Running head: Dwelling in Visual Search

Dwelling on Simple Stimuli in Visual Search

Gernot Horstmann

Bielefeld University

Stefanie I. Becker

University of Queensland

Anna Grubert

Durham University

Address: Gernot Horstmann

Department of Psychology

Bielefeld University

PO-Box 100 131

33 501 Bielefeld, Germany

Phone: (49) 521 - 106 12391

Fax: (49) 521 - 106 6422

E-mail: gernot.horstmann@uni-bielefeld.de

Abstract

Research and theories on visual search often focus on visual guidance to explain differences in search. Guidance is the tuning of attention to target features and facilitates search because distractors that do not show target features can more effectively ignored. As a general rule, the better the guidance is, the more efficient is search. Correspondingly, behavioral experiments often interpreted difference in efficiency as reflecting varying degrees of attentional guidance. But other factors such as the time spent on processing a distractor (dwelling) or multiple visits to the same stimulus in a search display (revisiting) are also involved in determining search efficiency. While there is some research showing that dwelling and revisiting modulate search times in addition to skipping, the corresponding studies used complex naturalistic and category-defined stimuli. The present study will test whether results from prior research can be generalized to more simple stimuli, where target-distractor similarity, a strong influence on search performance, can be manipulated in a principled way. Thus, in the present study, simple stimuli with varying degrees of target-distractor similarity were used to deliver conclusive evidence for the contribution of dwelling and revisiting to search performance. The results have theoretical and methodological implications. They imply that visual search models should not treat dwelling and revisiting as constants across varying levels of search efficiency and that behavioral search experiments are equivocal with respect to the responsible processing mechanisms underlying more versus less efficient search. We also suggest that eye-tracking methods may be used to disentangle different search components such as skipping, dwelling, and revisiting.

Dwelling on Simple Stimuli in Visual Search

Introduction

Finding a target among distractors can be easy or difficult, depending on the properties of the stimuli. It is easy in efficient search (also known as pop-out search), where the target is found at a glance, and where adding non-targets (distractors) to the search field does not affect search times. It is difficult in inefficient search, where a considerable amount of time is spent on checking distractors before the target is finally detected, so that search times are slower with more than with less distractors. According to Treisman's seminal Feature Integration Theory (FIT, Treisman, 1985; Treisman & Gelade, 1980), efficient or pop-out search has been attributed to the use of the output of a parallel feature extraction stage, whereas inefficient search indicates the involvement of a capacity-limited attentional stage that operates at least partly in a serial manner. This conception has been elaborated in more recent models of visual search, such as Guided Search (GS; Wolfe, Cave, & Franzel, 1989; Wolfe, 2004, 2007), the Target Acquisition Model (TAM, e.g., Zelinsky, 2008), the Dimensional Weighting Model (DW, e.g., Found & Müller, 1996) or the Saliency Model (IS, Itti & Koch, 2000, 2001), all of which explain varying levels of search efficiency in particular by differences in *target guidance*. GS2 (Wolfe, 2004) for example, suggests that during an initial (bottom-up) stage of visual processing, the visual field is decomposed into isolated spatial maps of basic features for color, orientation, luminance, etc. The output of these separate maps is spatiotopically organized and summed up to be represented in a single spatiotopic map of activation. To enable guidance, searched-for features are boosted on the activation map. The profile of the activation map schedules sequential shifts of attention within the visual field towards conspicuous locations, the attentional visiting of which is necessary for the final decision whether a stimulus at a particular location is the target or not. A single high peak in the activation map leads to a fast attention shift to the target’s location, and search is efficient. However, if some of the distractors share features with the target, multiple peaks arise in the activation map. Because there is also inherent noise in the system, the target location may in this case not always have the highest activation peak on the map. Accordingly, multiple stimuli in the display, sometimes even all of them, may be visited before the target is finally found, rendering search inefficient. Note that this view predicts that search efficiency is a continuum rather than a dichotomy. Depending on the signal-to-noise ratio, search is predicted to be more or less efficient. This is reflected by the fact that search slopes, that is, the slopes of the functions relating reaction times to set size (i.e., the number of stimuli in a display), can vary between values around zero and several hundred milliseconds per item (Wolfe, 1998).

Guidance based approaches to visual search are elegant and attractive because attentional guidance is effectively the only variable needed to explain search efficiency. These models predict that in efficient search, the ratio of activation for the target versus the distractors is very high, such that many of the distractors are not considered target candidates and will never be checked (skipping of distractors). Thus, prototypical efficient (pop-out) search is fast, because all distractors are skipped and only one stimulus, the target, is attended. Prototypical inefficient search is slow when all stimuli are considered target candidates, because attentionally investigating each of them is time consuming. Intermediate levels of efficiency result from possible, but imperfect guidance by the target: the activation ratio for target against distractors is lower, and noise inherent in the activation map leads to the selection of some of the distractors in some of the trials. Hence, the target is not found as the first item, but it is also not found as a result of a random sampling of stimuli. Rather a weak guidance signal renders search better than expected by chance. How much better, in turn, depends on the strength of the guidance signal.

On reflection, it is clear that other selection mechanisms should contribute to search difficulty (or its inverse: search efficiency) as well (Treisman & Souther, 1985; Wolfe, 2001). The first of such mechanisms which we consider here is the time spent checking candidate target items or *dwell time*. It has been acknowledged that irrespective of whether a single stimulus (Wolfe, 1989; Zelinsky, 2008) or groups of stimuli (Hulleman & Olivers, 2016) are attentionally investigated, the duration of this investigation process may be an effective cause rendering some searches inefficient, and that search is more efficient when only a short period of time is spent at the respective stimulus locations (e.g., Horstmann, Ansorge, & Scharlau, 2006; Horstmann, Herwig & Becker, 2016; Horstmann, Becker & Ernst, 2017; Wolfe, 2011; Wolfe & Horowitz, 2017). Correspondingly, search is less efficient if, for whatever reasons, attentional checking time is increased in a particular type of search. In some sense this is also an obvious interpretation of search slopes, which is measured as the rise in milliseconds over added distractors to the search display: time spent checking each (additional) item. Note, however, that this is not the only interpretation of search slopes, because search slopes can also be interpreted as indicating the (average) number of distractors selected during search, which are selected with a constant search rate (e.g., Chun & Wolfe, 1996) .

The influence of attentional checking time on search time is no secret and has been observed before (e.g., Hout et al., 2017; Walenchok et al., 2016; Wolfe, 2001; see Wolfe, 2018, for a comprehensible summary). However, guidance-based models tend to treat selection rate (e.g., Chun & Wolfe, 1996), or dwell time (Zelinsiky, 2008), as a constant that does not change for different search types. This does not necessarily mean that the processing requirements for stimuli are assumed to be the same for all variants of visual search. For example, Wolfe (2003; 2007) assumes parallel processing after selection (see also Venini et al., 2014) to reconcile the assumption of fast and constant selection rates in the range of 50 ms/item with findings that attentional dwell time is more in the range of 250-500 ms (e.g., Duncan et al., 1994). Yet current computational models of guided visual search such as GS or TAM, fix this time to a constant (see, for a different treatment, the CRISP model by Nuthmann, Smith, & Engbert, 2010, which is however not a model of guided visual search).

Treating attentional selection rate or dwell time as a constant implies that it is not useful in explaining differences in search. Moreover, yet another factor that might modulate search efficiency is the number of revisited stimuli in the visual field, i.e., stimuli that are checked multiple times. The classical notion in visual search models is that already visited locations are tagged in some way to ensure that each stimulus is investigated once and ideally only once (although revising has been suggested previously, e.g., by Horowitz & Wolfe, 1998). One possible way of implementing such a tagging is inhibition of return (IOR, Posner, Rafal, Choate, & Voaughan, 1985), where the locations on the activation map that correspond to already visited locations in the visual field, are transiently suppressed (Itti & Koch, 2000; Shipp, 2004; Wolfe, 1989; Zelinsky, 2008). Alternatively, already investigated locations in the visual field may be stored in a visuo-spatial memory store (i.e., VSTM, cf. Hulleman & Olivers, 2017) and prevent revisiting via higher level processes concerned with choosing the target for the next fixation. Unless memory is perfect, capacity limitations and storage decay are responsible for revisiting already scanned locations and thus can explain a decrease of search efficiency. Capacity limitations, if relevant, should become more important as set size increases, while memory decay should become a determining factor for less efficient search with increasing time elapsed between stimulus onset and response.

Focusing on the single principle of guidance to explain many aspects of visual search performance adheres to the criterion of parsimony in theory building. Moreover, letting some factors vary while fixing others to constants is obviously wise when beginning to build computational models and explore their behavior through simulations and experiments. Scientific models often seek complexity reduction and include only a few important variables. One might even argue that dwelling (the time spent on checking a stimulus) and revisiting (the frequency of repeated checking) only affect search efficiency in such a minor way that it can hence be ignored. However, there is still a lot of variation in search efficiency in highly inefficient search, when guidance is difficult or absent, indicating that dwelling and revisiting may be important variables to explain visual search efficiency. In a series of previous studies we have used eye-tracking to measure dwell time and revisiting most directly (Horstmann & Becker, 2019; Horstmann, Herwig, & Becker, 2016; Horstmann, Becker, and Ernst, 2017; see also Horstmann, Lipp, & Becker, 2012). Eye tracking was used to assess whether a distractor was looked at or not (skipping), and if so, for how long it was looked at during the first examination (dwelling) and whether the same distractor was looked at repeatedly (revisiting). Targets that were either similar or dissimilar to the distractors were used to induce different levels of search efficiency. We found that target-distractor similarity led to increased dwell time, which in turn increased search duration as indicated by RT. Correspondingly, distractors were more often revisited when searching for a similar (difficult) target than when searching for a dissimilar (easy) target.

These previous studies used naturalistic face stimuli, which is a drawback for two reasons. First, visual search studies are usually conducted with well controlled laboratory stimuli. While the importance of using more ecologically relevant common stimulus categories (e.g., faces or scenes) in studies of visual search have been substantiated recently (e.g., Alexander & Zelinsky, 2012; Einhäuser & Nuthmann, 2016), research has revealed differences in the processing of artificial and realistic stimuli (e.g., Neider & Zelinsky, 2006; Zelinsky & Schmidt, 2009; Jenkins, Grubert, and Eimer, 2018). Second, while we successfully induced differences in search efficiency with different target categories in the previous studies, with natural stimuli it is not possible to actively manipulate theoretically important dimensions, in particular, target-distractors similarity (but see Hout & Goldinger, 2010; Hout et al., 2016, for procedures to measure target-distractor similarity).

The aim of the present study was to test whether our previous results for faces generalize to more standard laboratory stimuli, and to manipulate target-distractor similarity in a systematic way while using artificial stimuli. To that aim, we adapted search stimuli from an eye-tracking study by Reingold and Glaholt (2014), which seemed ideal for our paradigm as they were already designed to manipulate target-distractor similarity. As illustrated in Figure 1 (top panel), these stimuli were irregular shapes where the target differed from the distractors (bottom row) either only in the spatial organization of one individual stimulus part (similar targets where in comparison to the distractor stimuli, the left stimulus arm is shifted downwards along the attaching vertical bar; middle row), or - in addition - in the orientation of the entire configuration (dissimilar targets which are mirror versions of the similar targets; top row).

