

Published in Journal of Experimental Psychology: human perception and performance, Vol 37(4), Aug 2011, 997-1006. doi: [10.1037/a0023099](https://doi.org/10.1037/a0023099)

A clash of bottom-up and top-down processes in visual search: the reversed letter effect revisited

Li Zhaoping & Uta Frith, University College London

Abstract:




It is harder to find the letter 'N' among its mirror reversals than the mirror reversal among 'N's (Frith, 1974). This asymmetry is problematic for a bottom-up saliency hypothesis based on V1 mechanisms (Li 2002), since the uniquely tilted oblique bar in the target should be equally salient in both searches. Experiment 1 used dense search arrays to reduce target shape recognition before gaze reached target. Observers' gaze typically located the target in about half a second, equally fast in both searches. However, subsequently, gaze sometimes abandoned the target to search elsewhere before returning, more often so for target 'N', causing long delays (longer for target 'N') before observers reported the target. We suggest that this delay was due to a clash between bottom-up saliency (leading gaze to the uniquely tilted bar in the target) and top-down shape recognition (confusing the target and distractors as they have the identical zigzag shape). Experiment 2 shows that this clash was enhanced in sparser (smaller set size) arrays, in which top-down target shape recognition can occur earlier and bottom-up saliency is weaker, manifesting search asymmetry even before gaze reached the target. Search time increased and decreased with set size, respectively, when the array was sparse and dense, since bottom-up saliency increased with density. Our results indicate that the asymmetry does not invalidate the V1 saliency hypothesis, but rule out previous explanations of the asymmetry in terms of stronger pre-attentive salience for the reversed target and faster rejection of distractors in familiar orientation.

Keywords: visual search, saliency, familiarity, object recognition.

What makes some targets highly salient in visual search tasks? An easy search, such as finding a red item among green ones, suggests a fast bottom-up process. Here, a pre-attentively salient feature, i.e., a unique red color among green colours, readily attracts attention. Accordingly, red is termed a pre-attentive feature, which by definition attracts attention when unique in a scene of other items lacking this feature (Treisman & Gelade 1980, Wolfe 1998). In this paper, the terms pre-attentive and bottom-up are identical in meaning, and the term saliency always means attentional attraction by bottom-up processes. Visual search studies have led to psychological models of how saliency depends on visual inputs (Koch & Ullman 1985, Duncan & Humphreys 1989, Wolfe et al 1989, Mueller et al 1995, Itti & Koch 2000). They have also led to a computational theory proposing that the primary visual cortex (V1) computes saliency, such that the receptive field of the most activated V1 neuron is the most salient location to attract attention (Li 1999, 2002). Although this V1 saliency hypothesis is yet to be fully tested, it has provided several testable predictions on visual search and segmentation, which have been confirmed experimentally (Zhaoping & Snowden 2006, Zhaoping & May 2007, Koene & Zhaoping 2007, Jingling & Zhaoping 2008, Zhaoping 2008). The V1 hypothesis can even account for the examples of visual search asymmetry observed by Treisman and Gormiclan (1988), such as searching for an ellipse among circles which is easier than finding a circle among ellipses, or searching for a curved line among straight lines which is easier than the reverse. At the same time, Saiki et al (2005) and Saiki (2008) provided evidence that top-down task-related knowledge is unnecessary for some examples of search asymmetry.

However, the reversed letter effect presents an inconvenient finding: 'N' among its mirror reversals is

harder to find than the mirror reversal among 'N's (Frith 1974). This effect depends on the subject's familiarity with the letter shape. In fact, familiarity based search asymmetry, with an easier search for the less familiar target, is general (Wolfe 2001). To account for the asymmetry, it has been proposed that a target that deviates from a familiar stimulus, as in the case of the reversed letter 'N', is a pre-attentive feature that makes the target more salient (Treisman & Gormican 1988). Another proposal is that familiar distractors, in this case, the familiar letter 'N', are easier to reject (Treisman & Southern 1985, Wang et al 1994). Both proposals predict that attention or gaze should arrive at the less familiar target faster. In contrast, the V1 hypothesis predicts that, in a mainly bottom-up process, attention be attracted to the target equally fast whether or not the letter is reversed. This is because V1 neurons are tuned to primitive image features (e.g., oriented bars) regardless of shape familiarity, and there is no V1 mechanism for different saliencies for objects from familiar and unfamiliar viewpoints. Asymmetries in visual search tasks have also been found in Chinese readers. In addition to showing the reversed (Chinese) word effect, Chinese readers were slower searching among Chinese characters and their mirror reversals than non-Chinese readers (Shen & Reingold 2001). The absence of a reversed letter effect has been noted in Slavic readers. This is due to the fact that N and reversed N shapes are part of the Cyrillic alphabet and therefore equally familiar for Slavic readers (Malinowski & Hübner 2001).

We asked to what extent bottom-up saliency is implicated in the reversed letter effect and to what extent top-down processes interfere with a saliency based search. We were guided by a previous study that involved search for a reversible shape (Zhaoping & Guyader 2007). Here a target  had to be found among distractors  and . As in the case of finding an 'N' among 'И's or vice versa, the target can be distinguished by a low level feature, the uniquely tilted oblique bar which is salient, but it is also a reflected or rotated image of the distractors. Observers, after initially locating the target by gaze, often abandoned the target to search elsewhere. The authors described this as a veto of the correct saliency based decision, and suggested that it was due to perceiving the target and distractors as having identical shapes. Of course, they only have identical shapes when viewpoint is temporarily ignored, and this happens at the point when targets come into the focus of attention (Stankiewicz, Hummel, Cooper, 1998). Indeed, the veto and subsequent delay in the decision was eliminated by slightly altering the target shape. The target shape was altered by tilting the oblique bar only 20 degrees from its non-oblique partner, such that the target was no longer a reflected or rotated image of the distractors. This meant that the confusion was removed and they were not then perceived as identical shapes. We hypothesized that the letter N search task would also involve such processing stages and that we would reveal interference from a temporary confusion of overall shape identity despite perfectly clear pick-up of the distinctive oblique feature via bottom-up saliency.

We present here two experiments where we track eye gaze and vary the influence of bottom-up and top-down processes respectively. In our first experiment we aimed to strengthen bottom-up and weaken top-down processes by using dense and large search arrays. Dense arrays result in crowding of the stimuli and this leads to impaired shape recognition at the periphery (Levi 2008). Dense presentation therefore serves to reduce shape recognition before the gaze first arrives at the target during search. At the same time crowding does not prevent visual feature detection by bottom-up saliency (Levi 2008). On the contrary, it facilitates the detection of the uniquely oblique bar in the target, since the saliency of an orientation singleton increases with texture density (Nothdurft 2000). As saliency effects are seen mainly in the fastest eye movements (van Zoest and Donk 2006), dense arrays (which should also be sufficiently large to allow eye movements) enable us to observe the saliency process in relatively pure form in the reaction time for the first landing of the gaze at target. As the target-distractor shape confusion is liable to occur in both searches for target 'N' in its reversals and vice versa, we can probe whether the reversed letter effect in fact resides in the top-down interference of the search task due to this shape confusion.

In our second experiment we used multiple set sizes including both dense and sparse search arrays so that bottom-up processes were made stronger and weaker respectively. This was to enable us to investigate interactions between top-down and bottom-up processes in a spectrum of mutual balance. In particular we wanted to find out to what extent top-down interference can penetrate into the time window before the

first gaze arrival to target. This penetration could occur when target shape is recognized by peripheral vision in sparse arrays, just like one can recognize words next to the current fixation during reading (Legge, Mansfield & Chung 2001). This experiment will additionally enable us to relate our findings to previous findings of the reversed letter effect as manifested in set size effects (Wang, Cavanagh, & Green 1994, Malinowski & Hübner 2001, Wolfe 2001).

