Errors in visual search: Are they stochastic or deterministic?

Aoqi Li[1], Johan Hulleman[1], and Jeremy M Wolfe[2,3]

1. University of Manchester, Manchester, UK
2. Brigham and Women's Hospital, Boston, MA
3. Harvard Medical School, Boston, MA

Abstract

In any visual search experiment, observers will make errors. Those errors can be categorized as "deterministic": If you miss this target in this display once, you will definitely miss it again. Alternatively, errors can be "stochastic", occurring randomly with some probability from trial to trial. To empirically categorize errors in a simple search task, our observers searched for the letter "T" among "L" distractors, with each display presented twice. When the letters were clearly visible (white letters on a gray background), the errors were almost completely stochastic (Exp 1). An error made on the first appearance of a display did not predict that an error would be made on the second appearance. When the visibility of the letters was manipulated (letters of different gray levels on a noisy background), the errors became a mix of stochastic and deterministic. Lower contrast targets produced more deterministic errors. (Exp 2). Using the stimuli of Exp 2, we tested whether errors could be reduced by a "mindless AI" intervention that guided attention around the display but knew nothing about the content of that display (Exp3a,b). This had no effect, but a slightly less mindless AI that knew the location of all items, did succeed in reducing deterministic errors (Exp3c). (205 Words)

1. Introduction

Individuals routinely fail to report or respond to visual stimuli that are clearly visible, "right in front of their eyes". In some cases, the missed item is unexpected. The Simons and Chabris (1999) gorilla is the most famous example of such "inattentional blindness" (Koivisto et al., 2004; Kuhn & Tatler, 2011; Mack & Rock, 1998; Macknik et al., 2008; Simons, 2000; Simons & Chabris, 1999). Inattentional blindness has been invoked as an explanation for some real-world errors; for example, how a driver may fail to notice an unexpected road user before a road accident or, more benignly, how an audience member may be induced to believe that something has materialized from nothing in a magic show. Various factors might contribute to these effects. For instance, some researchers have proposed that a failure to see some highly noticeable objects is due to an illusion that the space behind an occluding foreground object is experienced as empty ("the illusion of absence", Ekroll et al., 2021). As a result of this illusion, observers might have a misleading sense of security without checking whether the blind spot is really empty, or even be mistakenly convinced that a fixated spot is empty without taking a second look. Ekroll et al (2021) proposed this illusion as an important contributor to 'looked-but-failed-to-see' (LBFTS) errors in driving situations.

Missed gorillas and other examples of inattentional blindness are dramatic but they are far from the only type of LBFTS error. Clearly visible targets are routinely missed even when the searcher knows that such targets are part of their ongoing task. In typical LBFTS driving accidents, the driver will generally know that they should be watching for pedestrians, turning vehicles, etc. In medical settings, when a clinician fails to report an "incidental finding", it will not be a missed gorilla (Drew et al., 2013). It is more likely to be a secondary, but clinically significant finding that the clinician knows might occur in this setting (Lumbreras et al., 2010). Indeed, a missed item can be the actual target of a search (Hovda et al., 2022). The search for typos in a manuscript would be a relevant example for many readers of this paper of a clearly visible but not reported stimulus. In breast radiology, perhaps 70% of missed lesions on mammograms are visible enough to attract radiologists' visual attention but a plethora of different factors, including satisfaction of search, incorrect background sampling, and incorrect first impressions result in diagnostic errors (Gandomkar & Mello-Thoms, 2019).

Some efforts have been devoted to look at factors that influence the occurrence of miss errors within clinical settings. For example, Wolfe et al. (2017) developed the "mixed hybrid search" paradigm as a laboratory task with which to study why radiologists often miss clinically significant "incidental findings" like missing signs of lung cancer when they search for pneumonia. In a "mixed hybrid search" task, half of the targets are specific images (this butterfly in this pose, this shoe, etc) while the other half are categorically defined (any mammal, any piece of furniture). Their results showed that categorically defined targets were more likely to be missed than specific targets, analogous to what happens in clinical settings where the more specific object of search might be found while the more loosely-define incidental findings might be missed.

In addition to the specificity of the target, its prevalence also strongly influences whether a target will be correctly located or identified. A standard search experiment might have a target on 50% of trials (50% “prevalence”) while a task like screening mammography might have findings worth following up on 5-10% of cases, with actual cancer on just 3-4 per thousand cases. With decreased target prevalence, participants tend to quit the search earlier and make more conservative decisions about target presence (Hout et al., 2015; Peltier & Becker, 2016). Both faster quitting and more conservative decision criteria contribute to more miss errors. This is known as the low prevalence effect (LPE).

The presence of multiple targets introduces another path to error. Observers are more likely to miss a second target after a first one has been detected. These are so-called "satisfaction of search" or "subsequent search miss" errors (Adamo et al., 2013; Berbaum et al., 1990, 1991). Adamo et al. (2019) offer three accounts for the underlying mechanism. The “satisfaction” account suggests that additional targets are missed because observers are satisfied with the first found target and terminate the search early. The “perceptual set” account proposes that observers are more likely to miss dissimilar second targets because, having found one target, observers tend to look for targets similar to that first one. Finally, the “resource depletion account” suggests that after resources are allocated to the detection of the first target, reduced resources make it more likely that a subsequent target will be missed.