- Figure 1 about here -

In Experiment 1, we produced ten variations (distortions) of the basic stimulus shapes by replacing some pixels of the stimulus figures with a different color (Figure 1). This was done to mirror the stimulus heterogeneity of our previous studies where we used facial images of ten individuals (rather than ten times the same face). Experiment 1 was therefore designed to allow for a direct comparison of potential effects caused by the stimulus material (heterogeneous faces vs. heterogeneous simple stimuli) rather than any other changes in the methods or materials. To test whether the artificially introduced distractor heterogeneity affected the results of Experiment 1, we conducted Experiment 2, in which we used the three original stimulus types without any distortions. In Experiment 3, we varied the set size to assess search efficiency in a more traditional way with RT increase as a function of increased stimulus number. Experiment 4 was designed to test whether the previously observed effects of not only skipping, but also dwelling and revisiting, would hold under search conditions which very likely promote strong guidance.

We expect target-distractor similarity to influence search times, both in target-present and target-absent trials. If the effect of similarity on search times is exclusively due to guidance, we would expect more skipping with dissimilar than similar targets, whereas dwelling and revisiting should be unaffected by similarity. More skipping in target-present trials would be because the dissimilar target has a higher probability of being selected early than the similar target. More skipping in target-absent trials would depend on a stopping rule that is correlated to the guidance in present trials (Chun & Wolfe, 1996). In short, in target-absent trials, the activation map may contain multiple activations proportional to the similarity of the distractors with the searched for target, and in addition, noise. Search is stopped when the remaining activation falls below some threshold, under which activation is treated as effectively zero. The threshold is set to an optimal value, just high enough that the similar or dissimilar target, respectively, if present, is always above (note that this only applies to a design in which similar and dissimilar targets are presented in a blocked fashion). Because good guidance implies that the target has a much higher activation than a distractor, the threshold is well above zero in easy search. Because weak guidance implies that the target only statistically has a higher value than the distractors, the threshold is near zero in difficult search. In contrast, if the effect of similarity on search times is *not* exclusively due to guidance, but if dwelling and revisiting contribute to the effect of similarity on search times, we would expect to observe not only increased skipping rates, but also increased dwell times and revisiting rates when target-distractor similarity is high.

To validate our hypotheses and predictions that not only skipping, but also dwelling and revisiting are reliable determinants of search times, we traced the effects of these three variables by means of correlations and multiple regression analyses in which any potential impact will be registered in substantial regression weights. We focused on target-absent trials (cf. Horstmann et al., 2016, 2017), because they allow observations of distractor processing without interference from target-related processes. However, analyses for target-present trials are also reported for completeness. Note that it is not our aim to challenge guidance (here substantiated in skipping rates) as an important contributor to search efficiency in general. Rather, we aim to test whether differences in selection mechanism that are not directly related to guidance can modulate search efficiency in a non-trivial, substantial way.

Experiment 1

Methods

Participants. Sixteen students with normal or corrected to normal vision participated in the study. Each received €4 for their 30-minute participation. Three participants were excluded from analysis because of near-chance performance in the similar target-present condition. Mean age of the remaining 13 participants was 23.3 years (SD = 2.3 years); 10 were female. The study was approved by the Bielefeld University's ethics committee and performed in accordance with the approved guidelines.

Stimuli. Figure 1 A shows the stimuli used in Experiment 1. The basic shapes for the distractors (bottom row) and the similar targets (middle row; same as distractors, only that the right arm is shifted towards the bottom of the connecting vertical bar) were adapted from a publication of Reingold and Glaholt (2014). Their dissimilar targets, a C-shaped configuration, differed too much from the distractors for our purposes, and so we chose a horizontally flipped version of the similar target instead (top row). All stimuli were gray and were presented on a white background; each subtended 74x74 pixels. For Experiment 1, the basic shapes were distorted by randomly selecting seven positions within the inner 69x69 pixel matrix of a stimulus and flipping the grey foreground with blue. The selected locations to flip color each measured 11x11pixels, centered on a randomly selected position within the stimulus. No restrictions were applied to randomization (e.g., distortions were allowed to overlap), so that some of the distorted stimuli differed more from the basic shape than others.

Search displays consisted of ten stimuli presented at ten randomly selected locations from an imaginary grid of 5 horizontal ×3 vertical locations. Figure 1 B shows two examples of target-present trials, one with a similar target (left panel) and one with a dissimilar target (right panel). The central position of the grid was excluded, as it contained the fixation marker in the pre-stimulus display. Center to center distances of the grid position were 100 pixels (2.8° of visual angle) horizontally and 130 pixels vertically (3.6°). Each stimulus position on each trial was randomly jittered by 5±5 pixels horizontally and vertically. The fixation marker was the standard fixation stimulus for the SR-1000 eye tracker (a black disk with a small white center).

Apparatus. Stimuli were presented on a 19-inch display CRT-monitor (100-Hz refresh rate, resolution 1,024×768 pixels) at a viewing distance of 71 cm. A video-based tower-mounted eye tracker (EyeLink 1000, SR Research, Ontario, Canada) with a sampling rate of 1,000 Hz was used for the recording of eye movements. Participants’ heads were stabilized by a chin and forehead rest, and for all participants, the right eye was monitored. A 9-point eye tracker calibration was used before the start of the experiment. Stimulus presentation and manual response collection was programmed using Experiment Builder 1.10.165 (SR Research, Ontario, Canada). Eye tracking data were preprocessed using Data Viewer 2.2.1 (SR Research, Ontario, Canada).

Design. The experiment comprised of six experimental blocks with 20 trials each. There were three alternating blocks for each of the two target categories - similar versus dissimilar targets. Half of the participants started with a similar target block. Each block contained ten target-present and ten target-absent trials. Target-absent trials displayed ten distractors, which means that target-absent trials in the similar and dissimilar target condition were structurally identical. In target-present trials, one of the distractors was randomly chosen to be replaced with one out of the ten possible target shapes of the respective target category (similar vs. dissimilar targets). The target in each trial was selected pseudo-randomly, so that each individual target shape appeared equally often in each block. Before the experiment proper participants completed a 20-trial practice block, which did not enter analysis. The target category in this practice block was always different from the first experimental block.

Procedure. Each trial started with a fixation control, which was terminated with a left-hand key press that initiated the presentation of the search display. Participants' task was to indicate with a right-hand (index or middle finger) key press whether or not one of the ten possible target shapes was present in the search display. The search display was shown until a manual response was registered. A short beep was issued in case of an error. Prior to each block, the ten possible targets of the respective target category (similar vs. dissimilar) and the ten possible distractors were displayed until the participant initiated the start of the first trial. The experimental blocks were preceded by two practice blocks, one for each target, which were not analyzed. Instructions emphasized speed and accuracy.

Data preprocessing. Raw eye position data were parsed by the eye tracker software's standard experimental settings which used a speed threshold (30°/s) and an acceleration threshold (8,000°/s^2) for saccade detection. Rectangular 75x75 pixels areas of interest (AOIs) were defined that enclosed the stimulus shapes; outlier fixations were assigned to the nearest AOI. From these preprocessed data, four variables were derived for analysis. Each stimulus was classified as being fixated within a given trial or not. If a stimulus was fixated, dwell time was assessed, which is the sum of the fixation durations over the first continuous series of fixations on that stimulus. Of note, this measure often includes the duration of a single fixation, but in case several fixations were made on the stimulus, for example due to corrective saccades, the additional time was added. Please note also, that only gaze duration during the first continuous visit was used here, dismissing the gaze duration of possible revisits. This was done to avoid confounding the measures of dwelling and revisiting. Furthermore, we recorded whether a stimulus was visited only once, or whether it was revisited, that is, selected repeatedly during a trial after the first continuous run of fixations. A fixation was scored as a revisit if (a) the stimulus had been fixated before and (b) the last fixation of that stimulus was interrupted by at least one off-stimulus fixation. The basic variables of our analysis, however, were trial statistics (i.e., statistics for each trial). *Skipping* is defined as the proportion of stimuli that had not been fixated at all in a trial. Skipping is the variable which drives trial RTs, as assumed by guidance-based theories of visual search. *Dwelling* is the average dwell time in a given trial. We predict dwelling to be affected by similarity, whereas guidance-based theories treat dwelling as a constant. Also, we predict dwelling to substantially modulate trial RTs, i.e., to govern differences in search performance across trials. In addition to dwelling and skipping, we also assessed the proportion of stimuli that had been revisited, because *Revisiting* is a third possible source of variance in RTs. RT was measured as the time elapsed between display onset and key press of a correct answer.

Before analysis, raw data for the measures of time (i.e., RT, dwell times, and stimulus selection latencies) were filtered for outliers. Measures of time were log transformed before outlier analysis. Data points were identified as outliers when they exceeded the mean of their respective condition (target presence x similarity) by two standard deviations or more. In addition, a lower cutoff was used for RTs (300ms) and dwell times (40ms). Outlier analysis was performed separately for each participant. As a result of the analysis, 36 RTs, 298 dwell times, and 17 selection latencies were removed.

Results

In the presentation of the results, we will proceed as follows. First, we will present error rates and manual RTs (Figure 2, left column), along with the corresponding ANOVAs, to assess whether there was an effect of target-distractor similarity on overall search performance. Secondly, we will show correlations and multiple regression analyses, where trial-based search times were regressed on the trial statistics for distractor dwelling, skipping, and revisiting, to indicate whether all these three underlying mechanisms are substantial predictors for increased search times (as indexed by RTs) in similar relative to dissimilar target search. The respective means (panels a, b, and c) are depicted in Figure 3, top row. We will not present ANOVAs as we did in the previous studies (e.g., Horstmann, Becker & Ernst, 2017), because the central evidence, that is the effects of similarity on skipping, revisiting, and in particular, dwelling, is captured in the correlations between similarity on the one hand, and skipping revisiting, and dwelling on the other. Our main focus is on target-absent trials, because only in these trials skipping, dwelling, and revisiting can be observed independently of the processes that may lead to the selection of the target in target-present trials.

-- Figure 2 about here –

Error rates. Mean proportion correct on target-absent trials was .98 and .90 in blocks with dissimilar and similar targets, respectively, and .94 and .88 on target-present trials in those blocks. A repeated measures ANOVA with the factors target-distractor similarity (similar vs. dissimilar) and target presence (present vs. absent) revealed a significant main effect for presence, $F(1,12) = 123.99$, $p < .001$, $\eta_G^2 = .48$, and target-distractor similarity, $F(1,12) = 6.94$, $p = .021$, $\eta_G^2 = .20$, but no reliable interaction, F<1. All trials with errors were excluded from further analyses.

Reaction times. Figure 2 (left column) displays mean correct RTs, separately for target-absent and target-present trials in blocks with dissimilar and similar targets, respectively. A repeated measures ANOVA computed over mean correct RTs with the factors target-distractor similarity (similar vs. dissimilar) and target presence (present vs. absent) revealed significantly shorter RTs on target-present than target-absent trials, $F(1,12) = 168.56$, $\eta_G^2 = .38$, and in blocks with dissimilar as compared to similar targets, $F(1,12) = 33.42$, $\eta_G^2 = .31$; both *p*s < .001. The interaction just failed to be significant, F(1,12) =3.97, p=.069.

-- Figure 3 about here –

Contributions of distractor dwelling, skipping, and revisiting to increased search times

Search times are basically the product of the number of fixations and their average duration, and therefore it is reasonable to expect that skipping, dwelling, and revisiting together would predict search times very well. The crucial question is therefore not whether, but to which degree these variables predict RT. Table 1 (top panel) presents the bivariate correlations between manual RTs, the trial-based rates for revisiting and skipping, and the dwell times, separately for target-absent and target-present trials. The unit of analysis is the trial, and each participant provides measures from up to 120 trials (minus the trials where outliers were detected). Note that the bivariate correlations (first column) between similarity on the one hand and RT, dwelling, skipping, and revisiting on the other hand, correspond roughly to a test of the mean differences between the levels of similarity (i.e., the main effect of similarity), which would normally be done using a t-test. These significant correlations show that similarity impacts on RT, and also on dwelling, skipping, and revisiting.

