Interference with bottom-up feature detection by higher-level object recognition

In press for *Current Biology* 2006

Li Zhaoping and Nathalie Guyader
Department of Psychology, University College London, UK

Running head: Object-to-feature interference in visual search.

Corresponding author: Li Zhaoping, z.li@ucl.ac.uk

1 Summary

Drawing portraits upside down is a trick that allows novice artists to reproduce lower level image features such as contours with less interference from higher level face cognition. Limiting the time available for processing to be sufficient for lower but not higher level operations would be a more general way to reduce interference. We elucidate this interference using a novel visual search task requiring a target to be found among distractors. The search target had a unique lower level orientation *feature*, but was identical to distractors in its higher level *object shape*. Through bottom up processes, the unique feature attracted gaze to the target[1, 2, 3]. Subsequently, viewpoint invariant object recognition[4, 5] interfered, with the attended object being recognized as identically shaped as the distractors. Consequently, gaze often abandoned the target to search elsewhere. If the search stimulus was extinguished at time T after the gaze arrived at the target, reports of target location were more accurate for shorter (T$<$500 ms) presentations. This object-to-feature interference, though perhaps unexpected, could underlie common phenomena such as the visual search asymmetry that finding a familiar object, e.g., a letter, among its mirror images is more difficult than the converse[6]. Our results should enable additional examination of known phenomena and interactions between different levels of visual processes.

2 Results and Discussion

Among the 45^o left tilted bars in Fig. 1, a uniquely right tilted bar, 45^o or 20^o from vertical in conditions A_{simple} or B_{simple}, pops out. However, superposing a horizontal or vertical bar on each original bar makes the uniquely tilted bar much harder to find in condition A than condition B of Fig. 1. The target object in condition A, but not B, is a rotated and/or mirror reversed version of all distractor objects, easily confused with the distractors since object recognition is typically rotationally or viewpoint invariant. We suggest that the higher level perception of the object comprising the two intersecting bars interferes with the task of locating it based on its unique lower level orientation feature component.

Primitive features, like the orientations of small bars, of visual inputs are first extracted by primary visual cortex (V1) [7]. Then these features are combined into objects, e.g., of two intersecting bars[8, 9], by higher cortical areas, including inferotemporal (IT) cortex, whose neurons are selective to object shapes[10, 11, 12, 13, 14, 15]. V1 is not only a way station, but also its activities highlight salient items due to its sensitivity to unique low level features such as orientation[16, 17, 18]. In addition to driving the higher visual areas such as V4 which combines bottom-up and top-down factors[19, 20, 21], V1's saliency signal also evokes cognitive decisions by driving superior colliculus which controls saccades[3]. Behaviorally and pre-attentively, unique image features such as orientation and color can pop out[1], and an object's basic features like "vertical" and "red", but not its overall shape, can be obtained[22]. Meanwhile, an important characteristic of the progression from feature to object processing is making object recognition viewpoint independent[23], achieving object invariance. Some IT neurons are indeed insensitive to viewpoint[10, 11, 12]. IT activities also correlate with the planning of saccades[24]. There is thus a hierarchy of levels of cognition, and their consequent decisions and actions. Behaviorally, attentive exposure to an object's image can prime its subsequent recognition regardless of viewpoint, but only in the same view if the exposure was unattended[4].