Experiment 1

We tracked gaze during search in a dense and large array, 12 rows by 16 columns, using the letters N and Z and their mirror images, see Figure 1. In each search, the target is a rotated or reflected image of the distractors, but can be distinguished by a uniquely left or right tilted oblique bar tilted in the opposite direction from the uniformly tilted oblique bars in the distractors, so that neither object shape or letter recognition is necessary for the search. Let RT_{report} be the reaction time (RT) for subjects to report the search outcome. The reversed letter effect means that RT_{report} is longer for a target N among distractors ‘I’ than the reverse. If this effect is mainly caused by a pre-attentive feature in the target’s deviation from familiarity, or by easier identification and rejection of familiar distractors in search, then it should be manifest in the reaction time RT_{gaze} of the first gaze landing at target. If the effect however is mainly caused by top-down processes, then it should mainly be manifest in $RT_{lapse} = RT_{report} - RT_{gaze}$, the time interval between the first gaze landing and the final report. We also attempted to assess the role of target and distractor familiarity separately. To this end we compared the RTs between two searches which differed only in target familiarity (and have the same distractors), e.g., in Figure 1A and 1C for target ‘N’ and ‘Σ’ respectively in distractors ‘I’, or between two searches which differed only in distractor familiarity (and have the same targets), as in Figure 1B and 1D.

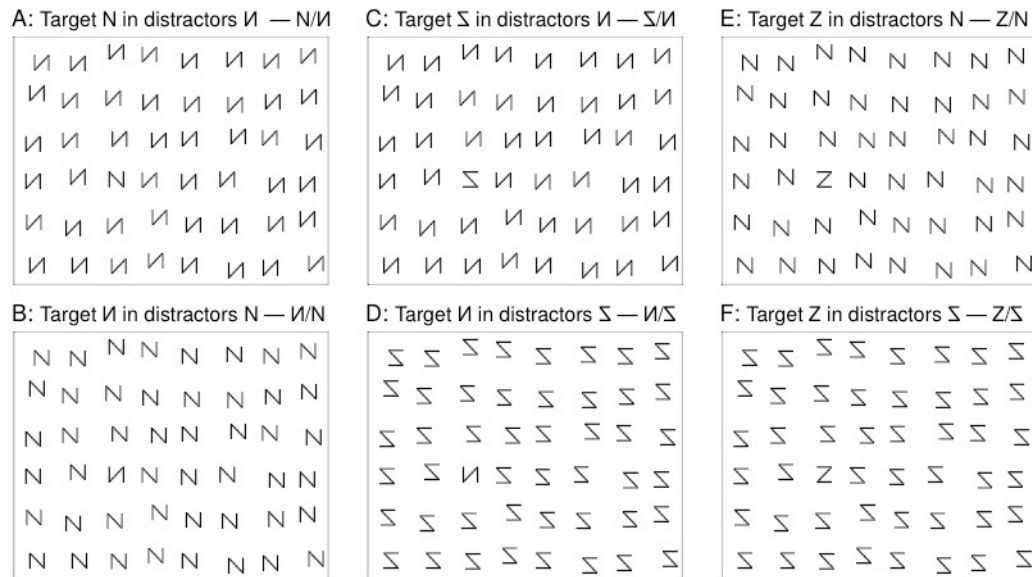


Figure 1. A-F: Illustrations of search stimuli in experiment 1. The actual stimuli contained 12 rows by 16 columns of items, as detailed in the text. A and B make the pair for the N vs. I asymmetry. A and C differ in targets but have the same distractor, similarly for the pair B and E, or D and F --- each to examine the role of target familiarity. Analogously, pair B and D and pair E and F enable one to examine the role of distractor familiarity.

Methods

Observers:

All observers (subjects) were literate in English, between 18-45 years old, had normal or corrected-to-normal vision, and were naïve to the research goal.

Stimuli:

Each search display was a 600x800 pixel image, viewed at a distance of 40 centimeters, and spanned 35° x 46° in visual angle. Each search item was at a position randomly displaced, up to 8 pixels, horizontally and vertically, from its grid location in a regular 12x16 grid with unit distance 50 pixels. Each item occupied 25x25 pixels, looked like N, Z, or their mirror images. It had three black bars of 2 pixels wide, one oblique tilted 45° from vertical, and the other two were either horizontal or vertical, on a white background. All items were rotated or mirror reflected images of each other. The target's grid location was randomly one of sixteen that were 215-276 pixels (about 12°-16°) radially, and at least 125 pixels (about 7.5°) horizontally (left or right), from the display center. A black disk of 0.3° diameter at the display center on a white background served as the fixation stimulus.

Procedure:

Without mentioning the words “letter”, ‘N’, or ‘Z’, we instructed the observers to find a target item containing a uniquely tilted oblique bar that was left or right tilted in the opposite direction from the uniformly oriented oblique bars in the distractors. They were told that the non-oblique bars were irrelevant, and that they should press a left or right button quickly to report whether the target, present in each trial, was in the left or right half of the display. In addition, to minimize other top-down influences such as search strategy, we informed the subjects not to search by looking around systematically (such as in reading text).

We denote each target-distractor condition by target/distractor, e.g., N/I means target N among distractors I, while Z/N means target Z among distractors N, see Figure 1. Each subject participated in one data collection session, involving randomly interleaved trials of N/I, I/N, and some other target-distractor conditions (see below). All target-distractor conditions employed about 30 trials each. Among the 11 subjects whose data will be presented later, four (SL, MJ, MM, DR) also searched Σ /I (Fig. 1C) and I/ Σ (Fig. 1D), and three (DL, HD, and AM) also searched I/ Σ (Fig. 1D), Z/N (Fig. 1E), and Z/ Σ (Fig. 1F). Each subject only practiced 1-2 trials for each search condition immediately before data collection. To avoid the build-up of a strategy, we restricted the total number of trials per subject, which meant that we did not interleave all six conditions shown in Fig. 1 for each subject. As Zhaoping and Guyader (2007) observed, subjects learn after a sufficient number of trials (regardless of conditions) that they could detect the salient target bar in their visual periphery by keeping their fixation at the center of the display. This strategy works since a salient pop out can be sensed in the visual periphery without recognizing the unique tilt orientation or the item shape at the pop out location. Obviously such a strategy, which avoids eye movements, would defeat our design to assess underlying mechanisms by examining gaze shift. To discourage this strategy, we randomly interleaved our trials of interest between trials associated with four control conditions that were more difficult. In the latter, there were two distractor types for each target, e.g., target N among distractors Z and I, making the target less salient than those in the conditions shown in Figure 1.

Gaze was tracked by the 50 Hz infra-red video eye tracker from Cambridge Research System (www.crs Ltd.com). Tracking calibration was performed just before data collection to a precision typically within 0.5° of visual angle. In each trial, subjects pressed a button for the fixation stimulus. Once gaze had stayed for 40ms within 3° of the central fixation point, a blank white screen replaced the fixation stimulus for 200 ms before the search stimulus was displayed till after the button response.

Data analysis:

We defined gaze arrival at the target when its position was within 1.2 times the unit grid distance, or 60

pixels, from the target's center position. RTs reported here were from trials in which the gaze had reached the target and a correct report had been given. The error bars in the plotted data designate the standard error of the mean. RT from one condition was deemed to be significantly larger than that of another when the one-tailed (matched sample when performed across subjects) t-test gave a p value less than 0.05.