The second column is informative, as it shows whether and how strongly dwelling, skipping, and revisiting co-vary with RTs. Figure 4 complements Table 1 with the respective scatterplots on target-absent trials, separately for low (left panel) and high (right panel) target-distractor similarity trials. The forth (bottom) row of Figure 4, for instance, shows the bivariate distributions of trial dwelling, trial skipping, trial revisiting, and trial RT for dissimilar (left) and similar (right) target blocks. Each dot represents one trial. The deep blue region represents the concentration of most trials. However, independently of the scatter, the positive linear regression slope of the main axis through the data points can clearly be visualized. Figure 4 also reveals that the linear relations between the variables are similar for blocks with similar and dissimilar targets. The same holds true for skipping and revisiting (middle and right plots of the fourth row). In contrast, dwelling showed no clear relationships to skipping (second row) and revisiting (third row, left plot), respectively. Finally, skipping and revisiting are slightly negatively correlated (third row, right plot), as the region above the diagonal is virtually devoid of data points. This has purely technical reasons and is due to the definition of revisiting as a per stimulus/trial statistic. Because a skipped distractor cannot be revisited, as skipping rates go up, revisiting rates go down. Note that due to the large number of observations; even small coefficients are significantly different from zero. The correlations should be thus evaluated with respect to their size, not their statistical significance.

The correlations between RT and dwelling, skipping, and revisiting, respectively, cannot easily be interpreted because the predictors are themselves correlated. Accordingly RT was regressed on dwelling, skipping, revisiting, and target-distractor similarity as predictors in order to obtain their statistically unique effects (i.e., when the other variables are statistically held constant). The binary variable target-distractor similarity was included to gauge the variance in RT that was caused by target-distractor similarity, but was not transmitted to RT via our main predictor variables. This includes, inter alias, interactions between the predictor variables, which are not included in this simple linear model.

-- Table 1 about here --

-- Figure 4 about here –

We used a linear multilevel regression with random intercepts to disentangle within-subject variations from between-subject variations in dwelling, skipping, revisiting, and RT. Metrical variables were *z*-transformed prior to analyses in order to make regression coefficients comparable among each other. With standardized coefficients, b = .5 means that when the independent variable is increased by one standard deviation, the dependent variables increases for half a standard deviation. For the experimental factor of target-distractor similarity, dissimilar was coded as zero and similar as one. Regression coefficients are tested against zero with t-tests. We interpreted empirical *t*-values exceeding a value of ±1.96 as significantly ($p<.05$) differing from zero.

Target-absent trials. Table 2 displays the results on target-absent trials, based on 744 observations. The *variance inflation factor* (*VIF*) was used to guard against collinearity among the predictor variables. It was acceptable with tolerances (i.e., 1/*VIF*) within > .70 (well above the critical tolerance level of .1). All predictors (i.e., dwelling, skipping, revisiting) were found to have a significant effect on target-absent RTs. The effect of dwelling was strongest, followed by revisiting and then skipping (see the second column, where the regression slope, indicated by coefficient b, is shown). Marginal R^2 was .90 (Nakagawa & Schielzeth, 2013). The effect for target-distractor similarity was significant but small, indicating that target-distractor similarity had little unique influence on RTs after the common variance of dwelling, skipping, and revisiting had been removed.

-- Table 2 about here --

Target-present trials. Only distractor fixations were analyzed for target-present trials to be consistent with the previous analysis of target-absent trials. Table 2 shows the results on target-present trials, which were based on 701 observations. Indications of collinearity were low, with all computed tolerances (1/*VIF*) > .75. All regression slopes were significant (see *t* -value). Skipping had the largest impact on target-present RTs, which is not surprising because the number of distractors that have been inspected before the target varies between zero and nine (all distractors in the display, ignoring possible revisits). Dwelling and revisiting had roughly the same beta weights, i.e., they can be considered to be equally important for the prediction of search RTs on target-present trials. Marginal R^2 was .79.

Discussion

Experiment 1 replicated previous findings (Horstmann, et al., 2016, 2017) and confirmed that in addition to skipping (the proportion of distractors that are excluded from inspection), dwelling (the time gaze is focused on a distractor) and revisiting (the frequency with which distractors are included for inspection repeatedly) were important predictors for search times. This is a notable finding because it shows that our previous observations are not limited to visual search for complex and naturalistic (face) stimuli. There are three main findings of Experiment 1. First, target distractor-similarity increases search times as expected, which is indicated by the significant correlation between similarity and RT, and the significant main effect of similarity in the ANOVA. The effect is actually quite strong, with search among ten items being almost a second longer with the similar target. Second, similarity increases dwelling and revisiting, and decreases skipping; these effects are also indicated by the corresponding correlation coefficients, and are visually presented in Figure 3, upper panel. These effects can be observed in absent and present trials as well. Third, dwelling, revisiting, and skipping, respectively influence RT independently from each other, as indicated by the coefficients in the multiple regression. The effects of dwelling and revisiting are strong in absent trials, and are actually the strongest effects in this experiment. The effects of dwelling and revisiting are weaker in present trials, whereas the effect of skipping increases. We will discuss this in more detail after we have reviewed the results of Experiment 2.

Experiment 2

Methodological aspects of stimulus presentation, experiment timing, and response collection were identical to Experiment 1, with the important exception that in Experiment 2, we employed the undistorted prototypes of the stimuli used in Experiment 1 (i.e., the grey basic stimulus types presented in Figure 1 without the colored distortions). The motivation for the stimulus distortions in Experiment 1 was to introduce a certain degree of stimulus variance (distractor-heterogeneity, cf. Duncan & Humphreys, 1989) that was expected to approximate the differences in the stimulus material used by Horstmann et al. (2017, 2018), where target and distractor stimuli were images of 10 individual faces, naturally introducing a certain level of stimulus variability. However, adding noise to the shapes may have had its own effect on the results. Moreover, the classical experiments concerned with the basic mechanisms of visual search often used little distractor heterogeneity (e.g., Treisman & Souther, 1987; Wolfe, Cave, & Franzel, 1989). Thus, Experiment 2 was designed to narrow the gap between more typical experiments on search efficiency and the protocols used in our studies. Apart from that, it is interesting in its own right to see whether the result pattern is changed when the distractors are more homogeneous (Duncan & Humphreys, 1989). In general, we expect search to be more efficient with more homogeneous displays.

Methods

Participants. Fourteen students participated in the study. Each received €4 for their 30-minute participation. Two participants were excluded because of near chance performance in the dissimilar target condition. Mean age of the remaining twelve participants was 26.4 years (SD = 2.8 years); 6 were female. The study was approved by the Bielefeld University's ethics committee and performed in accordance with the approved guidelines.

Apparatus, Stimuli, & Procedure. All aspects of the stimulus parameters, the apparatus used for testing, the study design employed, the procedures applied, and eye tracking data preprocessing were identical to Experiment 1, with the only exception that in Experiment 2, stimuli were the undistorted prototype shapes. To reiterate the most relevant variables: Experiment 2 was tested in 6 blocks with 20 trials each, ten of which were target-present trials (search displays showed one target shape and nine distractor shapes), the other ten were target-absent trials (search displays contained ten distractor shapes). Similar vs. dissimilar targets were tested block-wise and alternatingly. Two practice blocks preceded the actual experimentation.

Data pre-processing. This was analogous to Experiment 1. Outlier detection led to the exclusion of 42 RTs, 274 dwell times, and 14 latencies.

Results

Error rates. Mean proportion correct on target-absent trials was .99 and .98 in dissimilar and similar target blocks, respectively, and .95 and .87 on target-present trials of the respective blocks. A repeated measures ANOVA with the factors target-distractor similarity (similar vs. dissimilar) and target presence (present vs. absent) revealed significant main effects for target presence, $F(1, 11) = 19.81$, $p = .001$, $\eta_G^2 = 0.64$, and target-distractor similarity, $F(1, 11) = 10.05$, $p = .009$, $\eta_G^2 = 0.48$, as well as a significant interaction $F(1, 11) = 5.21$, $p = .043$, $\eta_G^2 = 0.32$. While the mean proportion correct was at ceiling on target-absent trials, target-distractor similarity affected accuracy on target-present trials such that there were more errors in blocks with similar as compared to dissimilar targets.

Reaction times. Figure 2, middle column, shows mean correct RTs, separately for target-absent and target-present trials in blocks with dissimilar and similar targets, respectively. Those mean correct RTs were subjected to a repeated measures ANOVA with the factors target-distractor similarity (similar vs. dissimilar) and target presence (present vs. absent), which revealed significant main effects for target presence, $F(1, 11) = 134.83$, $p < .001$, $\eta_G^2 = 0.92$, and target-distractor similarity, : $F(1, 11) = 20.01$, $p = .001$, $\eta_G^2 = 0.65$. The interaction was also significant, : $F(1, 11) = 3.37$, $p = .093$ $\eta_G^2 = 0.23$, because RTs were faster on target-present than target-absent trials (1754 ms vs. 2727 ms), and in dissimilar than similar target blocks (2027 ms vs 2454 ms). As confirmed by two independent follow-up *t*-tests, the target-distractor similarity effect (i.e., slower RTs in similar relative to dissimilar target blocks) was more pronounced on target-absent trials, but was substantial on both target-present and target-absent trials, both $t(11) > 3.79$, $p < .003$.

Impact of similarity on dwelling, skipping, and revisiting, and contributions of distractor dwelling, skipping, and revisiting to search times

As for Experiment 1, multiple regression analyses were trial-based, and search times were regressed on the trial statistics for distractor dwelling, skipping, and revisiting, to investigate the predictive power of each of these mechanisms on search time. Figure 3, bottom panel, illustrates the respective means (panels a, b, and c). Table 2 presents the bivariate correlations between RTs and the rates for revisiting and skipping together with the dwell time, separately for target-absent and target-present trials. The first column of Table 1 (lower part) shows the correlations between similarity and RT, dwelling, skipping, and revisiting. All correlations were significantly from zero, and the structure from Experiment 1 is well replicated. The second column shows the correlations between RT and dwelling, skipping, and revisiting with substantial effects as well. The two final columns show that correlations among dwelling, skipping, and revisiting.

Target-absent trials. Table 2 (lower top panel) shows the statistical values obtained in the regression analyses based on 632 target-absent trials, reflecting the separate prediction power of dwelling, skipping, revisiting, and target-distractor for the trial RTs. Collinearity among the predictor variables was acceptable, with all $1/VIF > .75$. All regression slopes, except target-distractor similarity were significant. Marginal R^2 was .93.

Target-present trials. The lower bottom panel of Table 2 shows the respective values based on 635 target-present trials. Indications of collinearity were low, $1/VIF > .75$. As on target-absent trials, all regression slopes were significant, except for the one on target-distractor similarity. Marginal R^2 was .82.

Discussion

The most important result in Experiment 2 is that similarity has an effect on skipping, dwelling, and revisiting, and that these variables in turn have an effect on search times. This indicates that the general results pattern is observed independently from the presence (Experiment 1) or absence (Experiment 2) of distractor heterogeneity. The effects of similarity are generally somewhat weaker in the present homogeneous distractors, as indicated by the correlations between similarity and the measured variables (see Table 1, first column); however, all correlations are still substantial. In the multiple regressions, the regression slopes of dwelling were reduced in comparison to Experiment 1, in particular for the absent trials; however, all regression slopes are still substantial. Overall, RTs seem to be shorter in Experiment 2 than in Experiment 1 (see Figure 2), which is consistent with the assumption that distractor-homogeneity makes search easier. As the standard errors also happened to be different in the two experiments (see Figure 2), it seems that increasing distractor homogeneity makes search less variable. To sum up, the present results shows that dwelling and revisiting (together with skipping) significantly contributed to explaining the increased search RTs in similar versus dissimilar target blocks, irrespective of the lower distractor heterogeneity in Experiment 2.