Even in a very basic laboratory visual search task like a search for a perfectly visible “T” among other distractor letters, observers will routinely miss 5% - 10% of targets. While the consequence of missing a “T” in a laboratory search are trivial, there are obvious and potentially serious consequences of miss errors in real-life contexts such as mammographic screening or road safety (Hovda et al., 2023; Pamme et al., 2018). Therefore, there is a need to investigate miss errors and to test interventions that might reduce their frequency (Wolfe et al, 2022).

In this paper, we are interested in the nature of miss errors in the simple case in which observers look for a letter “T” among “L”s with no uncertainty about the identities of the target or the distractors (no gorillas here). Still, as noted, targets are missed. Are those errors random (henceforth “stochastic”)? That is, if participants miss, let us say, 10% of targets, is that a random set of 10% of all target trials or are observers more likely to miss some specific targets in some specific displays? In the limit, would participants miss the same targets again, if asked to search the same displays? We will call such errors “deterministic”. To find the answer to this question, a set of T among L search displays was presented twice to each participant. We calculated the miss rate, $P1$, for the first time that the set of displays was shown and, $P2$, for the second time. We also calculated the proportion cases where both copies were missed: $P12$. If the errors are stochastic, then $P12 = P1 * P2$. If the errors are deterministic, $P12 = \min(P1, P2)$. If errors are a mix of stochastic and deterministic, $P12$ will fall between these two predictions. In addition to the analysis on the qualitative nature of these errors, it is possible to calculate the relative proportions of stochastic and deterministic errors, based on the three observable quantities: $P1$, $P2$ and $P12$. This calculation allowed us to evaluate the effect of "mindless" AI interventions. If an intervention was useful, did it reduce stochastic or deterministic errors? If these interventions reduce errors on a simple T-vs-L search task, it might be worth trying a similar strategy in socially important, real-life tasks.

2. Experiment 1: Basic search for a T among Ls

Experiment 1 consisted of a simple visual search task where white letters were presented against a gray background.

2.1 Participants

The experiment was run online on the Pavlovia platform (https://pavlovia.org). For Experiment 1, we tested 20 participants (6 males, 14 females, mean = 19.5, SD = 0.9, min = 18, max = 21) from the BSc Psychology programme at the University of Manchester. All participants reported normal or corrected-to-normal vision and gave their informed consent before they began the experiment. Participants received course credit for their participation. Ethics approval came from the University of Manchester (2023-16117-27175)

2.2 Stimuli & apparatus

The experiment was programmed in Python and translated into javascript by PsychoPy (Peirce et al., 2019). The online version was hosted on Pavlovia. Figure 1 shows the stimuli for Experiment 1. They consisted of an array of white letters (T and Ls) against a gray background. The length of vertical and horizontal line segments of the Ts and Ls was 0.03 screen height (note that because we were testing on-line, we had relative, not absolute control of the sizes of stimuli). The orientations of the letters were randomly and uniformly selected from rotations of 30, 60, 90, 120, 150, 180, 210, 240, 270, 300, 330, & 360 deg. The positions of the letters were randomly generated for each trial such that all items fit in a square region that had a side length of 0.7 screen height, centered on the middle of the screen. In addition, the minimum distance between any two letters was always larger than 0.1 screen height.

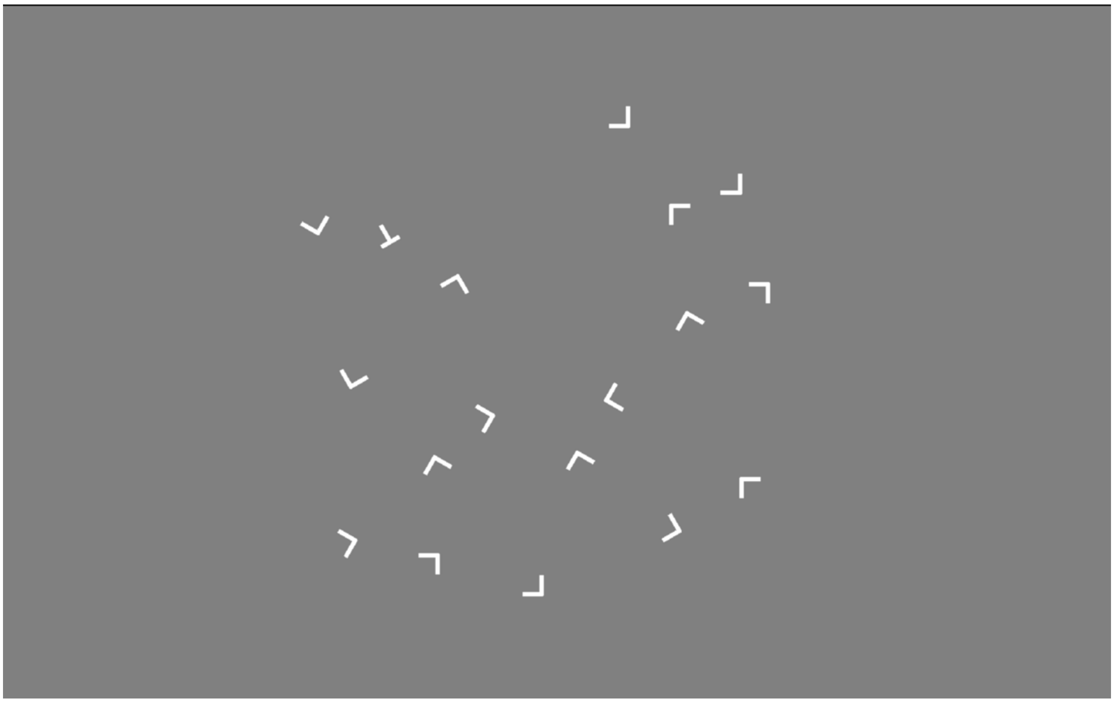


Figure 1: Sample stimulus for Experiment 1.