If gaze never reached the target in more than 5% of the trials in a session, poor tracking accuracy was suspected, and the data in this session were removed from analysis. Some other trials were also defined as bad: when gaze was untracked in more than 10% of the tracker video frames between the search display onset and button response, or if the button response reaction time $RT_{\text{report}} < 100$ ms. Data from a session were removed from further analysis if more than 10% of trials were bad.

We differentiate two target/distractor conditions, condition 1 and condition 2, by their RT difference $RT(1)-RT(2)$, normalized by their RT sum $RT(1)+RT(2)$, to give the RT differential index

$$\text{RT differential index} = \frac{RT(1) - RT(2)}{RT(1) + RT(2)}$$

Here the RT maybe RT_{report} , RT_{gaze} , or RT_{lapse} . For example, if condition 1 is N/I to find N in I's, and condition 2 is I/N to find I in N's, then the reversed letter effect should give a positive RT differential index for RT_{report} ; and if this effect is caused by top-down but not by bottom-up saliency processes, the RT difference index should be positive for RT_{lapse} but indifferent from zero for RT_{gaze} . For convenience, in using the RT differential index as above, condition 1 is always the one with the more familiar target, or the less familiar distractor, or both, so that a familiarity based differentiation expected from conventional knowledge should make this index significantly larger than zero.

RTs and their asymmetries in N versus I

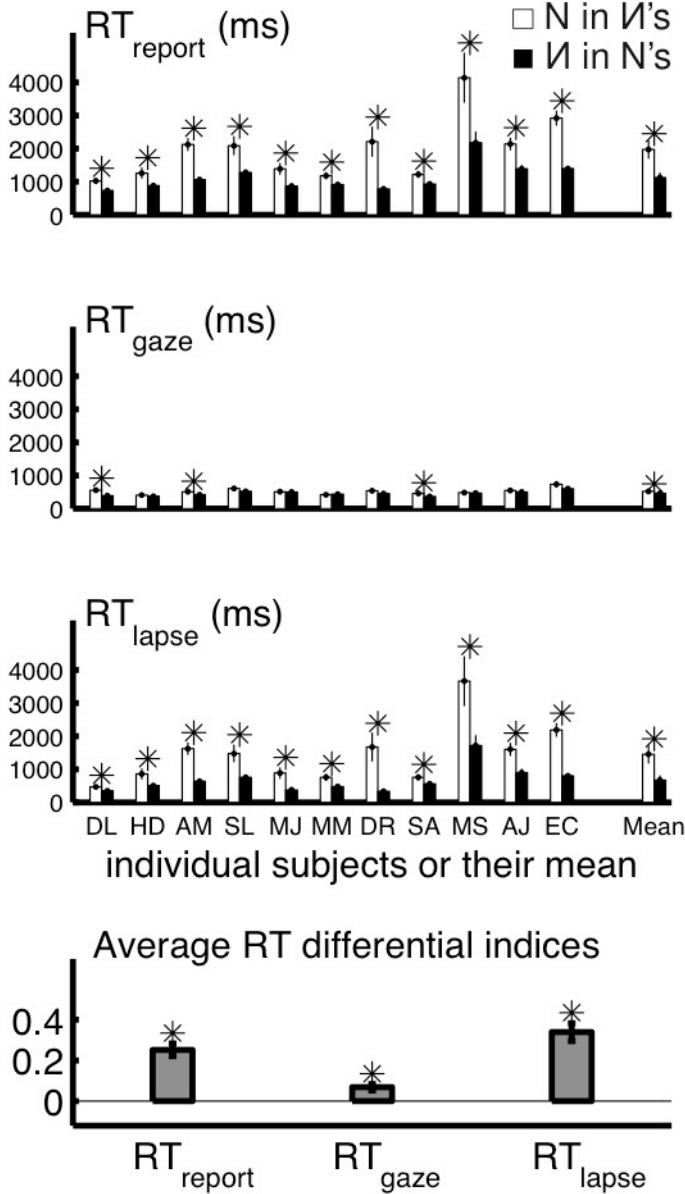


Figure 2. RTs and the asymmetry analysis in searches N/I (to find N in I's) and I/N (to find I in N's). In the top three rows are RT_{report} by button press (top) , its decompositions into RT_{gaze} (second row) , time for gaze to reach the target for the first time, and their difference RT_{lapse} = RT_{report} - RT_{gaze} (third row), the button press latency since the gaze arrival. These RTs are shown for 11 observers (DL, HD, ..., EC), and the mean RTs across these observers are shown at the right most position in the plots. White and black bars plot RTs for target N and target I searches respectively, as indicated by the legend. The bottom plot shows the RT differential indices $[RT(\text{target N}) - RT(\text{target I})] / [RT(\text{target N}) + RT(\text{target I})]$ averaged across subjects. In the plots, '*' denotes RT(target N) > RT(target I) significantly, or, an RT differential index as significantly positive. The fractions of bad trials for N/I and I/N are 0.014 and 0.027 respectively averaged across subjects.

Results

Of 15 subjects who searched both N/I (to find N in I's) and I/N (to find I in N's) conditions with sufficient eye tracking accuracy, 11 subjects had $RT_{\text{report}}(\text{N/I})$ significantly longer than $RT_{\text{report}}(\text{I/N})$. This confirms the robustness of the reversed letter effect. As we are only interested in the cause of this effect, we only analyze further the data from these 11 subjects in the various search conditions displayed in Figure 1. In none of the other four subjects was the lack of a significant reversed letter effect in RT_{report} caused by significant and opposite RT differentiations in RT_{gaze} and RT_{lapse} .

Asymmetry between N and I arises from search behavior after gaze arrival at target

All 11 subjects who showed the reversed letter effect had $RT_{\text{lapse}}(\text{target N})$ significantly larger than $RT_{\text{lapse}}(\text{target I})$, but only three had $RT_{\text{gaze}}(\text{target N})$ significantly larger than $RT_{\text{gaze}}(\text{target I})$, see Figure 2. For the RT difference $\Delta RT \equiv RT(\text{target N}) - RT(\text{target I})$ averaged across subjects, 92% of $\Delta RT_{\text{report}}$ is due to ΔRT_{lapse} . The RT differential index $[RT(\text{target N}) - RT(\text{target I})] / [RT(\text{target N}) + RT(\text{target I})]$ was largest in RT_{lapse} and smallest in RT_{gaze} , though all were significantly larger than zero. On average, $RT_{\text{lapse}} > RT_{\text{gaze}}$ in both target N and target I searches, suggesting that top-down processes involved in shape recognition dictate RT_{report} . This is particularly so in the target N search in which the average RT_{lapse} was more than double that of RT_{gaze} .

Figure 3 examines the causes for the long latencies RT_{lapse} . Fig. 3AB show an example trial. Here the gaze located the target within 1-2 saccades into the search. This was predicted since the uniquely tilted bar in the target is very salient and pops out to attract attention. The average RT_{gaze} is half a second in both target N and target I searches. However, in this example, the gaze dawdled hesitatingly around the target, and then, as if the decision by the bottom-up saliency was vetoed, it abandoned the target to continue searching elsewhere before it returned prior to another hesitation followed by the button press report. Trials with such arrival, abandon, and return gaze shifts before the button press are termed arrival-abandon-return (AAR) trials. A trial is categorized as an AAR trial only when, after reaching the target, the gaze deviated from the target by at least 2.4 times the average distance between neighboring items. This means that these trials are unlikely to be caused by subjects needing to compare the target with the immediately neighboring distractors to verify that the target was distinct. AAR trials occurred in both target N and target I searches, since the confusion, - both target and distractors seen simply as zigzags -, is possible in both cases, but they occur about twice as often in target N trials (Fig. 3C). Furthermore, among the non-ARR trials, RT_{lapse} in target N trials is also about twice as long as that in target I trials (Fig 3C bottom).