Experiment 3

Experiment 3 included a variation of set size to test (with a more traditional approach) whether search for a dissimilar rather than a similar target is more efficient. For RT studies set size modulations are critical to separate the slope of the search function (indicator of the actual search mechanism) from the intercept (reflecting response selection processes). One could argue that this would not be necessary in our experiments for which we employed eye movements to directly assess the search slopes. However, search slopes are a common currency in visual search and for the sake of between-study comparability it is thus informative to test search efficiency in this way. Furthermore, it might be possible that the impact of dwelling, skipping, and revisiting changes with set size. For example, revisiting may occur more frequently with higher set sizes because the decay of VSTM might be increased (Hulleman & Olivers, 2017) or inhibition of return might be less effective with larger samples. We used ANOVAs to test whether dwelling, skipping, and revisiting are influenced by set size.

Participants. Sixteen students participated in the study. Each received €4 for their 30-minute service. Two participants were excluded because of an excess of errors in at least one condition (33% or more). Age from one participant was missing, mean age of the remaining participants was 26.15 years (SD = 1.91 years); 9 were female, 5 were male. The study was approved by the Bielefeld University's ethics committee and performed in accordance with the approved guidelines.

Apparatus, Stimuli, & Procedure. All aspects of the stimulus parameters, the apparatus used, the study design employed, the procedures applied, and the eye tracking data preprocessing were identical to Experiment 1 (with stimuli being randomly distorted), with two differences. First, set sizes of five and ten were used, both of which were presented intermixed within the same block of trials, doubling the number of trials per block to 40. Second, stimuli were presented in a 5x5 grid, such that for both set sizes the arrangement of stimuli would be irregular on most of the trials.

Data pre-processing. Data were preprocessed as in Experiments 1 and 2. Outlier screening led to the exclusion of 116 RTs, 693 dwell times, and 47 latencies. As before, the screening was done case wise and separately for each combination of the variables similarity (similar vs. dissimilar), target presence (absent vs. present), and set size (5 vs. 10). Areas of interest (AOI) for the eye tracking data were defined somewhat differently from before, as every cell of the 5x5 grid was a single AOI. This was done because otherwise (i.e., using the nearest AOI for outlier fixations) AOI size would have been confounded with set size.

Results

For Experiment 3, we report ANOVAs in addition to the correlations and the multiple regressions.

Error rates. Mean proportion correct on target-absent trials was .99 and .98 in dissimilar and similar target blocks, respectively, and .95 and .87 on target-present trials of the respective blocks. A repeated measures ANOVA with the factors target-distractor similarity (similar vs. dissimilar), target presence (present vs. absent), and set size (5 vs. 10) revealed significant main effects for target presence, $F(1, 13) = 30.3$, $p < .001$, $\eta_G^2 = 0.7$, target-distractor similarity, $F(1, 13) = 21.93$, $p < .001$, $\eta_G^2 = 0.63$, and set size, $F(1, 13) = 5.11$, $p = .042$, $\eta_G^2 = 0.28$. The Target Presence x Similarity interaction was significant, $F(1, 13) = 25.85$, $p < .001$, $\eta_G^2 = 0.67$, as was the Target Presence x Set size interaction, $F(1, 13) = 25.85$, $p < .001$, $\eta_G^2 = 0.67$, (all other *F*s < 1). Less errors were made in target-absent than target-present trials (.99 vs. .93), with similar than dissimilar targets (.98 vs .94), and higher than lower set size (.96 vs. 95). The Target Presence x Similarity interaction reflected larger absent-present difference in correct responses with high similarity (.99 vs. .89) than with low similarity (.99 vs. .97). The Target Presence x Set Size interaction was due to more errors with set size 10 than 5 in target-present trials (.94 vs. 92), whereas there were no set size differences in target-absent trials (.99 vs. .99)

Reaction times. Figure 5 shows mean correct RTs. The same ANOVA as for error rates, revealed all main effects and interactions to be significant for reaction times (*F*s>10.47, *p*s<.006), with the exception of the three-way interaction (F=1.2). To follow this up, we focused on search slopes of the RT-set size function (difference in RT divided by difference in set size). An ANOVA with the variables target presence and similarity revealed main effects for target presence, $F(1, 13) = 88.78$, $p < .001$, $\eta_G^2 = 0.87$, and similarity , $F(1, 13) = 34.56$, $p < .001$, $\eta_G^2 = 0.73$. The interaction was not significant, $F(1, 13) = 1.2$, $p = .293$, $\eta_G^2 = 0.08$. Search slopes were steeper in target-absent than target-present trials (278 ms/item vs. 152 ms/item) and steeper with similar than dissimilar targets (273 ms/item vs. 158 ms/item).

-- Figure 5 about here --

-- Figure 6 about here --

Dwell times. An ANOVA of the dwell times (Figure 6a) with the variables target presence (present vs. absent), similarity (similar vs. dissimilar), and stimulus type (distractor in absent vs. distractor in present vs. target in present trial) rendered significant main effects for similarity, $F(1, 13) = 19.07$, $p = .001$, $\eta_G^2 = 0.59$, and stimulus type, $F(2, 26) = 67.57$, $p < .001$, $\eta_G^2 = 0.84$. The main effect of set size was not significant, $F(1, 13) = 2.49$, $p = .138$, $\eta_G^2 = 0.16$. Of the interactions, only the Stimulus Type × Set Size interaction was significant, $F(2, 26) = 4.62$, $p = .019$, $\eta_G^2 = 0.26$, other *F*s < 1.55, *p*s > .235. The main effect for similarity was due to longer dwell times in blocks with similar than dissimilar targets (391 vs. 455 ms). The main effect for stimulus type was in particular due to much longer dwell times on targets than distractors (see Figure 6a); t-tests revealed, however that all stimulus types differed from each other, *t*s >7.44, *p*s<.001.

The interaction between set size and stimulus type was followed-up by three t-tests, each comparing similarity for each stimulus type. There was no significant effect of set size on dwelling for targets, $t(13) < 1$, or distractors in target-present trials, $t(13) < 1$. However, for distractors in target-absent trials, $t(13) = 7.47$, $p < .001$, dwell time was longer with set size 5 than 10 (309 vs. 270 ms).

Proportion of fixated and skipped stimuli. An ANOVA of the skipping proportions (Figure 6b) with the variables target presence (present vs. absent), similarity (similar vs. dissimilar), and stimulus (distractor in absent vs. distractor in present vs. target in present trial) rendered significant results for all main effects and interactions, *F*s>4.94, *p*s<.05.

Follow-up tests revealed that for distractors in target-absent trials there was a significant main effect for similarity, $F(1, 13) = 51.45$, $p < .001$, $\eta_G^2 = 0.8$, and for set size, $F(1, 13) = 84.4$, $p < .001$, $\eta_G^2 = 0.87$. The two-way interaction was significant as well, $F(1, 13) = 30.02$, $p < .001$, $\eta_G^2 = 0.7$. Skipping was increased for dissimilar than similar targets (.10 vs.02), and with set size 10 than 5 (.09 vs. .03). This set size effect was larger for dissimilar than similar targets (difference .10 vs. .02).

A corresponding ANOVA for the distractors in target-present trials revealed main effects for similarity, $F(1, 13) = 50.81$, $p < .001$, $\eta_G^2 = 0.8$, and set size, $F(1, 13) = 60.99$, $p < .001$ $\eta_G^2 = 0.82$. There was no reliable interaction, F<1. More distractors were skipped with dissimilar than similar targets (.52 vs .34), and when set size was 10 rather than 5 (.49 vs. .37).

Revisiting. An ANOVA, investigating the effects of set size on revisiting (Figure 6c), revealed that revisiting is higher for similar than dissimilar targets (.35 vs. .30), $F(1, 13) = 6.17$, $p = .027$, $\eta_G^2 = 0.32$, higher for targets (.51) than for distractors in target-present (.10) and target-absent (.35) trials, $F(2, 26) = 89.13$, $p < .001$, $\eta_G^2 = 0.87$, and higher for set size 5 than 10 (.34 vs. .30), $F(1, 13) = 18.27$, $p = .001$, $\eta_G^2 = 0.58$. Only the Stimulus Type x Set Size interaction reached significance, $F(2, 26) = 24.1$, $p < .001$, $\eta_G^2 = 0.65$. The revisiting rate of the targets was virtually the same ($t < 1$) for set size 5 and 10 (.51), but it was increased for distractors in target-present trials with set sizes 5 than

10 (.12 vs .09), *t*(13) = 3.99, $p < .002$, and for distractors in target-absent trials with set sizes 5 than 10 (.41 vs. .29), *t*(15) =8.10, $p < .001$.

Impact of similarity on dwelling, skipping, and revisiting, and contributions of dwelling, skipping, and revisiting on search times

Target-absent trials. Correlations are shown in Table 3. As already indicated in the ANOVAs, the correlations of similarity with dwelling and skipping were substantial as before, the correlations of similarity and revisiting was significant but low here. The regression analyses corresponded to the previous ones of Experiments 1 and 2, with the exception that set size was included as a factor. Also, interactions of set size with dwelling, skipping, and revisiting are now included to examine whether the effects of dwelling, skipping, and revisiting change across set sizes. Set size was dummy-coded as 0 for set size 5 and 1 for set size 10. There is no difference computationally between treating set size as a metric or categorical variable. The advantage of dummy coding is that the simple regression coefficients can be read as showing the results for set size 5, and the interactions coefficients give the increments for set size 10. As before, the metrical predictors were standardized before regression analysis. In the following, we will concentrate on the target-absent trials, as these are most informative.

Table 4 (top panel) shows the statistical values obtained in the regression analyses based on 1680 target-absent trials. Collinearity among the predictor variables was stronger than in the previous experiments but still acceptable, with all $1/VIF > .29$. All predictors but similarity showed significant effects. Marginal R^2 was .79. Note that regression slopes for dwelling, skipping and revisiting can be derived from Table 4 by adding regression coefficients (i.e., the regression slopes) for set size 5 and 10 of a respective predictor. For instance, the regression slope for dwelling with set size 10 would be .63, as regression coefficients are .32 for set size 5 and .31 for the increment with set size 10 (.32+.31=.63). The interactions of set size with dwelling skipping, and revisiting, respectively, were significant. This indicates that the regression slopes differed significantly between the set sizes. More precisely, all regression slopes were more extreme in the set size 10 than in the set size 5 condition.

Target-present trials. Table 4 (bottom panel) shows the results from the regression analyses based on 1620 target-present trials. Collinearity was small, with all tolerances ($1/VIF$) > .37. All predictors had significant effects. Marginal R^2 was .66.