2.3 Design & procedure

Participants searched for the letter T among Ls. Participants were instructed to press ‘j’ if they found the target and ‘f’ if they did not. The stimulus was present until response. Targets were present on 50% of trials. Trial by trial feedback was not given, but after every block of 100 trials the proportion correct for that block was displayed. There were two set sizes; 18 and 36, fully crossed with target presence and target absence. For each participant, we generated 75 versions of each of the four resulting combinations for a total of 300 unique stimulus displays. Each of these was presented twice. The two copies of the 300 stimuli were randomly intermixed across six blocks of 100 trials for a total of 600 trials for the experiment. Thus, there were three factors in this design, each with two levels: repetition (first, second), set size (18, 36), and target (present, absent). Participants completed four practice trials before they started the experiment.

2.4 Analysis method

We focused on the RT data and miss rate data, with our primary interest being in the error data. The RT data was subjected to a three-way repeated measure ANOVA with target presence, set size and repetition as within-subject factors. Since the experiments involved a typical visual search task, we found the typical main effects of target presence and set size. Specifically, there were longer reaction times for absent trials and longer reaction times for larger set sizes. A two-way interaction between target and set size, showing steeper reaction time slopes for absent trials, also occurred. All of these effects are highly statistically reliable and will not be reported in detail in the Results section. The results of the full ANOVAs are shown in supplementary tables.

For miss rate data, we calculated the miss rate, $P1$, for the first time the set of displays was shown and, $P2$, for the second time. We also calculated the proportion of cases where both copies were missed, $P12$. If the errors are stochastic, then $P12 = P1 * P2$. If the errors are deterministic, $P12 = \min(P1, P2)$. If the errors are a mix of stochastic and deterministic errors, $P1 * P2 < P12 < \min(P1,\ P2)$. To get a quantitative estimate of the relative proportion of stochastic and deterministic errors, we modelled how the errors observed in round 1 and round 2 could be decomposed into different types of error, as shown in Figure 2. One complication worth noting is that a deterministic display may produce a stochastic error. A deterministic display is one that would produce a deterministic error. However, it is possible for an error to be produced on that trial for stochastic reasons. Imagine, for instance, that the observer is simply not paying attention on that would-be deterministic trial and pushes a response button at random.

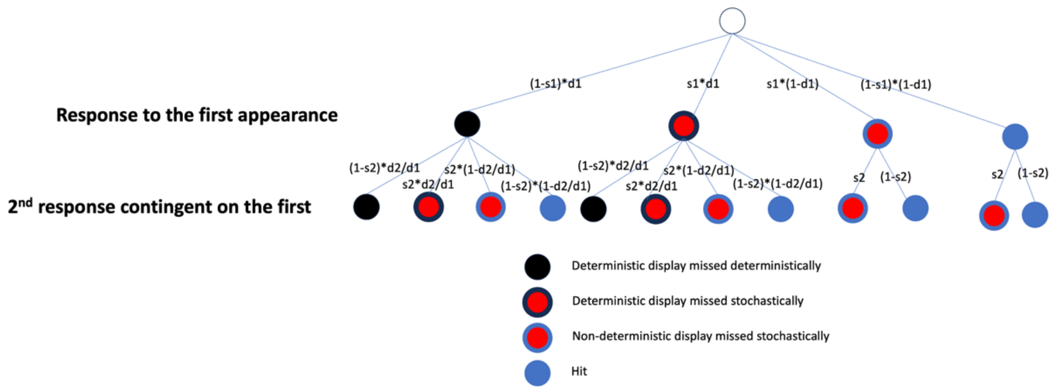


Figure 2: Observed errors decomposed into deterministic errors and stochastic errors.

In Figure 2, there are four possible states for a trial. The target in a trial is either fundamentally unfindable (black) or fundamentally findable (blue). A completely black circle represents the case where a deterministic error is made on a trial with an unfindable target. A black circle with a red centre means that the unfindable target was missed stochastically. A blue circle with the red centre represents the situation where a stochastic error is made on on a trial with a findable target. A blue circle represents a trial where the target is successfully found. If the target in a trial is fundamentally findable, this target cannot become fundamentally unfindable. This means that it is not possible to transition from a blue circle or blue circle with a red centre to a black circle or a black circle with a red centre. A transition in the opposite direction is possible though. For instance, if an AI manipulation works and reduces the number of deterministic errors, it is possible for a deterministic miss (black) on one trial to become a hit (blue) on its next appearance.