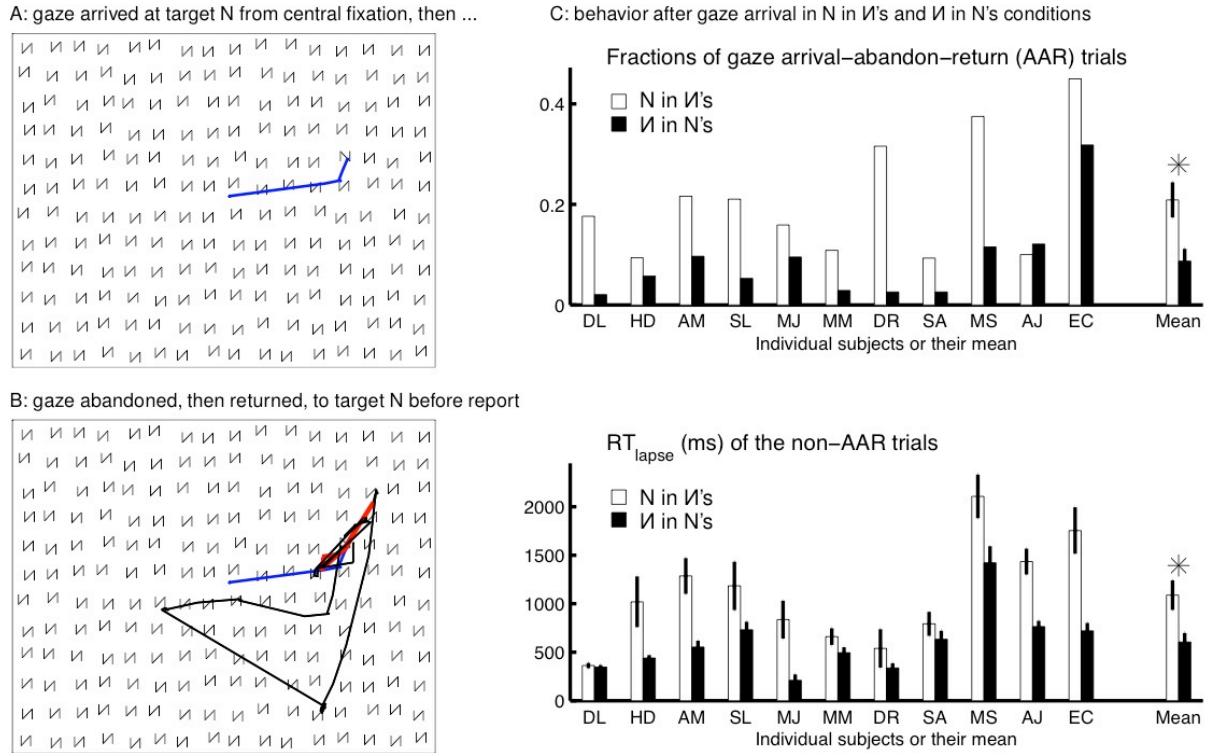


Figure 3 The N vs. I asymmetry after the gaze arrival to target. A and B: in a target N in I's trial, the gaze arrived (in blue scan path in A) from central fixation to the target after only 1-2 saccades upon stimulus onset, then (B), it hesitated (in red scan path), and, continuing in black scan path, abandoned, returned, and hesitated again before report. A trial having such target arrival, abandon, and return gaze shifts before button press is called an arrival-abandon-return (AAR) trial. C: for the 11 subjects and their averages (at the right most position), the target N search had a significantly larger fraction of AAR trials (top) and a longer RT_{lapse} in the non-AAR trials than the target I search. '*' denotes a significant difference between the subject means for the two searches.

Does familiarity of the distractors have a critical role in the search RT?

The effect of distractor familiarity could be isolated and assessed by comparing the RTs for two conditions having the same target but different distractors: one the familiar distractor N and the other the unfamiliar distractor Σ. This is shown in Fig. 4A and Fig. 4B, with I and Z as search targets respectively, using data from the ten observers, who had searched the corresponding conditions.

Except for RT_{report} by subject DL in target Z searches, the RT differences $\Delta RT = RT(\text{distractor } \Sigma) - RT(\text{distractor } N)$ is not significantly larger than zero for RT_{report}, RT_{gaze}, or RT_{lapse}. This is true for individual subjects and for the average across subjects. In comparison, as we noted in Figure 2, familiarity based RT difference in the reversed letter effect is significant for RT_{report} and RT_{lapse} within single subjects. There was another small and marginally significant difference in RT, which is not of central interest here, but deserves brief mention. The search for a target 'I' was slightly faster in distractors 'Σ's than in distractor 'N's. This is opposite to the expectation that search in familiar distractors is faster. Both distractors differ from the target 'I' by the 90 degree orientation difference of the oblique bar. However, distractor 'Σ' differs from the target 'I' additionally by the 90 degree orientation difference in the cardinal (horizontal/vertical) bars. This additional and yet redundant orientation feature difference may slightly raise the saliency of the target, and thereby shorten RT_{gaze}, a feature redundancy gain (Krummenacher, Miller, Heller 2001). This gain could also make RT(Z in Σ's) slightly longer than RT(Z in N's), as is in-

deed the case for subject DL (see Figure 4B). Nevertheless, none of the subjects needed more time to finally decide on the target; i.e., $RT_{\text{lapse}}(\text{Z in } \Sigma\text{'s})$ was not significantly longer than $RT_{\text{lapse}}(\text{Z in N's})$. The effect of such orientation feature redundancy gain on RT is known to be small (Zhaoping & May 2007). Averaged across subjects using our data in Fig. 4, it gives at most 19 ± 7 ms ($p=0.051$ for Fig. 4A) in the difference in RT_{gaze} or 100 ± 34 ms ($p=0.14$, for Fig. 4B) in the difference in RT_{report} . This is much smaller than the 844 ± 167 ms ($p=0.001$) in the difference in RT_{report} caused by the reversed letter effect (Fig. 2). Even though both the feature redundancy gain and the idea of a faster search in familiar distractors should make $RT(\text{Z in N's})$ smaller than $RT(\text{Z in } \Sigma\text{'s})$ (Fig. 4B), the difference between these two RTs is insignificant when averaged across subjects. Taken together these observations suggest that, with a given target, distractor familiarity plays a negligible role. They are consistent with previous finding that less familiar distractors do not require longer fixation durations during search (Greene and Rayner 2001).