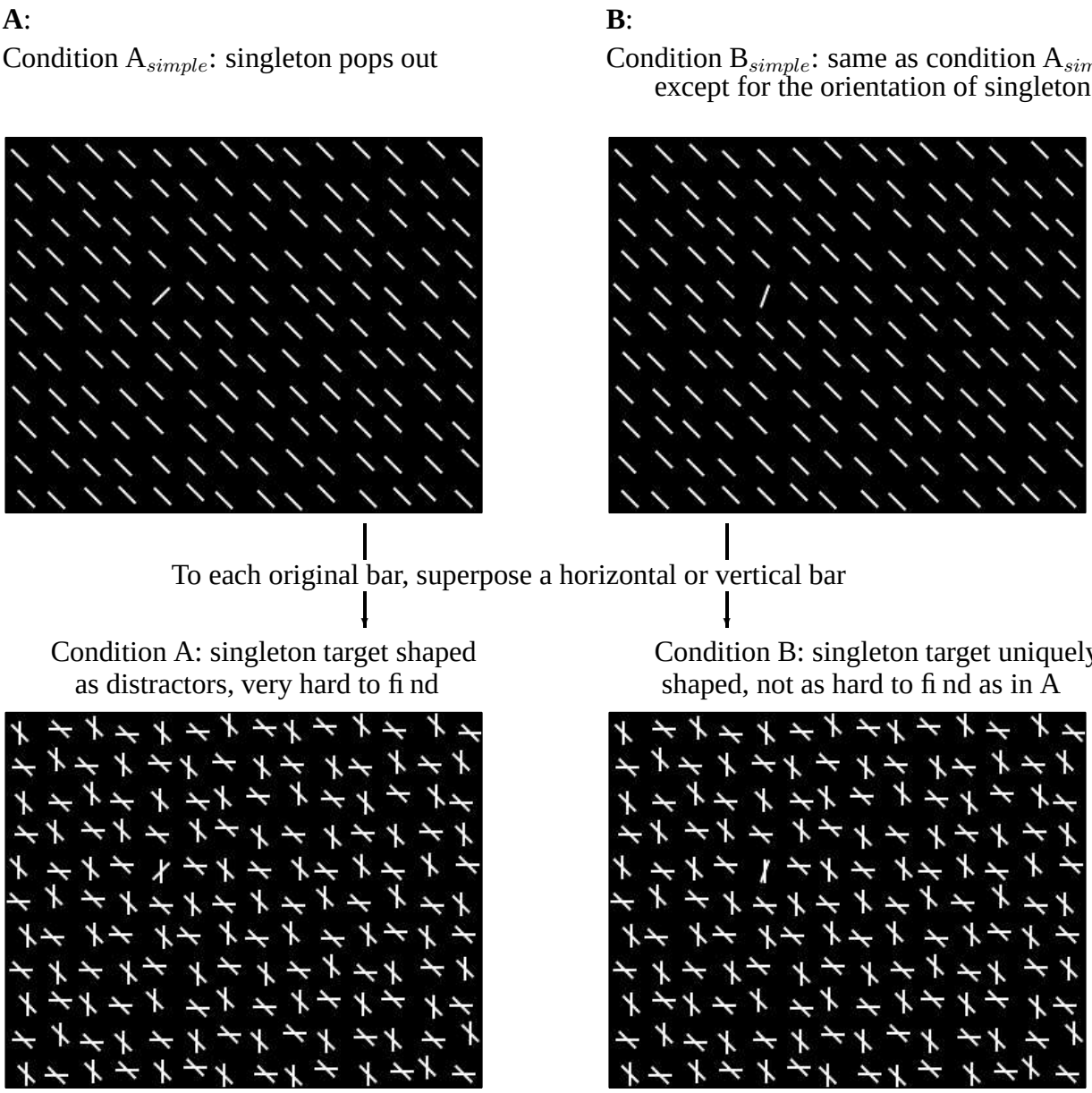


Figure 1: Small portions of visual search displays. The target possesses the uniquely left- or right-tilted (as in these examples) bar in the entire display. In conditions A_{simple} and B_{simple} (top), all bars were 45^o from vertical, except the target bar in B_{simple} was 20^o from vertical (in this example) or horizontal. Conditions A and B (bottom), derived from A_{simple} and B_{simple}, differed only in the angle, 45^o and 20^o respectively, between the two bars in the target. Task irrelevant, horizontal/vertical, bars made the orientation singleton much harder to find in condition A than in condition B.

The observations above suggest the following relevant processing stages: (1) an early pre-attentive stage processing image features, e.g., orientations of object components, and making unique features salient[1]; and, (2) a later, attentive[4], stage creating a viewpoint invariant object representation[1, 5], e.g., a shape from two intersecting bars. For locating a target possessing

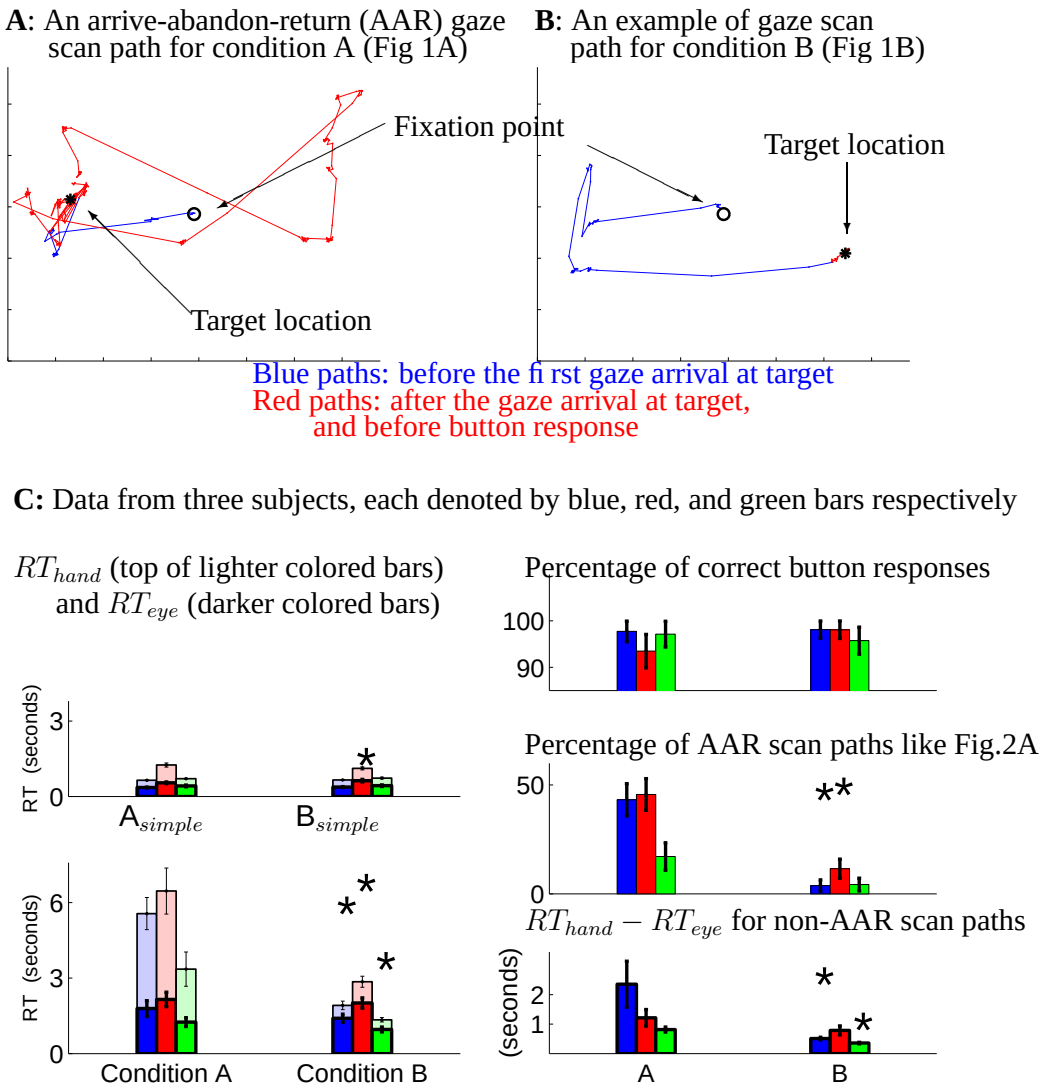


Figure 2: Hand and gaze responses in Experiment I. **A, B**: Examples of gaze scan paths. The '*'s and 'o's mark the locations for targets and fixation points respectively; the grid frames the spatial extent of the stimuli. **C:** Data for three subjects, denoted by red, green, blue hues respectively. '*'s indicate significant differences between conditions for the subject. Left: RT_{hand} and RT_{eye} (top of lighter and darker colored bars respectively) for button responses and first gaze at target respectively. Right: only for conditions A and B, task performances, percentage of arrive-abandon-return (AAR) scan paths (e.g., Fig. 2**A**), and eye-to-hand latencies in non-AAR trials. All error bars show standard errors of the means (s.e.m).

a uniquely oriented bar in the display, the early stage suffices, the salient unique orientation can attract gaze. The later, attentive, object processing stage is commonly expected to facilitate processing of the components of the objects through top-down feedback[25]. However, when differently oriented but otherwise identical distractors are present, as in condition A but not B of Fig. 1, viewpoint invariant object recognition could make search harder. If so, briefer viewings (within a time window) of the stimulus, preventing invariant object recognition, should improve target localization in condition A but not B. We show exactly this below.