To describe the proportions of deterministic and stochastic errors, four parameters are introduced: $d1$ and d2 represent the proportion of deterministic errors relative to the total number of stimuli in round 1 and round 2. $s1$ and $s2$ represent the stochastic error rates for a stimulus in round 1 and round 2 respectively. In Figure 2, *Row 0* with an empty circle represents the to-be-determined status of one trial. *Row 1* with four different types of circles represents the four possible outcomes of the first appearance of a trial with the notation for the corresponding probabilities in round 1. *Row 2* represents the possible outcomes with the notation for the corresponding probabilities in round 2. Therefore, the observed $P1, P2$ and $P12$ can be theoretically decomposed into the summed error probabilities in round 1 and round 2. The following three equations can be derived (The original version and the simplifying process can be found in the appendix):

$$P1 = d1 * (1 - s1) + s1$$

$$P2 = d2 + s2 * (d1 - d2) + (1 - d1) * s2$$

$$P12 = d2 + s2 * (d1 - d2) + s1 * (1 - d1) * s2$$

For Experiment 1, there was no AI intervention. A fixed deterministic rate was therefore assumed for round 1 and round 2, i.e., $d = d1 = d2$. With this additional assumption, there is a unique solution for the above equations.

$$d = \frac{P12 - P1 * P2}{1 - P1 - P2 + P12}$$
$$s1 = \frac{P1 - P12}{1 - P2}$$
$$s2 = \frac{P2 - P12}{1 - P1}$$

2.5 Data exclusion

Trials with RTs smaller or greater than 2.5 SD from the mean RT in each cell of the combination target x set size (3.47%) and trials where participants corrected their motor responses (1.07%) were removed for each observer. When one trial was removed, the other copy of the trial was also be removed (93.3% remained). After the removal of the above trials, we further checked the d' of all the participants. Participants with d' beyond 2.5 SD from the group mean for each individual experiment were excluded. One participant with a low d' = 1.02 was removed from Exp 1. For the remaining participants, min d' = 2.95, max d' = 5.84.

2.6 Results

2.6.1 RTs

Figure 3 shows RTs on correct response trials for Experiment 1. It is clear that the first and second repetitions of the stimuli produce very similar RTs with a slight speed-up on the second appearance. The three-way repeated measure ANOVA with target presence, set size and repetition as within-subject factors shows a main effect of repetition [F(1, 18) = 7.00, p = 0.016, η_p^2 = 0.28], suggesting that participants responded faster in round 2 than in round 1. The interaction between target presence and repetition [F(1, 18) = 0.37, p = 0.55, η_p^2 = 0.02] as well as the interaction between set size and repetition [F(1, 18) = 0.50, p = 0.49, η_p^2 = 0.03] was not significant. The three-way interaction among all the factors was not significant either [F(1, 18) = 0.004, p = 0.95, η_p^2 = 0.00]. The full results of the three-way ANOVA are presented in Figure S1 in the appendix.

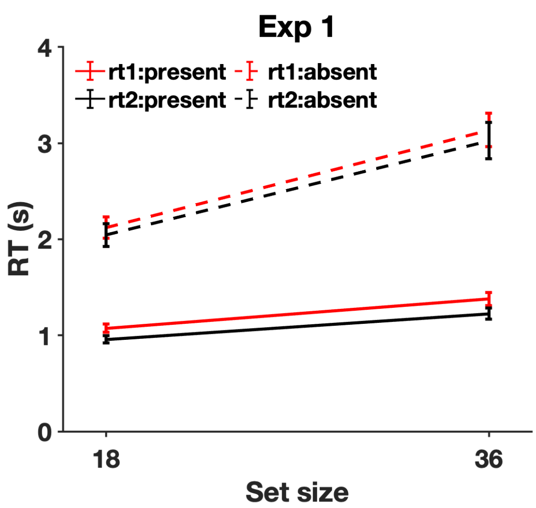


Figure 3: RTs from Experiment 1 as a function of set size, target presence, and repetition. Red lines: First presentation, Black lines: Second presentation. Full

lines: present trials, Dotted lines: absent trials.

2.6.2 Miss rates

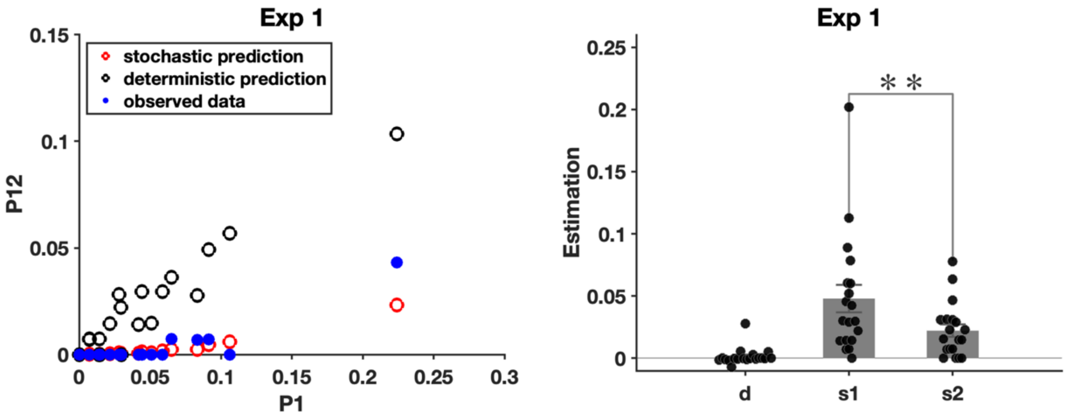


Figure 4: Miss rate analyses for Experiment 1. Left panel: comparison between observed human data, stochastic predictions and deterministic predictions for each observer. Right panel: deterministic error proportion d and stochastic error rates $s1$ and $s2$ calculated from human data.