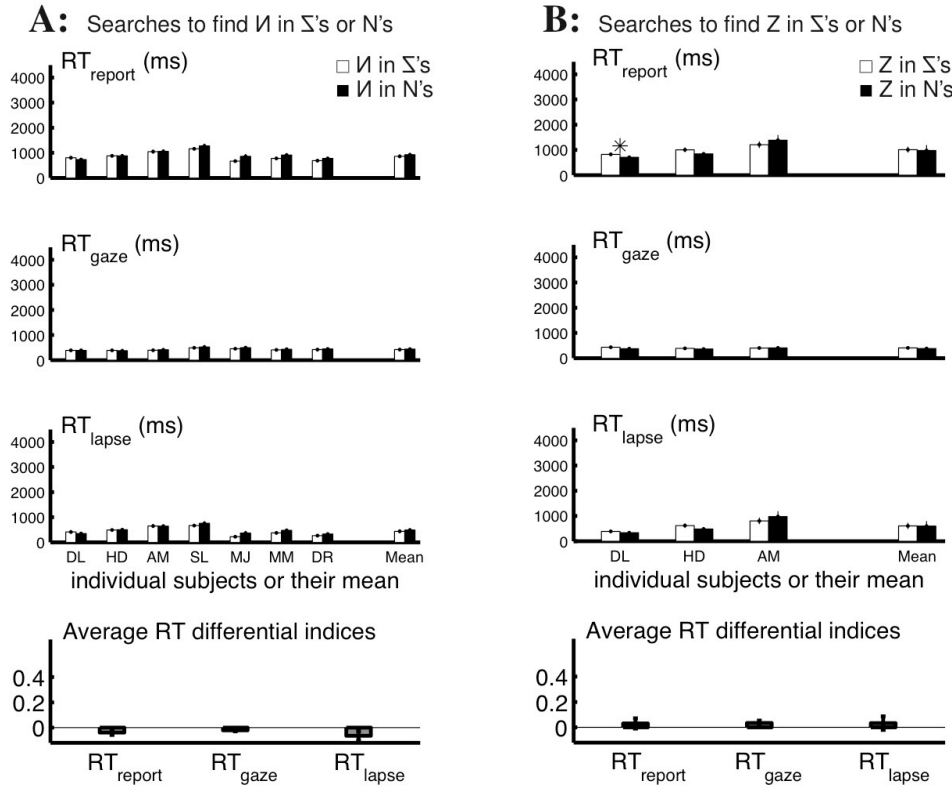


Figure 4. Little impact of distractor familiarity on RT. Each column of plots, A or B, is in the same format as the plots in Figure 2. Each column contrasts a search in unfamiliar distractors Σ with another in familiar distractors N for the same target. The target is I for column A and Z for column B. In each column, ‘*’ indicates that $RT(\text{distractors } \Sigma) > RT(\text{distractors N})$ significantly, or the RT differential index $[RT(\text{distractors } \Sigma) - RT(\text{distractors N})] / [RT(\text{distractors } \Sigma) + RT(\text{distractors N})]$ as significantly positive. For distractors Σ and N respectively, the fractions of bad trials averaged across subjects are 0.013 and 0.032 in A, and are 0.0133 and 0 in B.

The dominant role of target familiarity

Analogously, we contrasted search behavior for different targets among the same distractors to isolate and assess the role of target familiarity. Fig 5A contrasts a familiar target N with the unfamiliar target Σ in the same distractors I for all the (four) subjects who searched these conditions. The RT difference $\Delta RT = RT(\text{more familiar target}) - RT(\text{less familiar target})$ is significantly positive for RT_{report} and RT_{lapse} in

three out of four subjects or averaged across subjects, and is marginally significant ($p \leq 0.08$) in the fourth subject. On average, 95% of the $\Delta RT_{\text{report}} = 742 \pm 228$ ms arises from ΔRT_{lapse} . The RT differential indices (averaged across subjects) in RT_{report} and RT_{lapse} are comparable in magnitude to the corresponding RT differential indices in the N vs. I asymmetry (Fig. 2). Meanwhile, both Figure 5B and Figure 5C contrast the “familiar” target Z with the “unfamiliar” target I in the same distractors N (Figure 5B) or Σ (Figure 5C). The differential familiarity between the two targets remains doubtful. Consequently, none of the RT differences or the RT differential indices for target Z vs. target I were significantly positive, even when the effect of target familiarity should reinforce the effect of a feature redundancy gain when searching in distractors Σ (Fig. 5C). Our findings are consistent with the observation by Greene and Rayner (2001) that more saccades are involved in searching for a more familiar target, although these researchers did not distinguish between saccades before or after the gaze reaches the target.

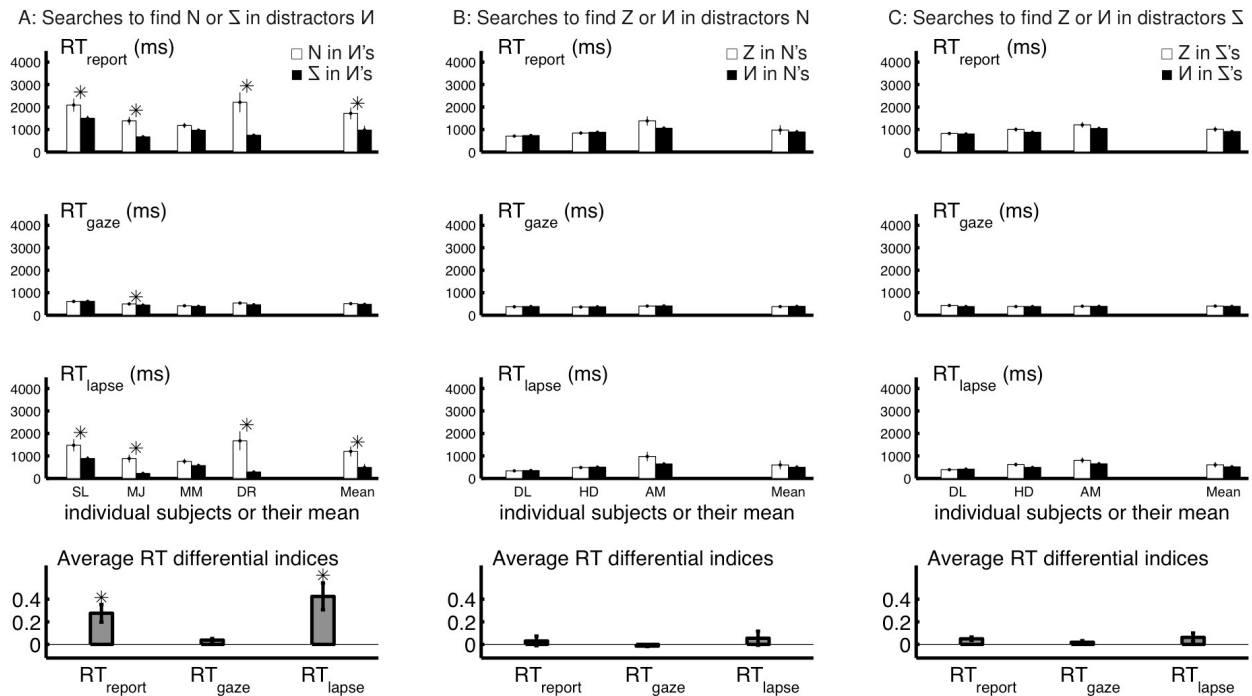


Figure 5. Big impact of target familiarity on RT_{lapse} but not on RT_{gaze} . Each column of plots, A, B, or C, is in the same format as that in Fig. 2. In each column, two search conditions in the same distractors, but for two different targets, are contrasted. A contrasts target N with target Σ among distractors I; B and C contrast target Z with target I, among distractors N in column B and among distractors Σ in column C. In each column, ‘*’ indicates that $RT(\text{more familiar target}) > RT(\text{less familiar target})$ significantly, or the RT differential index (between the two target conditions) as significantly positive. Averaged across subjects, the fractions of bad trials for the more and less familiar targets are, respectively, 0.026 and 0.057 in A, 0 and 0.038 in B, and 0.0133 and 0.014 in C.