In Experiment I, subjects searched among 660 objects, in a display extending $46^o \times 34^o$ of visual angle, for the one with the uniquely tilted oblique bar. The search stimulus was of conditions A_{simple}, B_{simple}, A, or B (Fig. 1) or control conditions. The non-oblique, task irrelevant, bar in the target of condition A or B was randomly either horizontal or vertical (the task relevant

bar in condition B was always 20^o from this irrelevant bar). The subjects were apriori informed about the uniquely oriented target bar, and that this unique orientation could be randomly tilted to the left or right in each trial. They were asked to press a left or right button quickly to indicate whether the target was in the left or right half of the display. Their eye positions were tracked.

Fig. 2 shows that reaction times (RT's) for the subject's first gaze arrival to the target, RT_{eye}, were comparable in conditions A and B. This is unsurprising since the target in both conditions had the uniquely oriented bar. This bar is salient pre-attentively[1, 2], attracting both attention and gaze, the latter due to the mandatory link between the directions of attention and gaze in free viewing[26]. These RT_{eye}'s were longer than those in conditions A_{simple} and B_{simple} mainly because the non-uniform orientations of the task irrelevant (horizontal and vertical) bars reduced the target saliency[27]. However, the RT's to report the target location by button press, RT_{hand}, were typically more than $1-2$ seconds longer in condition A than B, even though A and B had comparable button response accuracies. In condition A, after gaze first reached the target, it often dawdled around the target before the button presses, or even abandoned the target to search elsewhere before returning to it prior to button press (Fig. 2**A**). Such arrive-abandon-return (AAR) scan paths were much rarer in condition B. Even for the non-AAR trials, the eye-to-hand latency $RT_{hand} - RT_{eye}$ was much longer in condition A than in B. These observations are consistent with the hypothesis that decision processes veto-ed the first guess by the feature detectors in condition A because the attended object was recognized as having the same shape as the distractors, i.e., invariant object recognition could be interfering.

Alternative explanation consistent with the data could be that somehow targets in condition A, but not condition B, becomes less visible under foveal than peripheral viewing. To test between the hypothesis of interference by invariant object recognition and foveal visibility reduction, we examined conditions A and B in Experiment II in which the search stimulus was masked after a seemingly random time interval since its onset (Fig. 3A). The subjects button-pressed for target location as before, but could respond without time pressure, before or after the mask onset, guessing if they had to. The mask (Fig. 3B) covered each original object, target or distractor, with a star-shaped object, making the original object imperceptible. A random half of the trials in each session were gaze contingent trials, in which mask onset occured, reducing visibility of the original stimulus to zero, at one of several pre-determined time intervals T after the gaze first arrived at the target. The other trials had random mask onset times, some were gaze-opposite trials in which gaze and target positions at mask onset were such that one was on the left and the other on the right side of the display center, designed to prevent subjects' awareness of any link between mask onset and eye position (see Experimental Procedures).

Fig. 3**C** shows that for condition A, target localization worsened with longer gaze-to-mask viewing time T $\leq$ 1-2 seconds. This is not because the button presses tended to agree with the eye positions at mask onset, since, among the gaze-opposite trials, only 56% of the button presses agreed with the eye positions at mask onset. Furthermore, the performance for T$=0$, when target visibility became zero immediately upon foveal viewing, is comparable to that without the mask in Experiment I when the stimulus was viewed as long as deemed necessary by the subjects. This suggests that the extra viewing time T>0, or a longer duration of target visibility (even if reduced), is unnecessary, and can be detrimental for target localization for some T. Apparently, the subjects had a good first guess of the target location based on image features (orientations of the bars) alone, before they got confused by invariant object recognition which likely caused them to abandon target (the non-GSBM trials in Fig. 3C) and give incorrect responses. Eventually, their confusion subsided. Some subjects reported that sometimes they thought they found the target, only for it to disappear when they took a second look. In experimental sessions interleaving conditions A and B (for another group of subjects), extra viewing time $T > 0$ improved performance in condition B marginally, but worsened performance in A (Fig. 3**D**). Meanwhile, comparable performances for the two conditions for $T = 0$ is consistent with the comparable RT_{eye}'s in these conditions in Experiment I (Fig. 2).

While the trick about portrait drawing hints at something similar, our finding is the first we know of providing quantitative psychophysical data to suggest that deeper cognitive processing can be detrimental to some visual cognitive tasks. In particular, invariant object recognition interfered with lower level feature processes' abilities to detect unique salient features. Here, the later stage processes for object recognition are at best unnecessary for our task. Our findings suggest that they actually overwrite or interfere with the decisions of the necessary and earlier feature processes, even though, in principle, they do not have to do so. The uniqueness of the orientation of the target component bar is sufficient to make the target location salient. Previous physiological and computational studies[16, 17, 18, 2] have indicated that V1 can detect and highlight such a salient feature and direct gaze to it via superior colliculus[3].