Figure 4 shows the results of the miss rate analyses for Experiment 1. In the left scatter plot, the blue dots represent the observed data for each participant with $x = P1$ and $y = P12$ calculated from human data. Each observed data point (blue dot) is paired with the stochastic prediction (red circle) and the deterministic prediction (black circle) of $P12$ given the observed $P1$ and $P2$. Therefore, for one participant with observed $P1$, $P2$ and $P12$, each observed data point (blue dot) is $(P1, P12)$, the stochastic prediction (red circle) is $(P1,\ P1 * P2)$, and the deterministic prediction (black circle) is $(P1,\ \min(P1, P2))$. As can be seen, the observed data (blue dots) are almost overlapping with the stochastic predictions (red circles). The right bar plot shows the results of the parameters solved using the equations from the Methods section, above. The figure is based on the assumption that $d = d1 = d2$, resulting in three parameters to be computed i.e., the deterministic error proportion, d, the stochastic rate in round 1, $s1$, and the stochastic rate in round 2, $s2$. A one-sample t-test showed that the deterministic error proportion d was not significantly different from 0 [$t(18) = 0.91$, $p = 0.38$, Cohen's $d = 0.21$], demonstrating that errors in Experiment 1 were almost exclusively stochastic. A paired t-test comparing the stochastic rates, $s1$ and $s2$, shows that observers made fewer stochastic errors in round 2 than in round 1 [$t(18) = 3.81$, $p = 0.0013$, Cohen's $d = 0.87$], indicative of some learning effect over the course of Experiment 1.

2.6.3 Experiment 1 discussion

Experiment 1 consisted of a simple T-vs-L search task where all the white letters were presented on a gray background. Analyses of the RTs and miss rates showed that observers responded faster and made fewer errors in round 2 than in round 1, indicative of some learning effect. More importantly for present purposes, the proportion of deterministic errors, d, calculated from miss rates was not significantly different from 0, indicating that errors were almost purely stochastic in this experiment. The result would be different if there was a systematic bias in search. For example, if observers tended to ignore the lower left corner of the display, then targets in the lower left would be more likely to be missed on both their first and second appearances. This is not what is found with these simple and clear stimuli. However, in many real-world search tasks (mammography, airport security), search items are not so clearly visible. In the next two experiments we therefore tested whether stochastic errors still dominate when items become harder to distinguish from the background.

3. Experiments 2a and 2b: letters on a noisy background

In Experiments 2a and 2b, the uniform gray background was replaced by a noisy background. The letters were also of various grayscales. The only difference between the two experiments was that Experiment 2b used a more restricted set of target contrasts and target locations compared to Experiment 2a.

3.1 Participants

Both Experiments 2a and 2b were run online on the Pavlovia platform (https://pavlovia.org). For Experiment 2a, we tested 21 participants (6 males, 13 females, mean = 23.1, SD = 7.1, min = 18, max = 45, two participants did not provide the gender and age information). Thirteen of them were from the BSc Psychology programme at the University of Manchester and eight of them were recruited via Prolific. For Experiment 2b, we tested 21 participants (8 males, 8 females, 1 non-binary, mean = 28.8, SD = 10.9, min = 20, max = 57, four participants did not provide the gender and age information) recruited via Prolific. All participants reported normal or corrected-to-normal vision and gave their informed consent before they began the experiment. Participants received course credit (when recruited from the BSc

Psychology) or 8 GBP (when recruited via Prolific) for their participation. Ethics approval came from University of Manchester (2023-16117-27568 [credit version] and 2023-16117-28440 [payment version] for Exp 2a, 2023-16117-28499 for Exp 2b).

3.2 Stimuli & apparatus

In Experiments 2a and 2b, the stimuli consisted of an array of T and Ls against a background composed of $1/f^{1.3}$ noise. The noise was intended to roughly simulate the texture of a mammogram. Ts and Ls were of various grayscales. The length of vertical and horizontal lines of Ts and Ls was 0.03 screen height. The orientations of the letters were randomly selected from [30, 60, 90, 120, 150, 180, 210, 240, 270, 300, 330, 360]. The minimum distance between any two letters was always larger than 0.1 screen height to avoid overlapping.

In Experiment 2a, The grayscales for items were randomly generated by the formula (rand()-0.5)*2 for each trial. Therefore, in Experiment 2a, the distribution of item grayscale values was uniform. The default colour space in PsychoPy ranges from -1 to 1, so the recorded values were converted to 0-255 for subsequent analyses. The positions of the letters were randomly generated for each trial with the limitation that both x and y ranged from [0.15, 0.85] screen height. The noisy background was randomly selected from 10 noise images of 2000*2000 pixels for each trial and was cropped from the centre to fit the screen size during the online testing. Therefore, Experiment 2a required participants to use a screen smaller than 2000*2000.