Discussion

Our results rule out either of the two standard accounts (Treisman & Gormican 1988, Wolfe 2001) of the reversed letter effect. The first account assumes an advantageous pre-attentive saliency to the reversed letter N as target; the second assumes that more familiar distractors are easier to reject during search. Both accounts predict a reversed letter effect in the reaction time RT_{gaze} of the initial gaze landing and not the subsequent time RT_{lapse} of the time to reach decision to report. Instead, our data showed that the effect

resides almost entirely in the RT_{lapse} component of the RT after initial gaze landing. Thus the reversed letter effect appears to be located predominantly in top-down processes. We draw this conclusion notwithstanding the fact that we found a small but significant reversed letter effect even in RT_{gaze} . We believe this small effect was found because our dense arrays only reduced rather than completely blocked out top-down processes.

If we assume that RT_{gaze} mainly reflects saliency processes, then the present findings fit the V1 saliency account. The proposed V1 mechanism computes saliency via iso-feature suppression (Li 1999), and is mediated by neural connections that only stretch a short distance in the cortex or a short distance in the visual image. Only the uniquely tilted oblique bar in the target escapes iso-feature suppression, making it the most salient by the V1 hypothesis. From this account, gaze is expected to land equally fast on the oblique feature whether N or reserved N is the target.

The present findings have been obtained in dense search arrays. What would happen in less dense arrays? According to the V1 saliency hypothesis, in sparser search arrays target saliency will be weaker, since iso-orientation suppression is weaker because fewer neural connections are long enough to link neurons responding to different distractor items. More importantly, sparser arrays make it easier to recognize the target shape before the gaze reaches the target. In this case RT_{gaze} will be a poor indicator of the bottom-up saliency alone.

We conducted a second experiment to address the question what happens in less dense arrays and arrays of different set sizes. In particular we wanted to see what happens when top-down processes strongly interfere with bottom-up processes. Would this magnify the reversed letter effect?

Experiment 2

Previous studies of search asymmetry showed asymmetric set size effects (Wang, Cavanagh, & Green 1994, Malinowski & Hübner 2001, Wolfe 2001), such that RT increases more quickly with increasing number of distractors for target N than for reversed N. These studies used search arrays much sparser or smaller than those used in Experiment 1. We believe that sparser arrays are very different from denser ones. Consider the case when the array is so sparse that top-down processes concerned with shape recognition are strong and guide attention to the target, while bottom-up saliency of distinguishing features is weak. Top-down attentional guidance typically shifts attention to visual items serially, which explains why RT increases with set size. This set size effect should also hold for the second RT component when gaze reaches again the target after first abandoning it. In line with previous studies it is to be expected that this measure would show a stronger set size effect for target N than its mirror reversal in sparse arrays.

Consider the case when the arrays are dense enough, so that bottom-up saliency plays a dominant role to guide attention to target. Since saliency by V1 mechanisms increases with the density of the array, a denser array should have a shorter RT_{gaze} . Hence an inverse size effect should hold for RT_{gaze} . For the same reason as that for the sparser arrays, an inverse set size effect should also hold for RT_{lapse} , and consequently for RT_{report} , and it would be stronger for target N. It is likely that previous studies never saw such inverse set size effects as their search arrays were never dense enough. Exp. 2 was designed to test these predictions of a contrasting set size effects for sparse and dense arrays, with a stronger set size effect for target N; this allows a deeper probe into interactions between bottom-up and top-down processes during search.

Methods

Exp. 2 used the same eye tracking design and experimental procedures as Exp. 1. Eight new subjects participated and satisfied the tracking quality criteria for gaze tracking. Their task was the same as that in Experiment 1. Each subject searched for 600 trials, typically within 1.5 hours including 3 short breaks. These trials randomly interleaved about 50 trials each of the twelve conditions, which included six set sizes each for the two target/distractor conditions: N/I and I/N. The six set sizes are denoted by the number of rows x number of columns of the search items, including 2x2, 3x4, 6x8, 9x12, 12x16, and 18x24, such that, e.g., 3x4 means that the search array contains 12 search items. Note that here we are no longer worried about the build up of a search strategy which essentially takes advantage of bottom-up saliency, because the weaker saliency in sparse arrays discouraged such a strategy build up. All search items were 21x21 pixel images, slightly smaller than that Exp. 1 in order to accommodate all set sizes within the 600x800 pixel images. Thus, different set sizes differed by the average distances between nearby items. Each search item had its horizontal and vertical positions randomly jittered from the regular grid positions by an amount up to 35% ($D - d$) pixels, where D is the average distance between nearest items and $d = 21$ pixels is the size of each item. As most trials had sparser search arrays than that in Experiment 1, gaze was judged as having arrived at the target when it was within 120 pixels from the center of the target. All other aspects of this experiment are the same as those in Experiment 1.

Results

We found a reversed letter effect for all set sizes (though not significantly so for the smallest set size). In sparse arrays, the search RT increases with set size, and this set size effect is stronger for target N, in line with findings from previous studies (Wang, Cavanagh, & Green 1994, Malinowski & Hübner 2001, Wolfe 2001). In dense arrays, a hitherto unseen inverse set size effect appeared, i.e., search RT decreased with set size, and this effect was stronger for target N, as predicted from our analysis above. Hence, the reversed letter effect is amplified most in the intermediate set sizes. In smaller set sizes, the reversed letter effect in RT_{gaze} contributes to the reversed letter effect in RT_{report} by a larger weight.

We found that one subject, among our eight observers, could read a Slavic language, whose alphabet uses both N and reversed N as letter symbols. This is problematic because there is no reversed letter effect in Slavic readers (Malinowski & Hübner 2001). However, our subject had lived in a Slavic country until after the first year of primary school, and since her teens she had been living and studying in England with English as her dominant language. Accordingly, she showed a strong reversed letter effect just like the other subjects.

Fig. 6A shows the average RT_{report} , RT_{gaze} , and RT_{lapse} across the seven non-Slavic reading observers for the six set sizes. For comparison, Fig. 6B shows the corresponding result for the Slavic reader. Note that the scales for the vertical axes in Fig. 6A differ from that in Fig. 6B. The pattern of the RTs for the single Slavic reader is similar to that for other subjects, but her RTs are substantially longer. Furthermore, her longer RT_{report} are mainly due to the longer RT_{lapse} , and her RT_{gaze} are comparable to those of the non-Slavic readers. Thus it seems that she was not intrinsically slower, but had a stronger top-down interference. This is likely due to her extra familiarity with the target in both the N/I and I/N searches. In fact, variability between non-Slavic readers is also mainly manifest in their variability (the error bars on the averages across subjects) in RT_{lapse} rather than RT_{gaze} , and this was also the case in Experiment 1 (see Fig.2). Presumably, bottom-up processes are affected directly by the stimuli presented, and this is the same for all observers. On the other hand, top-down processes are affected by long term experience, and this differs between individuals.

Figure 6A also shows that RT_{lapse} is typically longer than 500 ms for target N of all set sizes, and longer than 400 ms for target reversed N except for the two densest search arrays. This confirms the notion that interference from shape confusion is stronger for target N, and is present regardless of the strength of bottom-up saliency. The reversed letter effect was enhanced for sparser arrays and was substantially manifested even in RT_{gaze} . This was predicted because without visual crowding the target shape could be recognized quickly, and hence top-down interference started even before gaze reached the target.

This experiment confirmed the extra strength of top-down processes in visual search for familiar targets and provides further evidence of the robustness of the reversed letter effect.

