. *Vision Research*, *41*, 3597-3611.

Melcher, D., & Morrone, M. C. (2003). Spatiotopic temporal integration of visual motion across saccadic eye movements. *Nature Neuroscience*, *6*, 877-881.

O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology*, *46*, 461-488.

O'Regan, J. K., Rensink, R. A., & Clark, J. J. (1999). Change blindness as a result of "mudsplashes". *Nature*, *398*, 34.

Rensink, R. A. (2000). Seeing, sensing, and scrutinizing. *Vision Research*, *40*, 1469-1487.

Rensink, R. A. (2002). Change detection. *Annual Review of Psychology*, *53(1)*, 245-277.

Schneider, W. X. (1995). VAM: A Neuro-Cognitive Model for Visual Attention Control of Segmentation, Object Recognition, and Space-based Motor Action. *Visual Cognition*, *2*, 331-375.

Schneider, W. X. (1999). Visual-spatial working memory, attention, and scene representation: A neuro-cognitive theory. *Psychological Research*, *62*, 220-236.

Shibuya, H., & Bundesen, C. (1988). Visual Selection From Multielement Displays: Measuring and Modeling Effects of Exposure Duration. *Journal of Experimental Psychology: Human Perception and Performance*, *14(4)*, 591-600.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs*, *74*.

Tatler, B. W. & Vincent, B. T. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research, 2*(2): 5, 1-18.

Tatler, B. W., Gilchrist, I. D., & Rusted, J. (2003). The time course of abstract visual representation. *Perception*, *32(5)*, 579-592.

Tatler, B. W., Gilchrist, I. D., & Land, M. F. (2005). Visual memory for objects in natural scenes: From fixations to object files. *The Quarterly Journal of Experimental Psychology*, *58A(5)*, 931-960.

Todd, J. J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature*, *428*, 751-754.

Võ, M. L.-H., & Schneider, W. X. (submitted). A glimpse is not a glimpse: Differential processing of flashed scene previews leads to differential target search benefits while leaving visual object memory unaffected. Manuscript submitted for publication.

Zelinsky, G. J., & Loschky, L. C. (2005). Eye movements serialize memory for objects in scenes. *Perception & Psychophysics*, *67(4)*, 676-690.