Although some forms of object recognition can occur quickly[28, 29] and without attention or awareness[5, 30], psychophysical data have indicated that viewpoint invariant object representation needs attention[4, 22]. Accordingly, our findings suggest that the later, interfering, stage does not only construct object from features, but also allows top-down attention to build

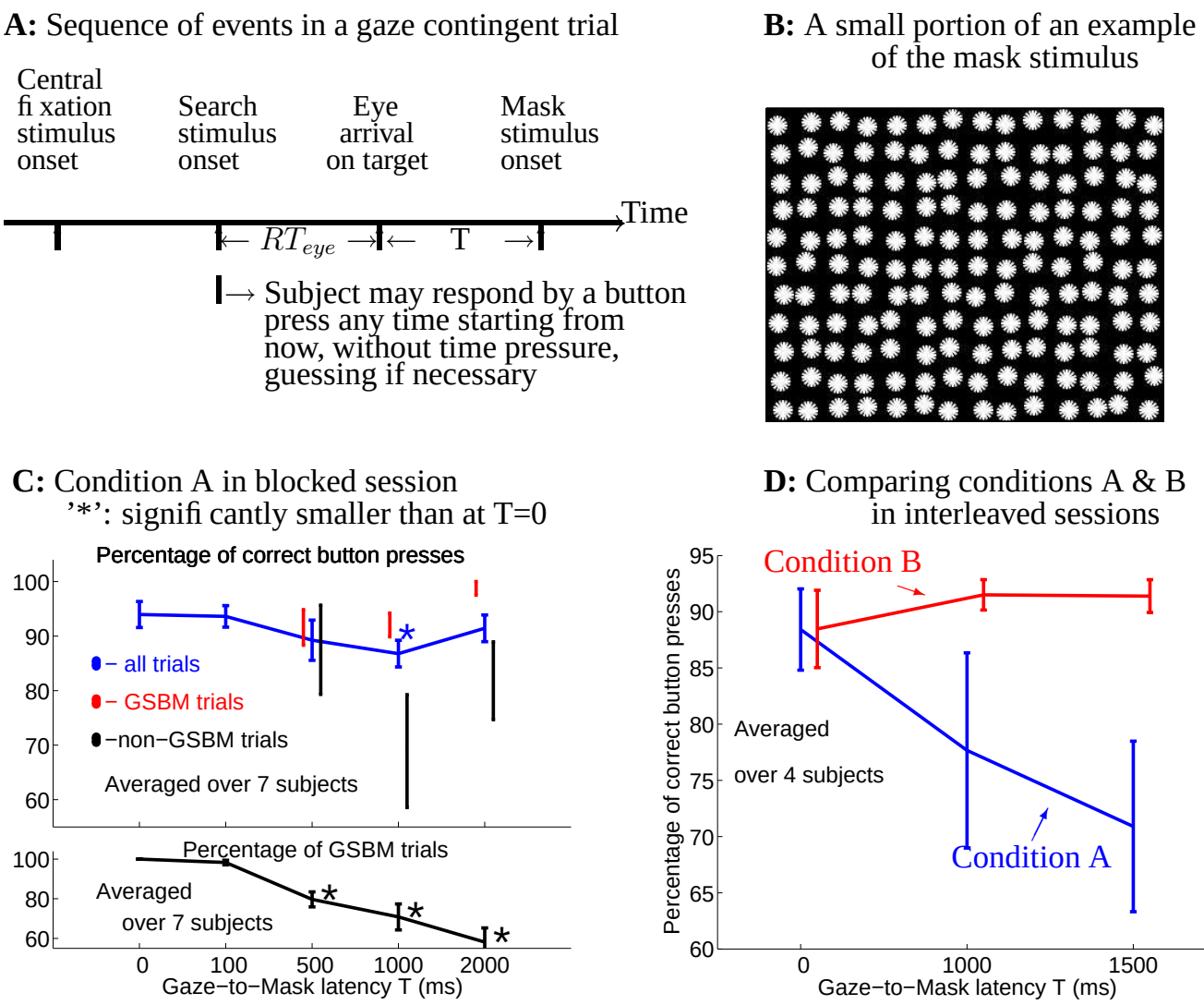


Figure 3: Experiment II: the longer one looks, the worse one "sees". **A:** Sequence of events in a gaze contingent trial. **B:** An example mask stimulus. **C:** With longer gaze-to-mask latency T, target localization in condition A (in blocked sessions) worsened, and the gazes are more likely to have abandoned the target before mask onset. GSBM trials are those in which *gaze stayed (at target) before mask* onset. **D:** In sessions interleaving conditions A and B (for another subject group), performances in conditions A and B are comparable for $T = 0$. Combining both $T > 0$ values, performance in B is significantly better than that in A (p=0.01). Error bars show s.e.m.

invariant object representations. This is consistent with the mandatory link between the directions of gaze and attention in free viewing[26]. Thus our finding can also be seen as top-down attentional processes interfering with the bottom-up processes, and introduces non-trivial complexity to the temporal and performance differences between higher and lower level processes[31, 32, 33]. Our finding also contrasts with backward visual masking[34] in which inattention enables a mask to impair object recognition. Fig. 3**C** suggests that building the invariant object representation requires at least 100 ms of attentive viewing for objects in our stimuli.

Our analysis suggests the following factors as being conducive to interference: (1) tasks being feature based, not requiring object recognition; (2) object recognition and/or top down knowledge introducing additional signals which has sufficient weight to counteract the low level feature's contribution to task-relevant decisions. Comparing condition A in blocked vs. interleaved (with condition B) sessions (Fig. 4**A**) suggests that an increased expectation for an unique target shape (in the interleaved session)