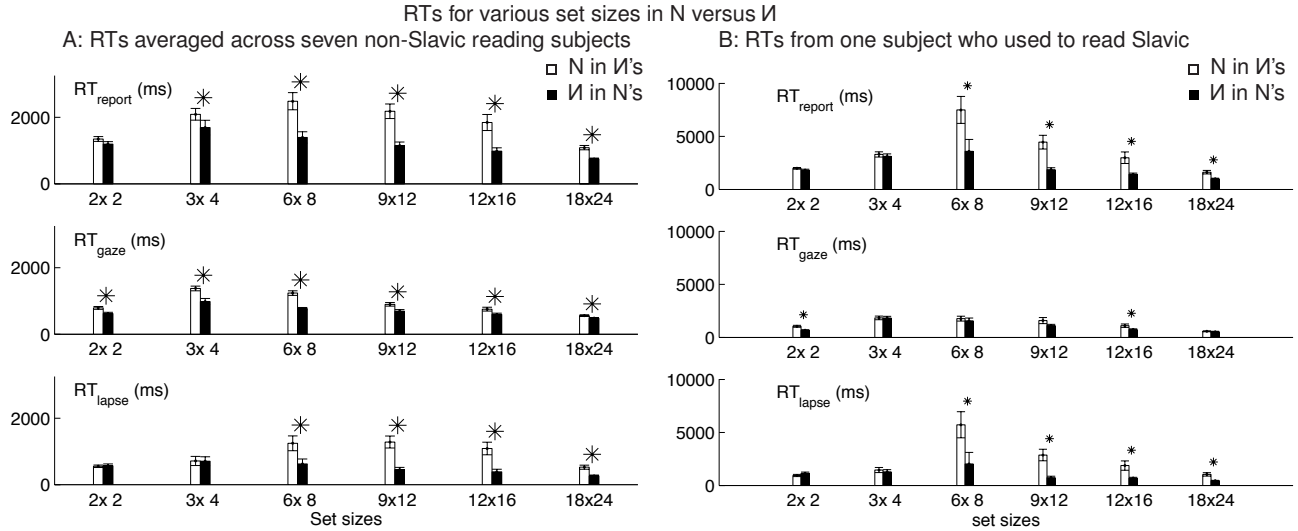


Figure 6. The conventional set size effect for sparse arrays and an inverse set size effect for dense arrays in searches N/I and I/N. A shows RTs averaged from the mean RTs of seven subjects who could not read Slavic. B shows the corresponding result from one subject who used to read Slavic. The set sizes are marked as the number of rows x number of columns of the search items. A ‘*’ marks a significant difference between the two RTs of the same set size but different target conditions. This significant difference is across subjects in A (by matched sample t-test), but is by a t-test over RTs for the corresponding trials in B. For searches N/I and I/N, the fractions of bad trials are, respectively, 0.038 and 0.030 in A, and 0.083 and 0.092 in B.

General Discussion:

Before we consider possible explanations of the reversed letter effect, we need to highlight the special nature of the present visual search task. As an observer in the task you cannot do better than relying on the bottom-up decision of your attention or gaze to land on the oblique feature in the target. Unfortunately for you, at the same time or very soon after, your top-down processes are also active and identify the shape of the target. You now find that it is confusingly the same as that of the distractors. Consequently, the extra time the observer takes to finally report the target is considerable. In our experiment 1, after your gaze first located the target, you would take on average an extra 500 ms or more to report the reversed N target, and an extra 1000 ms or more to report the familiar N target. This extra time is longer than typical inter-saccadic intervals of around 330 ms, and longer than it takes to report the appearance of a visual item anywhere in a blank field (300 – 400 msec, Koene & Zhaoping 2007).

We still have few clues as to what happens in the brain after first gaze landing. We speculate that there are at least the following steps in processing: first, the target is recognized as a zigzag shape with a particular orientation; second, the target’s orientation is ignored and its zigzag shape is extracted regardless of its orientation. This causes the letter N target to be confused with the reversed N as a distractor as both have the same zigzag shape. In a third step this confusion is overcome. The reversed letter effect may reside in one or all of these steps.

Why does the final decision in the search task take so much longer when the target is the familiar letter

N? Familiarity appears to play a key role, but a paradoxical one. One explanation might be that the first step to recognize the target's shape at a particular orientation is faster when this orientation is familiar. There is indeed ample evidence that objects in their more familiar or canonical orientations are identified or verified faster (Lawson 1999). This speeded recognition could lead to an earlier onset of the subsequent confusion or interference when the target orientation seems to be temporarily ignored. The small difference in the onset time of interference may have big consequences in decision time, since the earlier the interference starts, the stronger the effect, and the slower the decision. A second explanation might be that the reversed N may be more distinctive at a higher cognitive level, simply because its unfamiliarity makes it more novel. This distinctiveness might weaken the shape confusion, which is then easier to overcome. These explanations are not mutually exclusive. Both are consistent with our data in Fig. 5, which show that the interference was weaker for a less familiar target. Thus the final report was faster for the target reversed Z, which for English readers is less familiar than the target N. Unfortunately, until we have the techniques to segment gaze behavior and the accompanying neurophysiological processes further, the explanation to the reversed letter effect remains elusive.

By isolating the bottom-up processes in visual search through the use of dense search arrays, our study identified that the reversed letter effect is mainly caused by the target-distractor shape confusion in the top-down process when attention has reached the search target. More specifically, the target can be detected by bottom-up saliency of its uniquely oriented bar, guiding attention to the target without recognizing its shape. An almost absent reversed letter effect in the RT for gaze to reach the target in a dense array suggests that bottom-up saliency plays little role in this effect.

Our findings indicate that the reversed letter effect does not invalidate the V1 saliency hypothesis (Li 1999, 2002) which predicts no such effect by bottom-up saliency mechanisms in V1. They rule out previous accounts of this effect by pre-attentive saliency advantage for the unfamiliar target or easier rejection of the familiar distractors (Treisman & Gormican 1988, Wolfe 2001). Understanding the respective roles of the top-down and bottom-up processes in search also enables us to understand an inverse set size effect in visual search when the search arrays are sufficiently dense.

While more evidence is emerging for V1 as a neural basis for saliency, the neural basis for viewpoint invariant shape recognition is far from clear (Lawson 1999). Main phenomenological accounts include viewpoint invariant feature recognition (e.g., recognizing tigers by their stripes), multiple view representation (to match an image to the nearest view-specific stored representation of an object), and retinal image transformation (of input image to a view-specific stored representation). Physiological data indicate that neurons in inferotemporal (IT) cortex are tuned to specific viewpoints of objects (Logothetis, Pauls, & Poggio 1995) and only a very small number of IT neurons responded in a viewpoint invariant manner. Others reported that IT neurons have some degree of invariance for position and scale of the object (Tanaka 2003, Rolls 2003). A recent study (Oliveri, Zhaoping, Mangano, Turriziani, Smirni, & Cipolotti 2010) found that transcranial magnetic stimulation to the right parietal cortex (presumably impairing its function) reduced the RT for observers to find the target \nearrow which is a rotated version of the distractors \nwarrow and \searrow . This suggest that the right parietal cortex is involved in viewpoint invariant recognition leading to top-down interference, perhaps through its role in directing top-down attention necessary for viewpoint invariant recognition.

The present study points to an important limitation in testing the V1 saliency hypothesis by visual search behavior. The hypothesis concerns only bottom-up saliency, whereas search behaviors mostly result from a combination of both bottom-up and top-down factors. Hence, the V1 hypothesis should not try to explain other typical visual behaviors, including other examples of visual search asymmetries (Wolfe 2001, Rauschenberger and Yantis 2006). The current work also highlights a problem in the study of visual search if one only measures how long it takes a subject to make a report. This measure of reaction time is merely one outcome of multiple, dynamically interacting, brain processes, some involved in bottom-up attentional shift, some in top-down attentional guidance, and others in decision making. Additional measurements, such as the gaze shifts and brain waves, will hopefully permit the dissection of the complex

and interacting components behind this deceptively simple behavior.