Discussion

Experiment 3 yielded five informative results. First, not surprisingly, search was clearly inefficient, as gauged by the search slopes of the RT/set size function. Second, as predicted, search was less efficient for similar than dissimilar targets. Third, there was only a small effect of set size on dwell time, with slightly prolonged dwell times at the smaller set size. The main reason for this may be that dwelling not only includes the time to analyze a currently fixated stimulus, but also the time to select the next fixation location, and to plan and prepare the corresponding saccade. With set size 5, the spatial separation of the stimuli is necessarily higher in the present set up than with set size 10 and this might cause increased dwelling due to selection of a further away location (which may be complicated by degraded visual acuity) and preparation of spatially longer saccades. There is yet no consensus about which of the two variables is more affected by stimulus separations (Ludwig, Davies, & Eckstein, 2014; Unema, Pannasch, Hoos, & Velichkovsky, 2005; Antes, 1974, Viviani & Swensson, 1982). Alternatively, it might be that with small set sizes, participants feel less pressured to work quickly on each individual item, because with few items, working somewhat slower does not affect search time much. Fourth, there was more skipping with set size 10 than 5, and this was more pronounced during search for dissimilar compared to similar targets, i.e., skipping increases in more densely populated displays. The observation of more skipping with larger set sizes is in accordance with the assumption that more than one stimulus can be processed within one fixation (e.g., Hulleman & Olivers, 2015, Venini et al., 2014), but that the probability that an additional stimulus is in the functional field of view, is lower with sparsely populated displays (e.g., in a set size 5 display). Finally, revisiting rates were higher with set size 5 than 10. This is a rather unexpected result, because revisiting has been tied to memory overflow (e.g., Hulleman & Olivers, 2017) in that revisiting should occur more often for larger set sizes because participants forget the locations they already visited.

It is of note that the effects of dwelling, skipping, and revisiting in the regression analyses are remarkably stable, showing a similar pattern of effects as for Experiment 1 and 2: Dwelling and revisiting have both strong effects, and skipping a slightly weaker effect on search times in absent trials. All regression slopes were steeper with higher set sizes. This is not unexpected as dwelling, revisiting, and skipping are statistics per trial per stimulus. That is, an average dwell time of, for instance, 200 ms in a given trial means that each stimulus is looked at on average for 200 ms. This in turn implies that dwelling would have a stronger impact on RT with more stimuli in the display.

Experiment 4

Experiment 4 was designed to measure the effects of dwelling, skipping, and revisiting under conditions that most likely promote strong guidance (i.e., easy search). Given the inefficient searches employed in Experiment 1 to 3 (as indicated by the steep search slopes measured in Experiment 3, and the high search times in Experiment 1 and 2), it seems important to test whether dwelling, skipping, and revisiting still reliably contribute to search times when search might exclusively be controlled by guidance. We implemented guidance by allowing a subset search. Of the 10 stimuli presented, five distractor stimuli had a color (orange) that was never the target color. The target, if present, was always among the stimuli with the other color (blue). Note that the coloring (blue vs.

orange) uses a basic feature that is accessible before attentional deployment, and should thus be able to guide attention towards the subset of stimuli that possibly contains the target.

Methods

Participants. 16 students participated in the study. Each received €4-5 for their 30-40-minute service. Mean age (one age missing) was 25.21 years (SD = 3.24 years); 13 were female.

Apparatus, Stimuli, and Procedure. These were as in Experiment 1, with the exceptions that (1) the same 5x5 grid was used as in Experiment 3, (2) five of the distractors were shown in orange, while the other five stimuli (five distractors or four distractors and the target) were shown in blue, and (3) each block was twice as long and comprised 40 trials. Participants were informed about the target color (blue) at the beginning of the experiment and were therefore able to limit their search to five items in each trial.

Data pre-processing. Data were preprocessed as in Experiments 1 and 2. Outlier screening led to the exclusion of 103 RTs, 474 dwell times, and 62 latencies. As before, the screening was done case wise and separately for each combination of the variables similarity (similar vs. dissimilar), target presence (absent vs. present).

Results

Reaction times. An ANOVA of the RTs with the variables target presence (present vs. absent) and similarity (similar vs. dissimilar) revealed main effects for target presence, $F(1, 15) = 72.68$, $p < .001$, $\eta_G^2 = 0.83$, and similarity, $F(1, 15) = 20.9$, $p < .001$, $\eta_G^2 = 0.58$. The interaction was not significant, $F(1, 15) = 2.88$, $p = .111$, $\eta_G^2 = 0.16$. RTs were delayed in absent compared to present trials (1905 ms vs. 1414 ms), and they were slower with similar than with dissimilar targets (1456 ms vs. 1864 ms).

Error rates. The ANOVA on error rates revealed only a main effect of similarity, $F(1, 15) = 11.29$, $p = .004$, $\eta_G^2 = 0.43$ (other *F*s < 1). More errors were made with similar than with dissimilar targets (.15 vs. .06).

Contributions of distractor dwelling, skipping, and revisiting to search times

In contrast to the previous experiments, the search displays in Experiment 4 allow the participant to a priori exclude of half of all display items in each trial (all orange items). This subset search is assumed to provide a strong incentive for participants to use guidance, which should consequently be reflected in increased skipping rates. Figure 7 shows the means (panels a, b, and c) for dwelling, skipping, and revisiting for the factorial combinations of similar and dissimilar targets, and for distractors with or without the target color, for both target-absent and target-present trials.

Table 5 presents the bivariate correlations between RTs and the rates for revisiting and skipping together with the dwell time, separately for target-absent and target-present trials. The regression analyses corresponded in all details to those employed in the previous experiments.

-- Figure 7 about here --

-- Table 5 about here --

Target-absent trials. Table 6 (top panel) shows the results of the weights from the regression analyses based on 1731 target-absent trials. Collinearity among the predictor variables was acceptable, with all $1/VIF > .73$. All regression slopes were significant. The regression slopes of dwelling, skipping, and revisiting were of similar size. Marginal R^2 was .81.

-- Table 6 about here --

Target-present trials. The bottom panel of Table 6 shows the respective values based on 1589 target-present trials. Indications of collinearity were low, $1/VIF > .88$. All regression slopes were significant. Marginal R^2 was .57.

Discussion

Experiment 4 tested the replicability of the effects reported in the previous experiments under conditions when guidance is highly likely, due to the possibility to focus search on a subset of stimuli in each trial. As expected, search was relatively fast, because participants were able to ignore half of the distractors that did not share the color with the target (compare the RTs for Experiment 1 and Experiment 3 in Figure 2). Also, skipping had a strong impact on search time in target-absent trials. Importantly, however, the weights for dwelling and revisiting were high as well, indicating that the presence of strongly guiding features does not eliminate the effects of dwelling and revisiting on search time.

General Discussion

We conducted four experiments that were designed to investigate to which extent skipping, dwelling, and revisiting, determine behavioral search performance. Most generally, we found that apart from skipping, which can be interpreted as a measure of search guidance, dwelling and revisiting processes also contributed substantially to the additional time requirements in inefficient versus efficient search. More specifically, Experiments 1 and 2 showed that target-distractor similarity influenced skipping, dwelling, and revisiting, and that these variables in turn contributed to search performance in both experiments, and on both target-present and target-absent trials. Experiment 3 revealed that this contribution holds across sparely and densely populated search displays (and mirrored classical findings of search slope modulations by target-distractor similarity). Experiment 4 finally explored whether the effects of skipping, dwelling, and revisiting persist when search is more likely controlled by guidance (i.e., increased skipping rates).

We have pointed out that models of visual search often present elaborated theory on how guidance can be analyzed and how it affects search performance, and in particular search efficiency. Other aspects such as the contribution of attentional dwelling and revisiting are relatively underdeveloped, which implies that their role in search is generally assumed to be relatively minor in nature. The present results, however, do not seem to support this presumption. In the present experiments, the effects of similarity on dwelling and revisiting are significant, as well as the effects of dwelling and revisiting on search times. Our results thus point to an explanatory gap in the contemporary theories of visual search and indicate that a successful search model should include the search mechanism of dwelling and skipping to explain more or less efficient search.

Experiment 3 revealed that search for the distorted version of the stimuli used in this study is very inefficient, with search slopes in of 150 ms /item in target-present trials. Does this imply that the observed effects of dwelling, and revisiting on search times are limited to inefficient searches where guidance (as reflected in skipping rates) is less likely to begin with? Experiment 4 tested this hypothesis and revealed that in target-absent trials (RT without contamination of any target-related processing), skipping rates were increased in Experiment 4 as compared to Experiment, 1. This suggests that the sub-set search of Experiment 4 indeed allowed for a more guided search as compared to the full-set search employed in Experiment 1. Importantly, however, even under conditions that promote guidance, dwelling and revisiting were still substantially contributed to the search RTs in Experiment 4. On a greater scale, our results suggest that guidance by target features (selection) as measured by increased skipping rates on the one hand, and rejection of distractors as measured by decreased dwell times on the other hand, are two mechanisms that coexist in visual search.

We would like to begin a closer inspection of our search variables with skipping. Skipping was influenced by similarity and in turn influenced search times, in all experiments, and in both target-present and -absent trials. Skipping in target-absent trials can be explained as a result of a variable threshold for search termination. In the context of GS (2.0; Wolfe, 1994), Chun and Wolfe (1996) showed that search is terminated when activation falls below variable threshold. This threshold is set well above zero when target-distractor similarity is low (efficient search), however, with increased target-distractor similarity the threshold is set to a low value near zero. In consequence, search can be terminated early in the first case, but in the latter, it may happen that all stimuli of the display must be examined before search can be terminated. One might argue that skipping is not an exhaustive measure of ‘ignoring’ a search item. Clearly, a peripheral stimulus can covertly be attended while focal attention is fixated at a different location in the visual field. This has, for example, been put forward by the Functional View Field model (FVF; Hulleman & Olivers, 2016). The FVF is the region in space where a target can be detected among distractors with sufficient reliability in a single eye fixation. If the FVF includes more than one stimulus, this will result in the skipping of stimuli. The size of the FVF changes with search difficulty, and thus, depending on whether more or less stimuli are included in the FVF, skipping rates may differ for similar and dissimilar targets. However, correlations between eye and attention movements are undoubtedly very high (Deubel & Schneider, 1996), and we assume this to be sufficient to justify our methodological approach to employ fixations as an online marker of attentional selection (or de-selection, respectively).

In the context of skipping, we would like to state that despite the fact that we analyze and report search mechanisms in target-present trials, these results should be interpreted with caution. In particular when guidance is weak (inefficient search conditions), skipping rates will be dominated by random noise in the activation map. Finding the target as the first item (in which case the skipping rate is .90) or as the tenth item (skipping rate of zero) has a very large effect on RT, and the high regression weights for skipping are therefore not surprising. However, the regression analyses on target-present trials uncovered significant weights for dwelling and revisiting and therefore suggests a systematic contribution of search efficiency also on target-present and not just target-absent trials.

The second search time predictor, dwelling, was increased by target-distractor similarity, and in turn increased RT. It is important to note that we may have overestimated dwell times due to two factors. First, eye tracking procedures measure gaze fixations, which may be the sum of attentional dwell times on more than one stimulus included in one fixation (same logic as for skipping described above). Second, gaze dwelling does not only include the time of attentional dwelling, but in addition the time to select the next fixation target. While there is evidence that saccade planning is done in parallel to stimulus analysis (Ludwig, Davies, & Eckstein, 2014), some studies report that saccade amplitude register in fixation duration, indicating that the processes are either not completely parallel, or that saccade planning sometimes takes longer than stimulus analysis (Unema, Pannasch, Joos, & Velichkovsky, 2005). This important caveat, to our assessment, does not invalidate gaze dwell time as a valid indicator of the duration of stimulus analysis. For example, Becker et al. (2011) found that perceptual target difficulty (i.e., the acuity needed to detect the feature on which the response is based) increased dwell time independently from target-distractor similarity. While target-distractor similarity theoretically influences both the categorization of a stimulus (attentional dwelling) and the selection of the next candidate stimulus, perceptual target difficulty should influence categorization exclusively. Moreover, in the difficult searches of Experiments 1 to 3, dwelling had substantial effects, but here it is not plausible that this effect was dominated by the selection processes: Apparently guidance was rather weak, because otherwise search would not have been this inefficient.