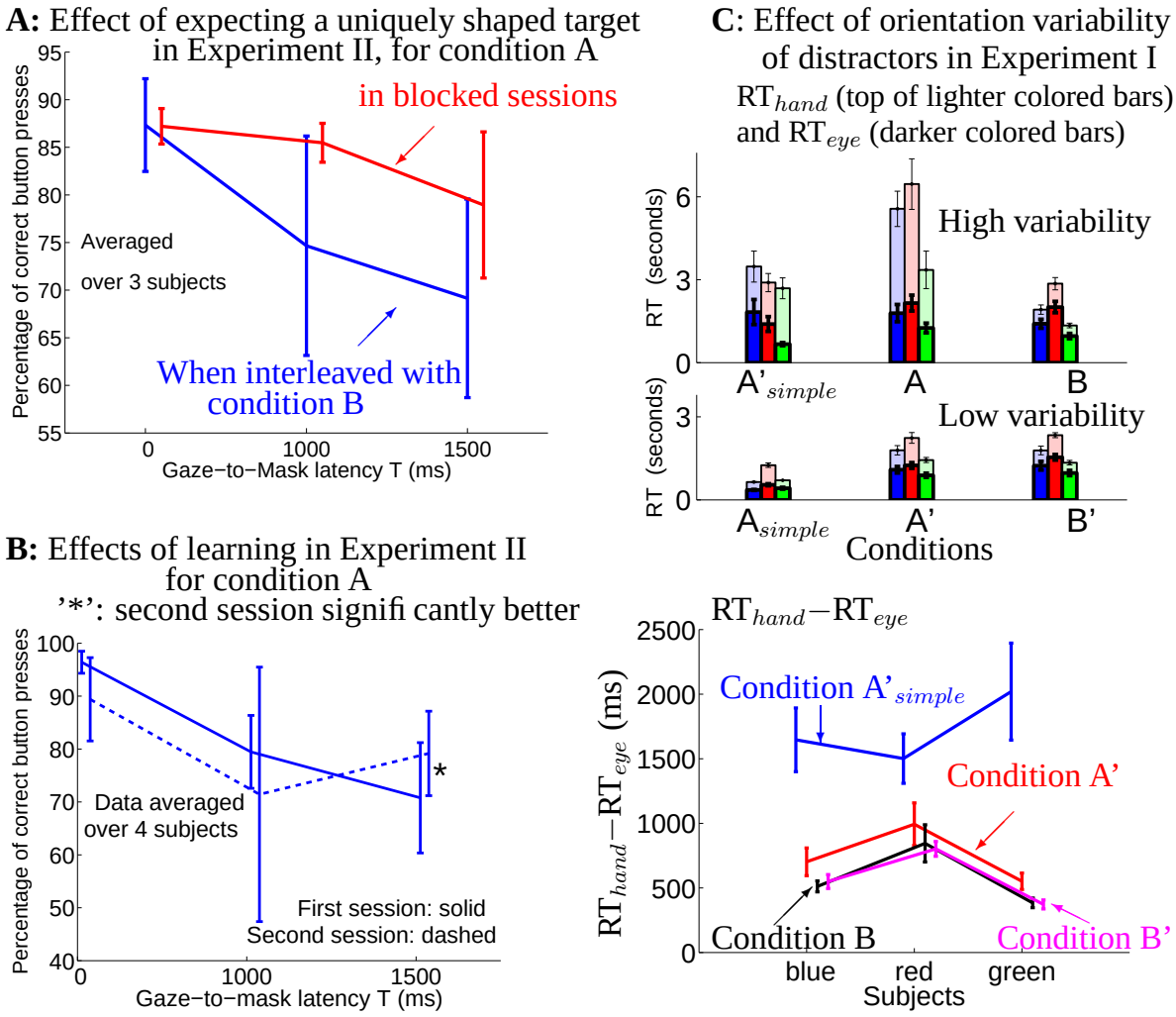


Figure 4: Factors affecting object-to-feature interference. **A:** Performance across $T > 0$ for condition A was somewhat better ($p = 0.08$) in blocked sessions (one session each subject) than in sessions interleaved with condition B (two sessions each subject). **B:** Reduction of interference with experience — for longer gaze-to-mask time T, performance for condition A improved in the second experimental session. The data in Fig. 3D were re-plotted here according to the two separate sessions. **C:** Stronger or weaker object-to-feature interference, manifested in $RT_{hand} - RT_{eye}$ in Experiment I, by, respectively, higher or lower orientation variabilities of the distractors to reduce or enhance bottom-up pop-out strength manifested in RT_{eye} (same three subjects as those in Fig. 2, denoted by blue, red, and green colors). The $RT_{hand} - RT_{eye}$ in condition A', while reduced from that of A, is significantly longer (p=0.002) than that of B'. In lower **C**, data points of different conditions are plotted in different colors. Stimulus examples of conditions A', B', and A'$_{simple}$ are shown in Supplementary data. Error bars show s.e.m.

increased interference. This is unsurprising since the expectation should increase the weight of factor (2) above. Analogously, interference can be reduced by increasing the weight of the bottom up factor, thus decreasing the relative weight of factor (2). For instance, when the task irrelevant bars in conditions A and B are all horizontal or all vertical to make distractors uniformly oriented, the target becomes more salient. We call these modified conditions A' and B' respectively. This reduces RT_{eye} significantly.

Consequently, the feature level influences could push the task decision process to reach decision threshold before object-to-feature interference becomes more significant. Hence, in Experiment I interleaving conditions A', B', A, and B, $RT_{hand} - RT_{eye}$ for A' is much shorter than for A, though still significantly longer than the two comparable $RT_{hand} - RT_{eye}$'s for B and B' (Fig. 4**C**). Conversely, when the orientation variability of distractors is increased in condition A_{simple}, such that randomly 1/3 of the distractor bars become oriented horizontally and another 1/3 vertically, making modified condition A'_{simple}, object-to-feature interference arises by a $RT_{hand} - RT_{eye}$ longer than that in condition B (Fig. 4**C**). This suggests that even a simple bottom up orientation feature can, given sufficient processing time, be treated as a viewpoint invariant object bar, making the target object bar a rotated version of all distractor objects.

Our data also suggest that one can quickly learn to remove the interference in condition A within two data sessions involving no more than 260 trials per subjects in Experiment II (Fig. 4**B**). Subjects reported discovering helpful strategies of trusting their instincts, or defocusing the image, or letting the target pop out while fixating on the center of display away from the peripheral target. Peripheral visual field is more heavily sampled by the magno celluar pathway, which, compared to the parvo cellular pathway, is faster and processes coarser resolution inputs[35, 36]. Hence, the magno pathway likely plays a greater role in detecting unique features and driving gaze in a bottom up manner. This is consistent with the idea that slower attentive process is associated with finer spatial resolution than the faster bottom up processes. Defocusing and peripheral viewing likely reduce the object-to-feature inteference by selectively emphasizing the magno pathway to speed up the bottom up process while removing the finer input details to attenuate the attentive object formation processes. Although removing finer resolution could make two intersecting bars resemble a single bar of the averaged orientation, the observed object-to-feature interference in condition A'$_{simple}$ (which has only disconnected bar stimuli) suggest that viewing the objects as single bars could not remove the interference if attentive object formation proceeded. Hence we predict that lesions (clinical or by transcranial magnetic stimulation) of the cortical areas responsible for attentive object processes (perhaps the parietal cortex which has been implicated in building objects from features[5]) could improve performance in our task. Our findings only reveal a fraction of the rich interactions between lower and higher level cognitive processes. The results of such interactions are unexpected if one assumes that deepening of processes should always lead to improved perception.