In Experiment 2b, the target contrast (defined by the difference between target grayscale and background grayscale [T-B]) was controlled to be [-105, -75, -45, -15, 15, 45, 75, 105] and the locations of the target were evenly distributed across four spatial quadrants (upper left: x, y from [0.15, 0.45]; upper right: x from [0.55, 0.85], y from [0.15, 0.45]; bottom left: x from [0. 15, 0.45], y from [0.55, 0.85]; bottom right: x, y from [0.55, 0.85]). To achieve this manipulation, we generated the stimuli before the experiment. Crossing target presence (2), set size (2), T-B (8) and target location (4) yielded 128 stimuli (2*2*8*4). Two search arrays were generated for each parameter combination, resulting in a total of 256 stimuli. Since all stimuli were presented twice, the total number of trials was 512. For each search array, the noisy background was randomly selected from 10 noise images of 1000*1000 pixels. The final stimuli were resized to fully occupy the screen height during the online testing. Figure 5 shows an example of the stimuli used in Experiments 2a and 2b.

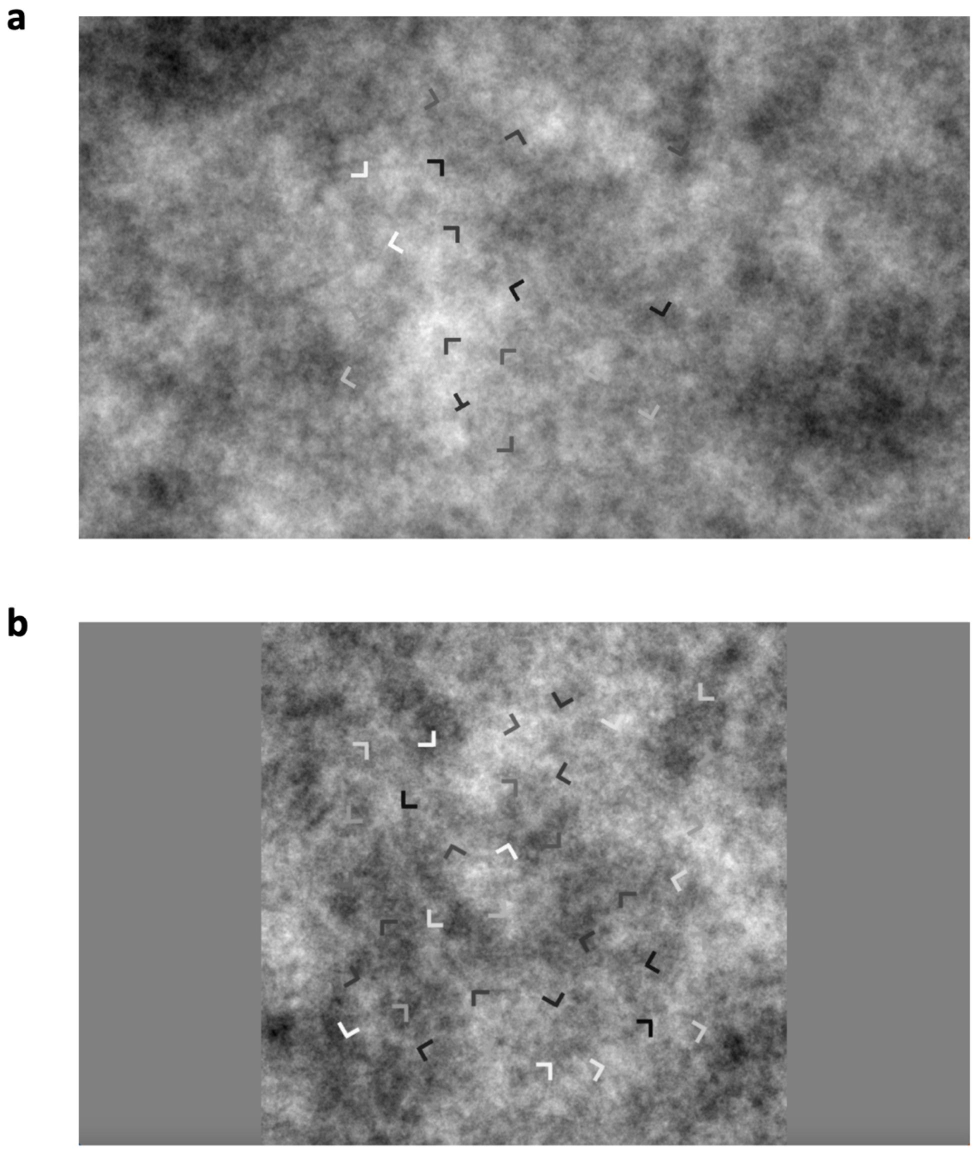


Figure 5: Sample stimuli from Experiments 2a and 2b. (a) In 2a, the background filled the entire screen. (b) In 2b, the background was a square with size determined by the vertical extent of the screen.

3.3 Design & procedure

Participants were instructed to press ‘j’ if they found the target T and ‘f’ if they did not. The stimulus was present until response. Trial by trial feedback was not given, but after each block, the percentage correct was displayed. Experiment 2a generated 300 stimuli online for each participant. The two copies of each of the the 300 stimuli were randomly intermixed across six blocks of 100 trials. Experiment 2b used pre-generated stimuli as described in the Stimuli & Apparatus section for all participants. Two copies of the 256 pre-generated stimuli were randomly intermixed across four blocks of 128 trials. As in Experiment 1, this design had three factors, each with two levels: repetition, set size, and target. Participants were required to finish a 12-trial practice session before the experiment and would only be able to begin the experiment with an accuracy higher than 0.75, otherwise, they had to repeat the practice.