Finally, revisiting was assessed for two reasons. First, revisiting is an empirical fact, as for example described in IOR studies, and should thus be included in the equation that explains differences in search times. Second, revisiting has been brought into focus recently by the FVF model (Hulleman & Olivers, 2016) which explicitly assumes that memory for recently checked positions is limited and that revisiting should thus be a general predictor of search behavior, in particular with larger set sizes. In line with these suggestions we find that revisiting is a substantial contributor to search performance. However, contrary to the suggestions in the FVF model, in Experiment 3, revisits were not more likely with the higher set size. Perhaps, set size in our experiments was still too small (10) to tease out such effects of IOR and VSTM limitations? Of all variables tested in our experiments, revisiting was least affected by target-distractor similarity (see correlation Tables 1,3, and 5). This fits the idea that skipping is related to mechanisms such as IOR-decay time or VSTM capacity, which are independent on the actual search task.

For our study, we have used a relatively abstract "guidance-based model" as a reference frame, with GS, TAM, DW, or SM, as possible exemplars. However, we appreciate that this simplification does not do justice to the individual models, which do differ from each other with many respects. In particular, GS and DW are primarily attentional models, while TAM and SM are rather models of eye-movements. Our results, which are based on eye movements, therefore relate more directly to TAM, and SM. Conversely, for GS and DW, some of our underlying assumptions – as specified above (for example the interpretation of fixations as markers of attentional selection or the possible overestimation of attentional dwell time by empirical gaze dwell time) - may be more difficult to accept. GS (2.0; Wolfe, 1994) and TAM, while similar in many respects, are different in that GS treats near and far stimuli the same, while TAM employs the concept of an inhomogeneous retina, where stimuli closer to fixation are represented with higher visual acuity than more peripheral stimuli. According to GS, the activation map is thus computed once for each display (and updated with IOR for a visited stimulus), while TAM assumes a new coding of the activation map with every fixation, because every fixation changes the spatial parameters between the stimuli of a display. Irrespective of these important details, though, all of these models make important predictions for visual search performance. They should therefore implement the search mechanisms of skipping, dwelling and revisiting, as with the present experiments, we have shown that these have a substantial impact on visual search performance.

On a broader scale, our results have important implications for the interpretation of behavioral search experiments that use RT to measure search performance. An idealized guidance-based model suggests that differences in search efficiency are solely caused by different guidance opportunities. More efficient search would be due to higher skipping rates, as efficient search allows for a priori rejections of distractors as possible target candidates. Correspondingly, less efficient search is due to low skipping rates, as multiple distractors have to be checked for target status. Because search efficiency is explained by guidance exclusively, it is tempting to interpret differences in search efficiency as directly indicating different degrees of guidance (given that other possible influences such as differential amounts of crowding can be excluded; cf. Vlaskamp & Hooge, 2006). However, our series of experiments suggest that skipping is not the only mechanism determining visual search performance: dwelling and revisiting also significantly modulate search efficiency. The effects on dwelling and skipping, however, cannot be disentangled in RT data. RT measures, and in particular the measures of search efficiency derived from search function slopes, are completely blind as to whether these effects are driven by longer dwelling or less skipping or more frequent rescanning. Eye-tracking data, as used in the present experiments, are very useful to assess dwelling and skipping separately. At present, eye tracking seems to be the only method that allows distinguishing between skipping, dwelling, and revisiting.

Taken together, search slopes are a reliable measure of search efficiency, but they are not informative as to the underlying search mechanisms. Eye tracking data are required to disentangle contributions of guidance/skipping, dwelling and revisiting on search performance. Any comprehensive model of visual search must incorporate all three of those selection mechanisms: skipping, dwelling, and revisiting.

References

Alexander, R. G., & Zelinsky, G. J. (2012). Effects of part-based similarity on visual search: The Frankenbear experiment. *Vision research, 54*, 20-30.

Becker, S. I. (2011). Determinants of dwell time in visual search: Similarity or perceptual difficulty? *PLoS ONE* 6(3): e17740. doi 10.1371/journal.pone.0017740

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

Chun, M. M., & Wolfe, J. M. (1996). Just say no: How are visual searches terminated when there is no target-present? *Cognitive Psychology, 30*, 39-78.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96*, 443-458.

Found, A., & Müller, H. J. (1996). Searching for unknown feature targets on more than one dimension: Investigating a “dimension-weighting” account. *Perception & Psychophysics*, 58(1), 88-101.

Godwin, H. J., Reichle, E. D., & Menneer, T. (2017). Modeling Lag-2 revisits to understand trade-offs in mixed control of fixation termination during visual search. *Cognitive Science, 41(4)*, 996-1019.

Gould, J. D. (1967). Pattern-recognition and eye-movement parameters. *Perception & Psychophysics, 2*, 399-407.

Hoffman, J. E. (1978). Search through a sequentially presented visual display. *Perception & Psychophysics, 23,* 1-11.

Hooge, I. T. C., & Erkelens, C. J. (1998). Adjustment of fixation duration in visual search. *Vision Research, 38,* 1295–1302.

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

Horstmann, G., Herwig, A., & Becker, S. I. (2016). Distractor dwelling, skipping, and revisiting determine target-absent performance in difficult visual search. *Frontiers in Psychology,7*, 1152.

Horstmann, G., Lipp, O. V., & Becker, S. I. (2012). Of toothy grins and angry snarls - Open mouth displays contribute to efficiency gains in search for emotional faces. *Journal of Vision. 12* (5), 7.

Hout, M. C., *Robbins, A., Godwin, H. J., Fitzsimmons, G, & *Scarince, C. (2017). Categorical templates are more useful when features are consistent: Evidence from eye-movements during search for societally important vehicles. *Attention, Perception, & Psychophysics, 79*, 1578-1592

Hulleman, J. & Olivers, C. N. L. (2016). The impending demise of the item in visual search. *Behavioral and Brain Sciences*, in press.

Humphreys, G. W., & Müller, H. J. (1993). Search via Recursive Rejection (SERR): A connectionist model of visual search. *Cognitive Psychology, 25*, 43-110.

Jenkins, M., Grubert, A., & Eimer, M. (2018). Category-based attentional guidance can operate in parallel for multiple target objects. *Biological Psychology, 135*, 211-219.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology, 4,* 219–227.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4*(2), 133-142.

Nako, R., Wu, R., Smith, T. J., & Eimer, M. (2014). Item and Category-Based Attentional Control During Search for Real-World Objects: Can You Find the Pants Among the Pans? *Journal of Experimental Psychology: Human Perception and Performance, 40*, 1283-1288.

Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research, 46*(5), 614-621.

Neisser, U. (1964). Visual search. *Scientific American, 201*, 94-102.

Olivers, C. N. L., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: When it guides attention and when it does not. *Trends in Cognitive Sciences, 15,* 327–334.

Pashler, H. (1994). Overlapping mental operations in serial performance with preview. *Quarterly Journal of Experimental Psychology Section A. Human Experimental Psychology*, *47*, 161-191.

Posner, M. I., Rafal, R. D., Choate, L. S., & Vaughan, J. (1985). Inhibition of return: Neural basis and function. *Cognitive Neuropsychology*, 2(3), 211-228.

Remington, R. W., Lewis, R., & Wu, S.-C. (2006). Scheduling mental operations in a multiple-response sequence: Modeling the effects of a strategy to minimize variance in the timing of saccades. In D. Fum, F. Del Missier, & A. Stocco (Eds.), *Proceedings of the 7th International Conference on Cognitive Modeling* (pp. 250-255) Trieste, Italy: Edizioni Goliardiche .

Rosch, E. H. (1973). Natural categories. *Cognitive psychology*, *4*(3), 328-350.

Shipp, S. (2004). The brain circuitry of attention. *Trends in Cognitive Sciences, 8*(5), 223-230.

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, *81*(3), 214.

Tottenham, N., Tanaka, J., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., et al. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research, 168*, 242-249.

Treisman, A. (1985). Preattentive processing in vision. *Computer vision, graphics, and image processing*, *31*(2), 156-177.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97-136.

Treisman, A., & Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: General, 114*(3), 285-310.

Verghese, P. (2001). Visual search and attention: A signal detection theory approach. *Neuron, 31*, 523-535.

Venini, D., Remington, R.W., Horstmann, G., & Becker, S.I. (2014). Centre-of-gravity fixations in visual search: When looking at nothing helps to find something. *Journal of Ophthalmology, 237812*, 1-14.

Walenchok, S. C., Hout, M. C., & Goldinger, S. D. (2016). Implicit object naming in visual search: Evidence from phonological competition. *Attention, Perception, & Psychophysics*, *78*, 2633–2654

Wolfe J. M. (2007). Guided search 4.0: Current progress with a model of visual search. In W. Gray (Ed.), *Integrated models of cognitive systems (pp. 99–119)*. New York: Oxford.

Wolfe, J. M. (1994). Guided Search 2.0: A revised model of guided search. *Psychonomic Bulletin & Review, 1*, 202-238.

Wolfe, J.M. (1994b). Visual Search in Continuous, Naturalistic Stimuli. *Vision Research, 34*(9), 1187-1195.

Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Science*, *9*, 33-39.

Wolfe, J. M. (2001). Asymmetries in visual search: An Introduction. *Perception and Psychophysics, 63*(3), 381-389.

Wolfe, J. M. (2013). When is it time to move to the next raspberry bush? Foraging rules in human visual search. *Journal of Vision*, *13*(3), 10-10.

Wolfe, J. M. (2018). Visual Search. In J. Wixted) (Ed.), Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience (Vol. II. Sensation, Perception & Attention: John Serences (UCSD)): Wiley.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance, 15(3)*, 419.

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience, 5*, 495–501.

Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review, 115*, 787-835.

Zelinsky, G. J., Adeli, H., Peng, C., & Samaras, D. (2013). Modeling eye movements in a categorical search task. *Philosophical Transactions of the Royal Society B, 368(1628)*, 1–12.

Zelinsky, G. J., & Schmidt, J. (2009). An effect of referential scene constraint on search implies scene segmentation. *Visual Cognition, 17*(6), 1004-1028.

Funding

The work was supported by the Cluster of Excellence – Cognitive Interaction Technology ‘CITEC’ (EXC 277) at Bielefeld University, which is funded by the German Research Foundation (DFG), and by DFG grant HO 3248/2-1 to Gernot Horstmann.

Figure Captions

Figure 1. A depicts the dissimilar targets (top row), the similar targets (middle row), and the distractors (bottom row), used in Experiment 1. B shows two example search displays with a similar (left panel) or dissimilar target (right panel), respectively. The right part includes a simplified scan path in yellow. The "lines" are saccades and the "knots" are fixations, where the eyes drift slowly on a small spot. The figure also illustrates the main variables. The time the gaze foveates a stimulus is the dwell time; this may include one fixation, as on the first stimulus visited, or two fixations, as for example on the second stimulus visited. The stimuli on the right side of the screen are not visited during this trial; that is the skipping proportion on this trial is 0.5. The target is visited as the fourth stimulus, but it is revisited after fixating the fifth stimulus. As this is the only stimulus revisited on this trial, revising proportion for the target is 1.0 and revisiting proportion of the distractors is 0.0

Figure 2. Mean correct RTs on target-absent and target-present trials for Experiments 1, 2 and 4, separately for blocks with dissimilar and similar targets. Results for Experiment 1 are on the left, for Experiment 2 in the middle, and for Experiment 3 on the right. Error bars are standard errors (i.e., SD / √N) of the means.