Different degrees of object-to-feature interference may underlie common observations of visual search asymmetry between familiar and unfamiliar targets. For example, searching for a familiar letter 'N' among its mirror reversals is slower than searching for a mirror reversal among normal N's[6, 37, 38]. Both searches require the same low level processes to detect orientation contrast between left and right tilted bars, and do not require letter recognition. However, familiarity of the letters should affect the object rather than feature level processing. Hence, the object-to-feature interference, manifested in our task and likely behind the portrait drawing trick, can enable additional examination of many known phenomena.

3 Experimental Procedures:

Stimuli: Each stimulus display, viewed at a distancce of 40 centimeters, had 660 = 22x30 object items, each at a position randomly displaced, up to $\pm 0.24^o$ visual angle, horizontally and vertically, from its corresponding position in a regular grid of 22 rows $\times$ 30 columns, spanning correspondingly $34^o \times 46^o$ in visual angle. Each stimulus bar was $0.12^o \times 1.1^o$ in visual angle, and 48 cd/m^2 in brightness. The background was black. The target's grid location was randomly one of those closest to the circle of about 15^o eccentricity, and beyond 12^o of horizontal eccentricity, from the display center. The fixation stimulus was a bright disk of 0.3^o diameter at the display center.

Procedures: Gazes were tracked by the 50 Hz infra-red video eye tracker from Cambridge Research System (www.crsltd.com). Tracking calibration was performed before each data session to a precision typically within 0.5^o of visual angle. After being shown two examples of each stimulus condition, untrained subjects were instructed to fixate centrally until stimulus onset and to freely move their eyes afterwards for target searching. The sequence of events in a trial was as follows. (1) With the fixation stimulus, the subject pressed a button to start a trial and eye tracking. (2) After 0.6 second, upon the subject's continuous fixation for 40 ms within 3^o of the fixation point, a blank screen replaced the fixation stimulus for 200 ms, followed by the onset (designated as time zero) of search stimulus. (3) In Experiment I, the search stimulus remained till after the subject's button press. In Experiment II, a mask replaced the search stimulus at a time determined as follows. In a gaze contingent trial, the mask onset occured at time T after the first gaze arrival at the target. The criteria for the arrival was when the gaze was within 2.3^o in visual angle from the target's center position. T was randomly chosen from the set T = (0, 100, 500, 1000, 2000) ms for data sessions contributing to

Fig. 3**C**, and for a different group of subjects, from the set T = (0, 1500) ms or T = (0, 1000, 1500) ms for sessions contributing to Fig. 3**D** and Fig. 4. For each non-gaze-contingent trial, a time τ was chosen randomly and uniformly from the time window 200-1700 ms. The mask onset occured upon the first gaze arrival at the opposite (laterally from the center) side of the target since 200 ms after stimulus onset, or at time τ, whichever was sooner. The mask, once displayed, remained till after the subject's button press. Each session of Experiment I had 200 trials, randomly interleaving conditions A_{simple}, B_{simple}, A, B, A', B', A'_{simple} and other control conditions. In Experiment II, each blocked session for condition A had 130 or 60 trials, each interleaving session (of conditions A and B) had 100 trials. After each session of Experiment II, we verified that subjects did not notice any links between the mask onsets and the gaze positions. Different subjects participated in Experiments I and II.

Data analysis: A trial is defined as a bad trial and removed from further analysis if gaze was untracked in more than 10% of the video frames of the eye tracker within the time window $(0, RT_{hand})$, or if $RT_{hand} < 100ms$. Data from a subject or session when bad trials comprised more than 10% of all trials are removed from further analysis. Sufficiently large gaze tracking error can lead to failures to detect gaze arrivals at the target. A trial is called a non-arrival trial if the gaze never arrived at the target by our arrival criteria using the tracker measurements. We thus remove from further analysis subjects and data sessions having more than 11% of non-arrival trials among all trials in Experiment I, or among the gaze contingent trials in Experiment II. Results in figures were based on the gaze arrival trials only. The RTs plotted were based on trials with correct button responses. The error bars plotted are the standard errors of the means (s.e.m). Statistical tests for differences between different conditions in Fig. 2 were by two-tailed t-test, while those in Fig. 3 and 4 were by one tail matched sample t-test.

Acknowledgement: Work supported by the Gatsby Charitable Foundation. We thank Keith May for help in programming the stimulus, and he, Peter Dayan, Chris Frith, Uta Frith, Sheng He, Li Jingling, and Alex Lewis for conversations and comments on our works, manuscripts, and references. Comments by the three anonymous reviewers are also much appreciated.

References

[1] Treisman A. M., Gelade G. (1980) A feature-integration theory of attention. *Cognit Psychol.* 12(1), 97-136.

[2] Li Z. (2002) A saliency map in primary visual cortex. *Trends Cogn. Sci.* 6(1):9-16.

[3] Tehovnik EJ, Slocum WM, Schiller PH (2003). Saccadic eye movements evoked by microstimulation of striate cortex. *Eur J. Neurosci.* 17(4):870-8.

[4] 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.

[5] Treisman AM, Kanwisher NG. (1998) Perceiving visually presented objects: recognition, awareness, and modularity. *Curr Opin Neurobiol* 8(2):218-26

[6] Frith U. (1974) A curious effect with reversed letters explained by a theory of schema *Perception & Psychophysics* 16(1):113-116.

[7] Hubel DH Wiesel TN (1968) Receptive fields and functional architecture of monkey striate cortex. *J Physiol.* 195(1):215-43.

[8] Kahneman D, Treisman A, Gibbs BJ. (1992) The reviewing of object files: object-specific integration of information *Cognit. Psychol.* 24(2):175-219.