Acknowledgements We thank Chris Frith for linking the two authors together. We also like to thank Chris Frith, Eyal Reingold, Jeremy Wolfe, and an anonymous reviewer for very helpful comments on the manuscript. We particularly like to thank Jeremy Wolfe for suggesting to examine the set size effect. This work is funded by a grant from the Gatsby Charitable Foundation and a Cognitive Science Foresight grant BBSRC #GR/E002536/01, and by the Aarhus University Research Foundation.

References

- Duncan, J. & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, 96, 433-458. [PubMed]
- Frith U. (1974) A curious effect with reverse letters explained by a theory of schema. *Perception & Psychophysics* 16(1):113-116.
- Greene HH & Rayner K (2001) Eye movements and familiarity effects in visual search. *Vision Res.* 41:3763-73.
- Hoffman J.E.(1998) Visual attention and eye movements. in Attention Ed. H. Pashler, Psychology Press, page:119-154
- Jingling L, & Zhaoping L. (2008) : Change detection is easier at texture border bars when they are parallel to the border: evidence for V1 mechanisms of bottom-up saliency. *Perception*. 37(2):197-206.
- Koene A.R. & Zhaoping L (2007) Feature-specific interactions in saliency from combined feature contrasts: Evidence for a bottom-up saliency map in V1. *Journal of Vision*, 7(7):6,1-14. <http://journalofvision.org/7/7/6/>, doi:10.1167/7.7.6 [pubmed]
- Koch, C. & Ullman, S. (1985) Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*. 4(4), 219-227. [PubMed]
- Itti L. & Koch C. (2000) [A saliency-based search mechanism for overt and covert shifts of visual attention](#). *Vision Research* 40(10-12):1489-506
- Krummenacher J, Miller HJ, & Heller D 2001. Visual search for dimensionally redundant pop-out targets: evidence for parallel-coactive processing of dimensions. *Percept Psychophys*. 63(5):901-17. [PubMed].
- Lawson R. (1999) Achieving visual object constancy across plane rotation and depth rotation. *Acta Psychologica* 102:221-245.
- Legge GE, Mansfield JS, & Chung STL. (2001) Psychophysics of reading XX. Linking letter recognition to reading speed in central and peripheral vision. *Vision Research*, 41:725-743
- Levi DM (2008). Crowding --- an essential bottleneck for object recognition: a mini-review. *Vision Research* 48:635-654.
- Li Z. (1999) Contextual influences in V1 as a basis for pop out and asymmetry in visual search. *Proc. Natl Acad. Sci USA*, 96(18):10530-5. [PubMed]
- Li Z. (1999b) Visual segmentation by contextual influences via intra-cortical interactions in the primary visual cortex. *Network* 10(2):187-212 [PubMed]
- Li Z, (2000) Pre-attentive segmentation in the primary visual cortex. *Spatial Vision*, 13(1) 25-50. [PubMed]
- Li, Z. (2002). A saliency map in primary visual cortex. *TRENDS in Cognitive Sciences*, 6, 1, 9-16. [PubMed]
- Logothetis N.K. Pauls J., & Poggio T (1995) Shape representation in the inferior temporal cortex in monkeys. *Current Biology*, 5, 552-563.
- Malinowski P., & Hübner R (2001) The effect of familiarity on visual search performance: Evidence for learned basic features. *Perception & Psychophysics*, 63:458-463.
- Motter BC & Belky EJ (1998) The zone of focal attention during active visual search. *Vision Research*. 38(7):1007-22

- Nothdurft HC (2000) Saliency from feature contrast, variations with texture density. *Vision Research* 40:3181-3200.
- Oliveri M, Zhaoping L, Mangano GR, Turriziani P, Smirni D, & Ciolotti L. (2010). Facilitation of bottom-up feature detection following rTMS-interference of the right parietal cortex. *Neuropsychologia*. 48(4):1003-10.
- Rauschenberger R & Yantis S. (2006) Perceptual encoding efficiency in visual search. *J. Exp. Psychol. Gen.* 135:116-131.
- Rolls, E.T. (2003) Invariant object and face recognition. In the *Visual Neurosciences* Vol. 2. L. M. Chalupa and J. S. Werner, eds. (Cambridge, MA: MIT press), pp. 1165-1178.
- Saiki J (2008) Stimulus-driven mechanisms underlying visual search asymmetry revealed by classification image analysis. *J. of Vision*, 8(4):30.1-19.
- Saiki J, Koike T, Yakahashi K, & Inoue T. (2005) Visual search asymmetry with uncertain targets. *J. Exp. Psychol. Hum. Percept Perform.* 31:1274-87.
- Shen J & Reingold EM (2001) Visual search asymmetry: The influence of stimulus familiarity and low-level features. *Perception & Psychophysics*, 63:464-475.
- Stankiewicz BJ, Hummel JE, & Cooper EE (1998) The role of attention in priming for left-right reflections of object images: Evidence for a dual representation of object shape. *Journal of Experimental Psychology* 24(3): 732-744.
- Tanaka K (2003) Inferotemporal response properties. In the *Visual Neurosciences* Vol. 2. L. M. Chalupa and J. S. Werner, eds. (Cambridge, MA: MIT press), pp. 1151-1164.
- Treisman, A. & Gelade, G. (1980) A feature integration theory of attention. *Cognitive Psychology*. 12, 97-138. [PubMed]
- Treisman A & Gormiclan S (1988) Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, 95:15-48.
- Treisman A & Souther J. (1985). Search asymmetry: a diagnostic for pre-attentive processing of separable features. *Journal of Experimental Psychology: General*, 114:285-310.
- Van Zoest W & Donk M. (2006) Saccadic target selection as a function of time. *Spatial Vision* 19:61-76.
- Wang Q, Cavanagh P, & Green M (1994) Familiarity and pop-out in visual search. *Perception & Psychophysics* 56:495-500.
- Wolfe J.M. (1998) [Visual Search](#), In: Pashler H., editor. *Attention*. London UK: University College London Press
- Wolfe, J. M., Cave, K. R. & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model of visual search. *Journal of Experimental Psychology: Human Perception and Performance*. 15, 419-433. [PubMed]
- Wolfe J. M. (2001) Asymmetries in visual search: an introduction. *Perception & Psychophysics*, 63(3):381-389.
- Zhaoping L & Snowden RJ (2006) A theory of a saliency map in primary visual cortex (V1) tested by psychophysics of color-orientation interference in texture segmentation. *Visual cognition*. 14(4/5/6/7/8):911-933.
- Zhaoping L. & Guyader N. (2007) Interference with bottom-up feature detection by higher-level object recognition. *Curr. Biol.* 17(1):26-31. [PubMed]
- Zhaoping L. & May K. A (2007), Psychophysical tests of the hypothesis of a bottom-up saliency map in primary visual cortex. *Public Library of Science, Computational Biology* 3(4):e62. doi:10.1371/journal.pcbi.0030062. [PubMed]
- Zhaoping L. (2008) Attention capture by eye of origin singletons even without awareness --- a hallmark of bottom-up saliency map in the primary visual cortex. *Journal of vision*, 8(5):1, 1-18, <http://journalofvision.org/8/5/1/>, doi:10.1167/8.5.1. [Article]