Figure 3. Mean dwell times (a) , proportion of skipped stimuli (b), and proportion of revisited stimuli (c) for non-target distractors in target-absent trials, for non-target distractors in target-present trials, and for targets in target-present trials, separately for blocks with dissimilar and similar targets, respectively. Results for Experiment 1 are in the top row, and for Experiment 2 in the bottom row. Error bars are standard errors (i.e., SD / √N) of the means.

Figure 4. Bivariate relationship between trial search times (RTs), revisiting rates (Revisiting), skipping rates (Skipping), and dwell times (Dwelling), for distractors in target-absent trials from Experiment 1, separately for the dissimilar target condition (left side) and the similar target condition (right side). Each dot represents a single trial.

Figure 5. Mean correct RTs for Experiment 3. Error bars are standard errors (i.e., SD / √N) of the means.

Figure 6. Mean dwell times (a), proportion of skipped stimuli (b), and proportion of revisited stimuli (c) for Experiment 3. Error bars are standard errors (i.e., SD / √N) of the means.

Figure 7. Averages for Experiment 4. Mean dwell times (a), proportion of skipped stimuli (b), and proportion of revisited stimuli (c) for distractors, separately for similar and dissimilar targets, and for distractors with or without the basic feature characterizing the target. Error bars are standard errors (i.e., SD / √N) of the means.

Table 1. Correlation matrix for the variables target-distractor similarity, RT, skipping, dwelling, and revisiting in target-absent and target-present trials of Experiment 1 and 2.

		Similarity	RT	Skipping	Revisiting
Experiment 1					
Target-absent trials	RT	.43			
	Skipping	-.40	-.66		
	Revisiting	.21	.49	-.27	
	Dwelling	.43	.80	-.46	.07
Target-present trials	RT	.34			
	Skipping	-.23	-.81		
	Revisiting	.25	.51	-.48	
	Dwelling	.21	.23	.09	-.10
Experiment 2					
Target-absent trials	RT	.34			
	Skipping	-.35	-.74		
	Revisiting	.13	.69	-.36	
	Dwelling	.28	.55	-.19	.12
Target-present trials	RT	.26			
	Skipping	-.22	-.86		
	Revisiting	.24	.54	-.45	
	Dwelling	-.12	-.11	.34	-.20

Note. Correlations were calculated on trial measures. All coefficients are significantly different from zero ($p < .05$).

Table 2. Linear multilevel regression of target-absent and target-present trial reaction times on dwelling, skipping, revisiting, and similarity as fixed effects, and random intercepts for participants based on the data of Experiment 1 and 2.

Experiment 1

Target-absent trials	*b*	*SE(b)*	*t*
Intercept	0.10	0.04	2.65
Dwelling	0.74	0.02	39.91
Skipping	-0.24	0.02	-14.99
Revisiting	0.44	0.01	30.68
Similarity	-0.10	0.03	3.59
Target-present trials			
Intercept	-0.04	0.05	-0.69
Dwelling	0.25	0.02	11.92
Skipping	-0.73	0.02	-37.71
Revisiting	0.22	0.02	11.15
Similarity	0.12	0.04	3.34

Experiment 2

Target-absent trials			
Intercept	0.06	0.02	2.87
Dwelling	0.43	0.01	37.16
Skipping	-0.49	0.01	-38.81
Revisiting	0.52	0.01	43.54
Similarity	-0.01	0.02	-0.49
Target-present trials			
Intercept	0.04	0.05	0.77
Dwelling	0.20	0.02	11.61
Skipping	-0.84	0.02	44.75
Revisiting	0.26	0.02	13.75
Similarity	0.03	0.03	1.04

Notes. *b* = regression coefficient; *SE* = standard error of regression coefficient; models allowed for random intercepts between subjects; estimation method was full maximum likelihood; with the exception of similarity, all metrical variables were *z*-transformed prior to analyses; for similarity “dissimilar target” was coded as zero and “similar target” as one; coefficients are statistically significant where $t > |1.96|$.

Table 3. Correlation matrix for the variables target-distractor similarity, set size, RT, skipping, dwelling, and revisiting in target-absent and target-present trials of Experiment 3.

		Similarity	Set size	RT	Skipping	Revisiting
Target-absent trials	Set size	.00				
	RT	.35	.51			
	Skipping	-.40	.30	-.21		
	Revisiting	.12	-.27	.32	-.31	
	Dwelling	.48	-.24	.36	-.24	.11
Target-present trials	Set size	-.01				
	RT	.28	.31			
	Skipping	-.29	.25	-.53		
	Revisiting	.07	-.10	.43	-.43	
	Dwelling	.24	-.01	.29	-.14	.10

Note. Correlations were calculated on trial measures. Coefficients < |.07| are not statistically significant ($p < .05$).

Table 4: **Linear multilevel regression of target-absent and target-present trial reaction times on dwelling, skipping, revisiting, similarity, and set size as fixed effects, and random intercepts for participants for Experiment 3**

Target-absent trials		b	SE(b)	t
	Intercept	-0.71	0.05	-13.01
	Dwelling	0.32	0.01	21.71
	Skipping	-0.07	0.02	-3.76
	Revisiting	0.29	0.01	21.29
	Similarity	0.04	0.03	1.47
	Setsize	1.56	0.02	68.13
	Dwelling:Setsize	0.31	0.02	13.94
	Skipping:Setsize	-0.11	0.02	-4.80
	Revisiting:Setsize	0.29	0.02	12.82
Target-present trials				
	Intercept	-0.48	0.07	-7.11
	Dwelling	0.07	0.02	3.95
	Skipping	-0.26	0.02	-11.66
	Revisiting	0.14	0.02	8.54
	Similarity	0.18	0.03	6.39
	Setsize	0.87	0.03	32.32
	Dwelling:Setsize	0.17	0.03	6.49
	Skipping:Setsize	-0.39	0.03	-12.21
	Revisiting:Setsize	0.32	0.03	9.79

Notes. *b* = regression coefficient; *SE* = standard error of regression coefficient; models allowed for random intercepts between subjects; estimation method was full maximum likelihood; with the exceptions of similarity and set size, all metrical variables were *z*-transformed prior to analyses; for similarity “dissimilar target” was coded as zero and “similar target” as one; for set size, set size 5 was coded as zero and set size 10 was coded as one; coefficients are statistically significant where $t > |1.96|$.

Table 5. Correlation matrix for the variables target-distractor similarity, RT, skipping, dwelling, and revisiting in target-absent and target-present trials of Experiment 4.

		Similarity	RT	Skipping	Revisiting
Target-absent trials	RT	.35			
	Skipping	-.22	-.57		
	Revisiting	.09	.56	-.23	
	Dwelling	.36	.63	-.05	.10
Target-present trials	RT	.26			
	Skipping	-.16	-.66		
	Revisiting	.07	.45	-.32	
	Dwelling	.25	.37	-.06	.06

Note. Correlations were calculated on trial measures. Coefficients ≥ |.05| are statistically significant (p < .05

Table 6. Linear multilevel regression of target-absent and target-present trial reaction times on dwelling, skipping, revisiting, and similarity as fixed effects, and random intercepts for participants based on the data of Experiment 4.

Experiment 1

Target-absent trials	*b*	*SE(b)*	*t*
Intercept	0.05	0.05	>1
Dwelling	0.51	0.01	44.96
Skipping	-0.40	0.01	-38.75
Revisiting	0.40	0.01	39.99
Similarity	-0.09	0.02	4.52
Target-present trials			
Intercept	-0.08	0.09	>1
Dwelling	0.28	0.01	18.93
Skipping	-0.54	0.02	-35.54
Revisiting	0.25	0.01	17.32
Similarity	0.20	0.03	7.44

Notes. *b* = regression coefficient; *SE* = standard error of regression coefficient; models allowed for random intercepts between subjects; estimation method was full maximum likelihood; with the exception of similarity, all metrical variables were *z*-transformed prior to analyses; for similarity "dissimilar target" was coded as zero and "similar target" as one; coefficients are statistically significant where *t* > |1.96|.