[9] Riesenhuber M. and Poggio T. (2003) How the visual cortex recognizes objects: the tales of the standard model. *The visual neurosciences*, Chalupa LM and Werner JS. (Eds) vol 2. page 1640-1653.

[10] Tanaka K. (2003) Inferotemporal response properties. *The visual neurosciences*, Chalupa LM and Werner JS. (Eds) vol 2. page 1151-1164.

[11] Rolls E.T. (2003) Invariant object and face recognition *The visual neurosciences*, Chalupa LM and Werner JS. (Eds) vol 2. page 1165-1178.

[12] Logothetis NK, Pauls J, Poggio T. (1995) Shape representation in the inferior temporal cortex of monkeys. *Current Biology* 5(5):552-63

[13] Humphreys GW, Riddoch MJ, Price CJ. (1997) Top-down processes in object identification: evidence from experimental psychology, neuropsychology and functional anatomy *Philosophical Transactions of the Royal Society B: Biological Sciences* 352(1358):1275-82.

[14] Kourtzi Z, Kanwisher N. (2001) Representation of perceived object shape by the human lateral occipital complex. *Science* 293(5534):1506-9

[15] Grill-Spector K, Kourtzi Z, Kanwisher N. (2001) The lateral occipital complex and its role in object recognition. *Vision Res.* 41(10-11):1409-22.

[16] Knierim JJ, Van Essen DC. (1992) Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *J. Neurophysiol.* 67(4): 961-80.

[17] Sillito AM, Grieve KL, Jones HE, Cudeiro J, Davis J. (1995) Visual cortical mechanisms detecting focal orientation discontinuities. *Nature* 378(6556):492-6.

[18] Nothdurft HC, Gallant JL, Van Essen DC. (1999) Response modulation by texture surround in primate area V1: correlates of "popout" under anesthesia. *Vis. Neurosci.* 16, 15-34.

[19] Schiller PH and Lee K. (1991) The role of the primate extra-striate area V4 in lesion. *Science* 251(4998):1251-1253.

[20] Mazer JA, Gallant JL. (2003) Goal-related activity in V4 during free viewing visual search. Evidence for a ventral stream visual salience map. *Neuron* 40(6):1241-50.

[21] Ogawa T and Komatsu H. (2004) Target selection in area V4 during a multidimensional visual search task *J. Neurosci.* 24(28): 6371-6382.

[22] Wolfe JM, Bennett SC. (1997) Preattentive object files: shapeless bundles of basic features. *Vision Res.* 37(1):25-43.

[23] Ungerleider, L. G. and Mishkin, M. (1982) Two cortical visual systems. In D. J. Ingle, M. A. Goodale, and R. J. W. Mansfield (Eds.), *Analysis of Visual Behavior* The MIT Press: Cambridge, Mass. 1982, pp. 549-586.

[24] Chelazzi L, Miller EK, Duncan J, Desimone R. (1993) A neural basis for visual search in inferior temporal cortex. *Nature* 363(6427):345-7.

[25] Motter BC. (1993) Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *J Neurophysiol.* 70(3):909-19

[26] Hoffman J. E. (1998) Visual attention and eye movements. *Attention* Ed. H. Pashler, Psychology Press. Page:119-154.

[27] Duncan J, Humphreys GW. (1989) Visual search and Stimulus similarity. *Psychol. Rev.* 96(3):433-58.

[28] Potter M. C. (1976) Short-term conceptual memory for pictures. *J. Exp. Psychol. Hum. Learning Memory* 5:509-522.

[29] Thorpe S, Fize D., Marlot C. (1996) Speed of processing in the human visual system. *Nature* **381**:520-522.

[30] Luck SJ, Vogel EK, Shapiro KL. (1996) Word meanings can be accessed but not reported during the attentional blink. *Nature* **382**:616-618.

[31] van Zoest W, Donk M. (2004) Bottom-up and top-down control in visual search *Perception* 33(8):927-37

[32] Gilchrist ID, Heywood CA, Findlay JM. (2003) Visual sensitivity in search tasks depends on the response requirement. *Spat Vis.* 16(3-4):277-93

[33] Fang F, He S. (2005) *Nat. Neurosci.* 8(10):1380-5.

[34] Enns JT and Di Lollo V. (2000) What's new in visual masking? *Trends in Cognitive Sciences* **4**(9):345-352.

[35] Frazor RA, Albrecht DG, Geisler WS, Crane AM. (2004) Visual cortex neurons of monkeys and cats: temporal dynamics of the spatial frequency response function. *J. Neurophysiol.* 91(6):2607-27.

[36] Li Z. (1992) Different retinal ganglion cells have different functional goals. *International Journal of Neural Systems* 3(3):237-248.

[37] Richards, J T; Reicher, G M (1978) The effect of background familiarity in visual search - An analysis of underlying factors *Perception and Psychophysics* 23(6)p. 499-505.

[38] Shen J.; Reingold E. M. (2001) Visual search asymmetry: The influence of stimulus familiarity and low-level features *Perception & Psychophysics*, 63(3) p. 464-475(12)