3.4 Analysis method

The analysis of the RT data and miss rate data was the same as in Experiment 1. The RT data was subjected to a three-way repeated measure ANOVA with target presence, set size and repetition as within-subject factors. As in Experiment 1, there was no AI intervention in Experiments 2a and 2b and therefore the same unique set of solution could be obtained for d, $s1$ and $s2$.

3.5 Data exclusion

Experiment 2a required a screen resolution smaller than 2000×2000 so that the noisy background would cover the whole screen. In Experiment 2a, one participant whose screen resolution did not meet the 2000×2000 requirement was excluded. Next, the same exclusion criteria as in Experiment 1 were applied. Trials with 2.5 SD outlier RTs (3.33%) and motor correction (0.91%) were excluded for each observer. After removing both copies of the aforementioned trials, 93.52% remained. One participant was removed from Exp 2a based on a d' of -0.04 (guessing). For the remaining participants, min d' = 1.66, max d' = 3.94.

For Experiment 2b, 2.80% and 0.39% of the trials were excluded due to outlier RTs and motor correction. 94.48% of the trials remained after removing both copies of those trials. No participant was removed based on the d' calculated from the remaining trials (min d' = 2.10, max d' = 4.80).

3.6 Results

3.6.1 RTs

Figure 6 shows RTs on correct response trials from Experiments 2a and 2b. As can be seen, the second repetition of the trial is somewhat faster than the first, especially for absent trials. In Experiment 2a, the results from the three-way ANOVA suggest that there was a main effect of repetition [$F(1, 18) = 11.84$, $p = 0.003$, $\eta_p^2 = 0.40$]. The interaction between target presence and repetition [$F(1, 18) = 4.80$, $p = 0.042$, $\eta_p^2 = 0.21$] was also significant. Post hoc analyses suggest that the effect of repetition was significant on both target present [$t(18) = 3.32$, $p = 0.004$] and target absent trials [$t(18) = 2.92$, $p = 0.009$], but the effect of repetition was larger on target absent trials (Mean Difference = 554 ms) than on target present trials (Mean Difference = 148 ms). The interaction between set size and repetition was not significant [$F(1, 18) = 3.49$, $p = 0.08$, $\eta_p^2 = 0.16$]. The three-way interaction among all the factors was not significant either

[F(1, 18) = 3.16, p = 0.09, η_p^2 = 0.15]. The full results of the three-way ANOVA are presented in Figure S2 in the appendix.

In Experiment 2b, the results from the three-way ANOVA were similar. They show that there was a main effect of repetition [F(1, 20) = 18.06, p < 0.001, η_p^2 = 0.47]. The interaction between target presence and repetition was also significant [F(1, 20) = 9.78, p = 0.005, η_p^2 = 0.33]. Post hoc analyses found that the effect of repetition was significant on both target present [t(20) = 2.27, p= 0.034] and target absent trials [t(20) = 4.09, p < 0.001], but it was larger on target absent trials (Mean Difference = 970 ms) than on target present trials (Mean Difference = 224 ms). The interaction between set size and repetition was not significant [F(1, 20) = 0.48, p = 0.50, η_p^2 = 0.02]. The three-way interaction among all the factors was not significant either [F(1, 20) = 0.004, p = 0.95, η_p^2 = 0.00]. The full results of the three-way ANOVA are presented in Figure S3 in the appendix.

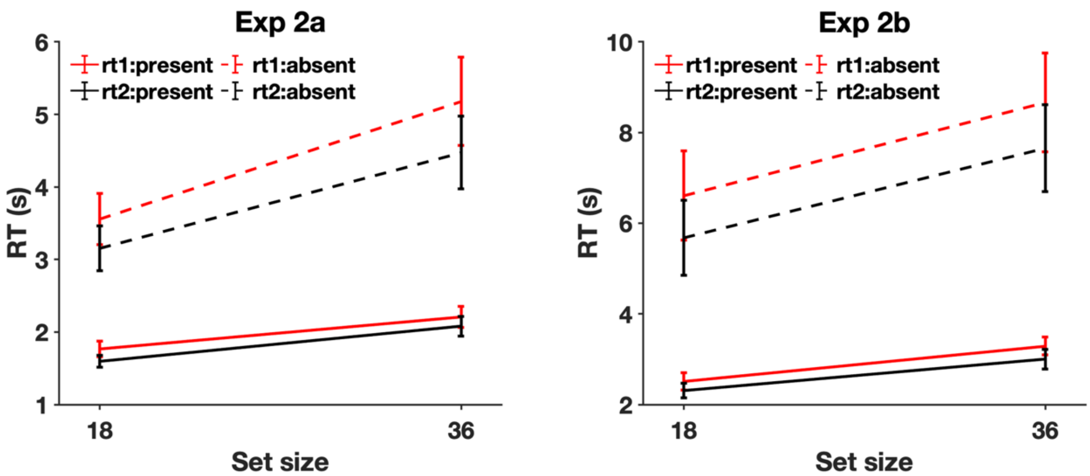


Figure 6: RTs from Experiments 2a and 2b as a function of set size, target presence, and repetition. Red lines: First presentation, Black lines: Second presentation. Full lines: present trials, Dotted lines: absent trials.