Figure 1

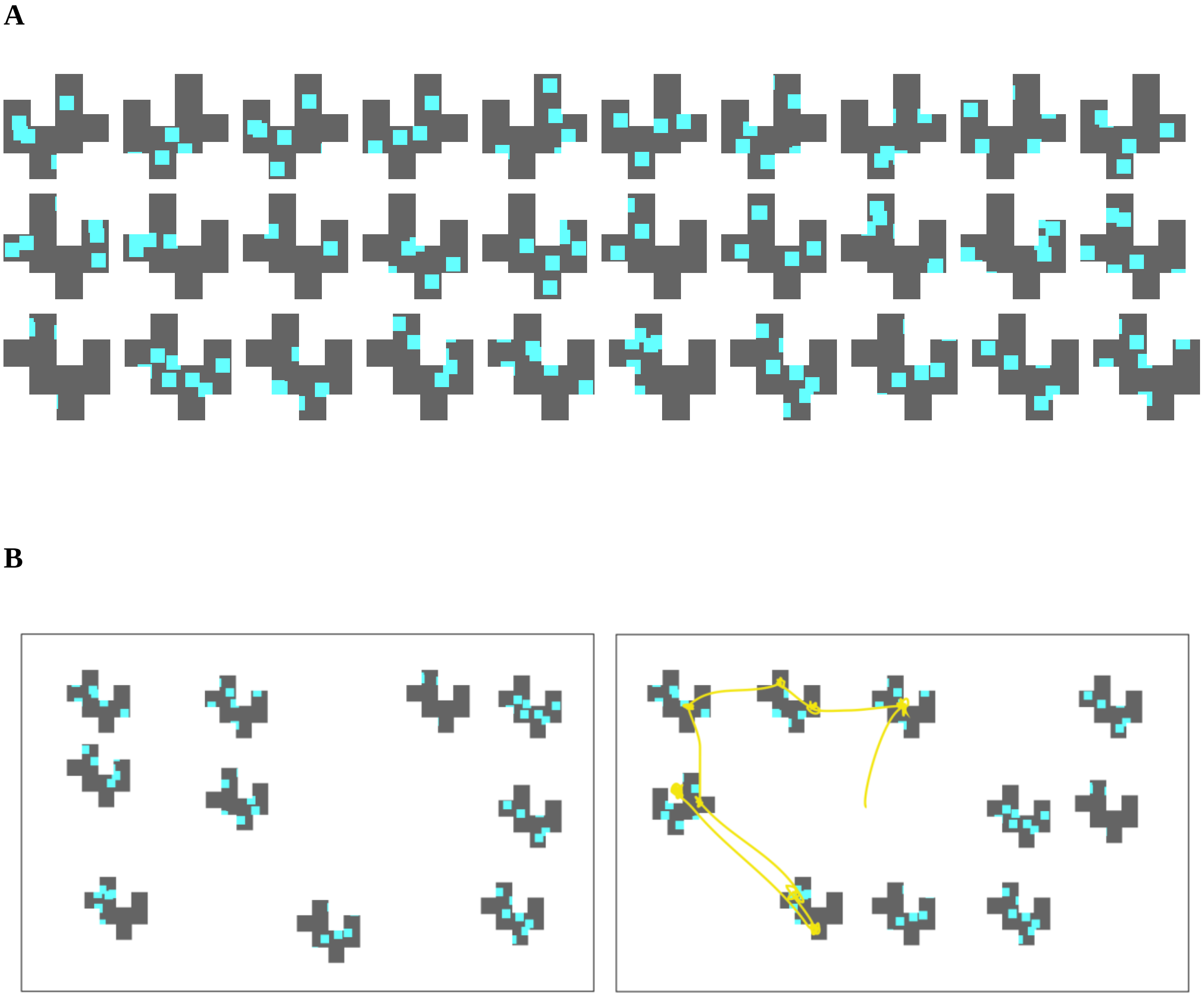


Figure 2

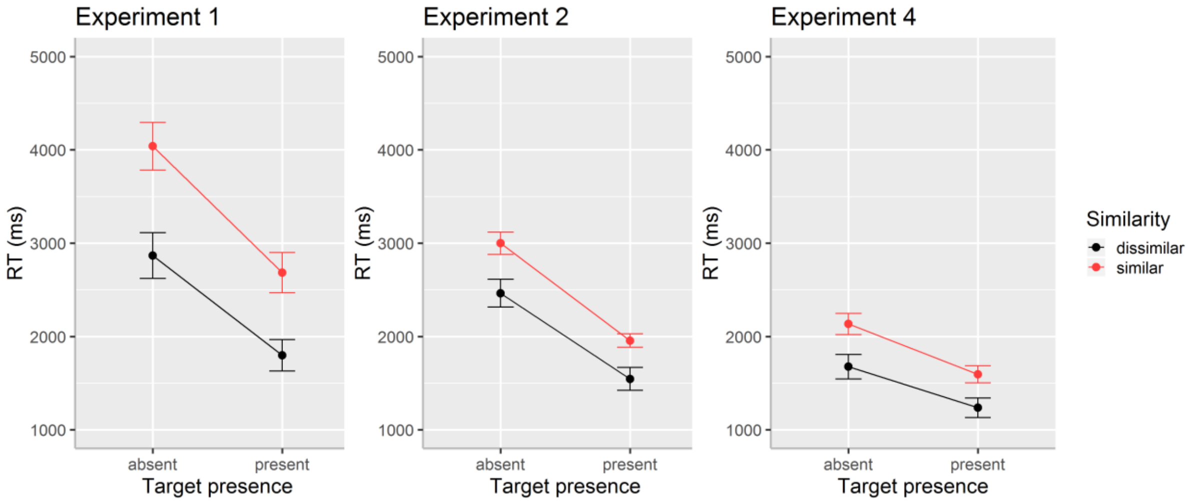
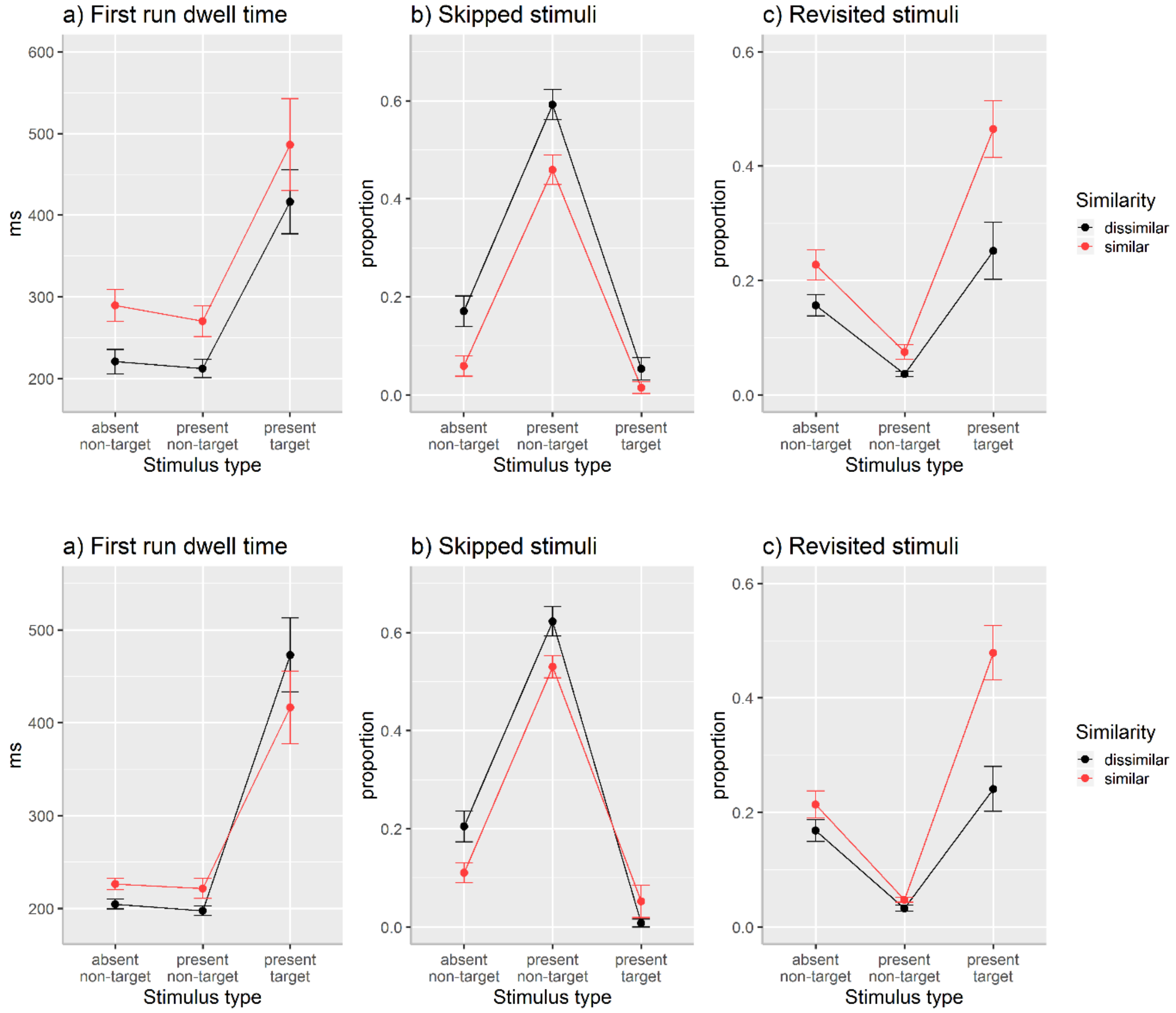

Experiment 1
Experiment 2
Experiment 4
RT (ms)
5000
4000
3000
2000
1000
absent
present
Target presence
Similarity
dissimilar
similar

Figure 3



a) First run dwell time

b) Skipped stimuli

c) Revisited stimuli

a) First run dwell time

b) Skipped stimuli

c) Revisited stimuli

Figure 4

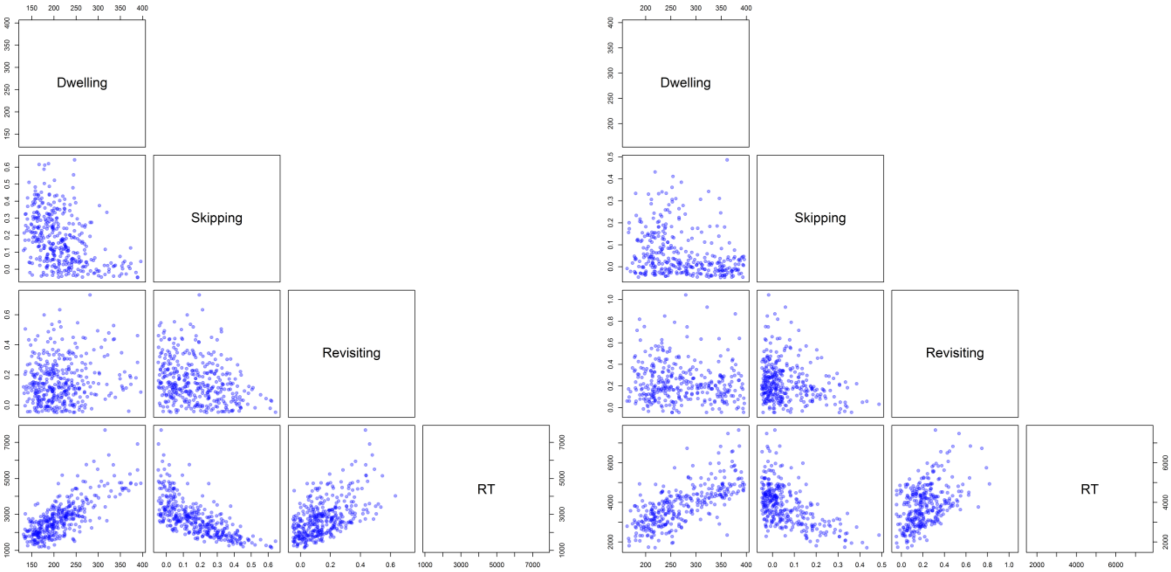

Dwelling
Skipping
Revisiting
RT
Dwelling
Skipping
Revisiting
RT

Figure 5 (Experiment 3)

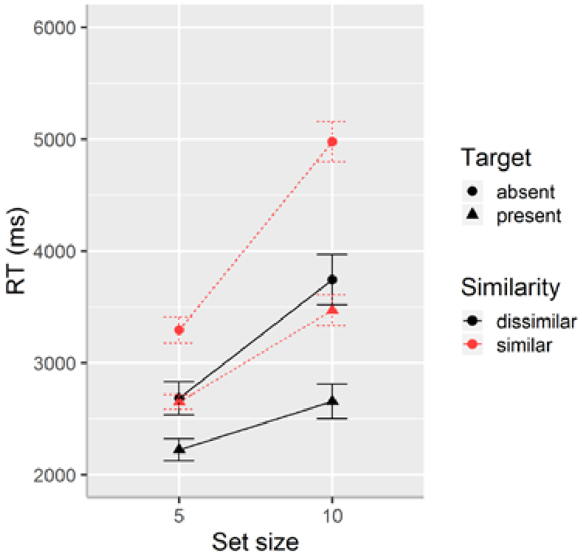


Figure 6 (Experiment 3)

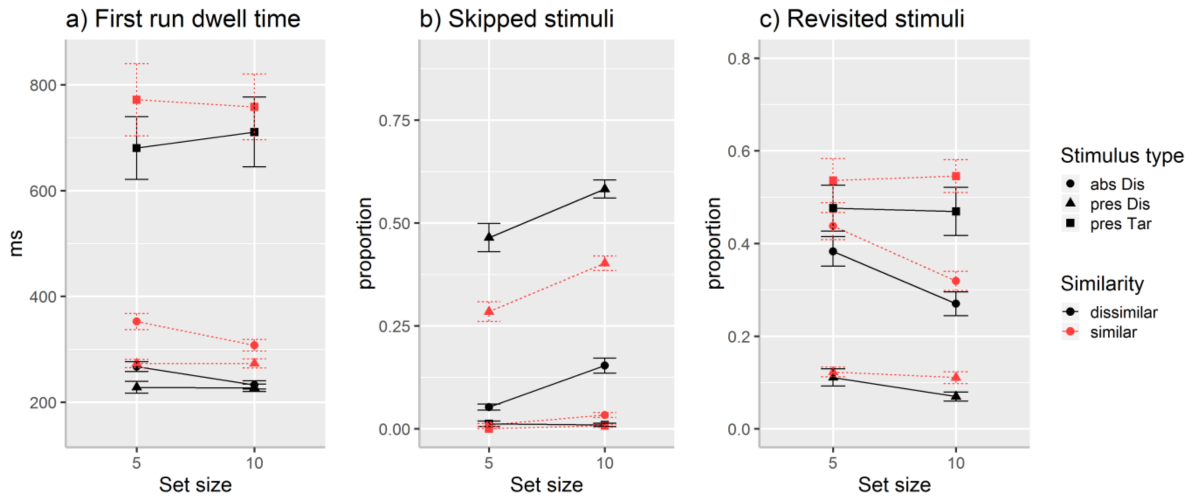


Figure 7 (Experiment 4)