4 Supplementary data

More details on experimental design and procedure

Subjects were all naive, did not practice for the task before data taking, and were only shown two stimulus examples for each relevant stimulus condition to know the task before the actual data taking. Experiment I took two sessions, of 200 trials each, on each subject, interleaving conditions A, B, A$_{simple}$, B$_{simple}$, A', B', A'$_{simple}$, and other control conditions. In all conditions, the target is the only one that had a left (or right) tilted bar in the whole display regardless of the degree of tilt from vertical (i.e., each distractor could either contain a bar tilted to the other direction from vertical, or have no left or right tilted bars). This gives about 44 trials for each subject on each condition. Experiment II had two designs. The first design did a finer resolution scan on the Gaze-to-Mask latency $T = (0, 100, 500, 1000, 2000)$ ms on condition A only for two sessions (of 130 trials each) on each subject, each gaze contingent trial was randomly assigned to one of the T values with equal probability. This gave about 26 trials per subject for each T, contributing to results in Fig. 3C. The second design did a coarser resolution scan of the Gaze-to-Mask latency $T = (0, 1000, 1500)$ ms or $T = (0, 1500)$ ms for two sessions (100 trials each) interleaving conditions A and B and one session (of 60 trials) of condition A only on each subject. Each gaze contingent trial was randomly assigned to each T either with the same probability or $T = 0$ was twice as likely as another $T > 0$. This gave about 23/13/15 trials per subject per condition for $T = 0/1000/1500$ ms from the sessions interleaving conditions A and B, and about 17/7/9 trials per subject for $T = 0/1000/1500$ ms in the blocked sessions. The results contributed to Fig. 3D and Fig. 4AB. Each subject participated only in Experiment I, or only in the first or second design of Experiment II. No subject participated in more than 3 sessions of experiments or did a total of more than 400/260 trials in Experiment I/II, since we observed (by using ourselves as subjects in pilot experiments) that experience with the task reduced or eliminated the object-to-feature interference. In Experiment II, to verify that the subjects did not notice any link between their gaze positions (relative to the target) and the mask onset time, we asked them at the end of the sessions whether they had any comments and observations. None of their answers mentioned this link, most comments about the target positions were either irrelevant (e.g., "targets from two successive trials often appear on the same sides") or mentioned that targets seemed equal distance from the display center. Roughly half of the subjects could not have their eye positions calibrated sufficiently well to proceed to data taking.

More details on data analysis

By examining RT_{eye} to other possible target positions not having the target, we verified that the subjects did not locate the target by *randomly* scanning only the possible target positions. Since head and body movements, and excessive blinking, often worsen the eye tracking quality during a data taking session, we carried out data analysis to identify poorly tracked trials as defined in Experimental Procedures. These trials are removed from further data analysis, and subjects or data sessions with too many (as defined in Experimental Procedures) of these trials are also removed from further data analysis. As a result of these analysis, and of the experimental design to assign the trial conditions in each session, the number of trials included for analysis can vary somewhat from subject to subject and from condition to condition.

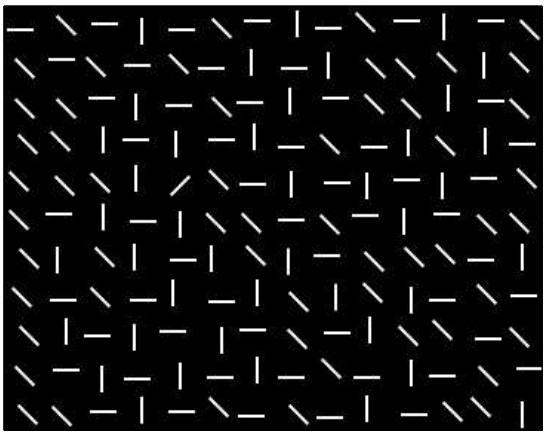
The RT results (for each subject) presented are the means (and the standard errors of the means) of RTs from gaze arrival trials in each subject. The percentage values in Figs. 3 and 4 are the means (and s.e.m's) of those from individual subjects. To obtain for an individual subject the percentage values and their errors in Fig. 2C, we calculate the mean and variances in the Beta distribution

(see http://mathworld.wolfram.com/BetaDistribution.html) $P(p) = (N+1)!p^n(1-p)^{N-n}/n!/(N-n)!$ which gives the posterior probability that the true fraction of certain trial type (e.g., correct button press) is p when n such trials have been observed in a total of N trials.

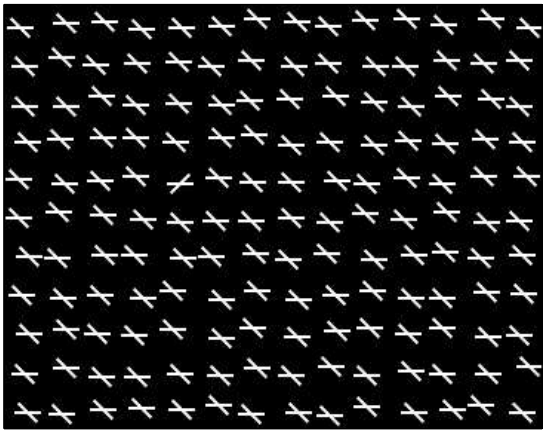
Illustrating the additional stimulus conditions in Fig. 4C

Fig. (5) illustrates stimulus conditions in Fig. 4C that are not shown in Fig. 1. These additional conditions demonstrate that increasing background variability, as when modifying conditions A$_{simple}$, A', and B' to conditions A'$_{simple}$, A, and B, respectively, can decrease the bottom up component in task decision, prolonging RT_{eye}. Consequently, object-to-feature interference increases in conditions A and A'$_{simple}$ than that in conditions A' and A$_{simple}$ respectively, as manifested in prolonged $RT_{hand} - RT_{eye}$. However, as condition B is changed to B', $RT_{hand} - RT_{eye}$ remain unchanged even though RT_{eye} shortened, since object-to-feature interference remains absent when the target is uniquely shaped.

Condition A'$_{simple}$, modified from condition A$_{simple}$
by increasing distractor orientation variability.
Note that target bar is still uniquely oriented



Condition A', modified from condition A
by orienting distractors uniformly



Condition B', modified from condition B
by orienting distractors uniformly

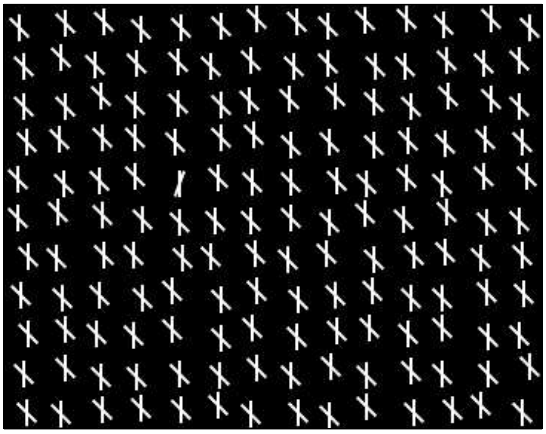


Figure 5: Small portions of examples of visual search displays for conditions A', B', and A'$_{simple}$ whose results are shown in Fig. 4C

.