3.6.2 Miss rates

The miss errors are the main focus of interest here. Figure 7 shows the results of miss rate analyses for Experiments 2a and 2b. Compared to Experiment 1, it is clear that there were more errors and that those errors were less strictly stochastic. For Experiments 2a and 2b, the scatter plots show that the observed data (blue dots) lie between the deterministic (black circles) and the stochastic (red circles) predictions, indicating that the errors were a mix of both types. In Experiment 2a, a one-sample t-test showed that the deterministic error proportion d was significantly different from 0 [$t(18) = 10.45$, $p < 0.001$, Cohen's d = 2.40], suggesting the existence of deterministic errors in Experiment 2a. No learning effect was observed as suggested by the nonsignificant difference between $s1$ and $s2$ [$t(18) = 1.52$, $p = 0.15$, Cohen's d = 0.35]. Experiment 2b essentially replicated the results from Experiment 2a. The deterministic error rate was significantly different from 0 [$t(20) = 6.26$, $p < 0.001$, Cohen's d = 1.37]. The difference between $s1$ and $s2$ was not significant [$t(20) = 0.34$, $p = 0.74$, Cohen's d = 0.07].

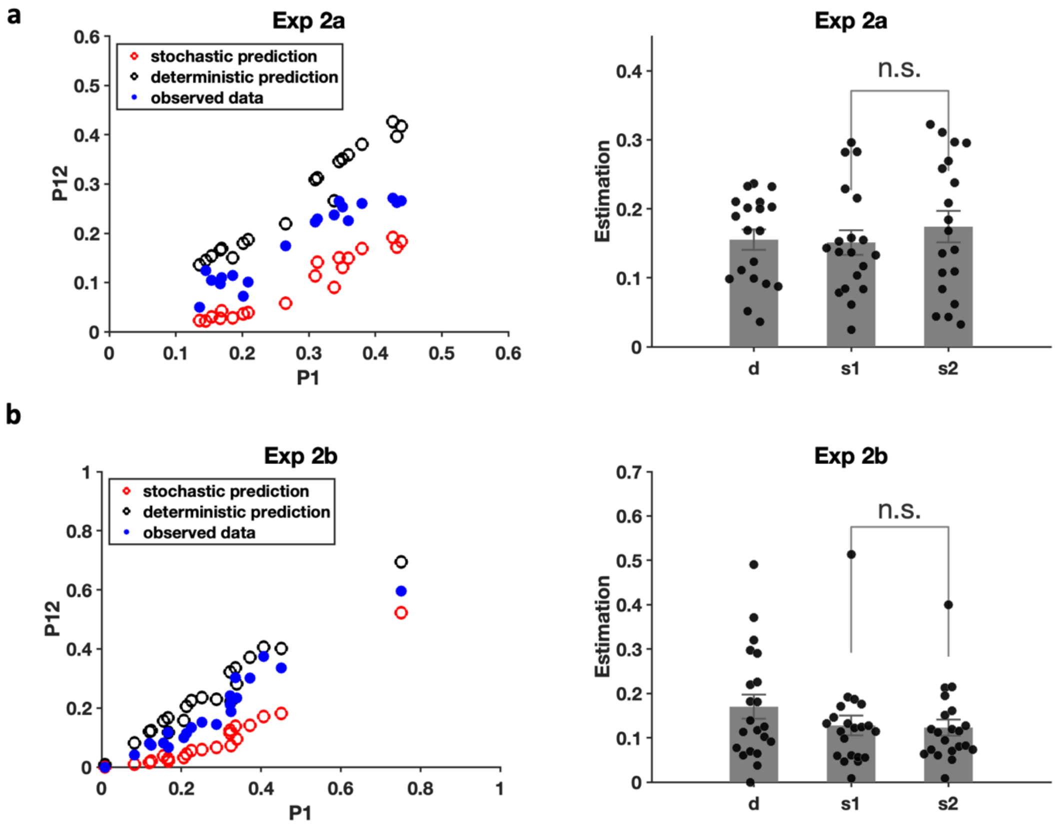


Figure 7: Miss rate analyses for Experiments 2a and 2b. (a) data from Experiment 2a. (b) data from Experiment 2b.

As can be seen in Figure 5, the letters in Experiments 2a and 2b were of varying contrast. Contrast on a non-uniform background can be defined in several ways (Peli, 1990). The precise details are not critical here. As cartooned in Figure 8, what matters is that low contrast items are harder to see and find than high contrast and, as shown in 8b, those low contrast images generate lower accuracy. For purposes of analysis, we computed the contrast as the target-background dissimilarity, i.e., target gray minus the average background gray in a square region surrounding the target. The exact size of the background region does not matter much based on the calculation results, so we chose the background patch outlined by the small blue square for the following analyses.

The box's size is twice the length of the lines composing the letter T or L. As the “low contrast Ls” in Figure 8 should illustrate, on a non-uniform background a letter may be detectable, even when T-B is near zero.

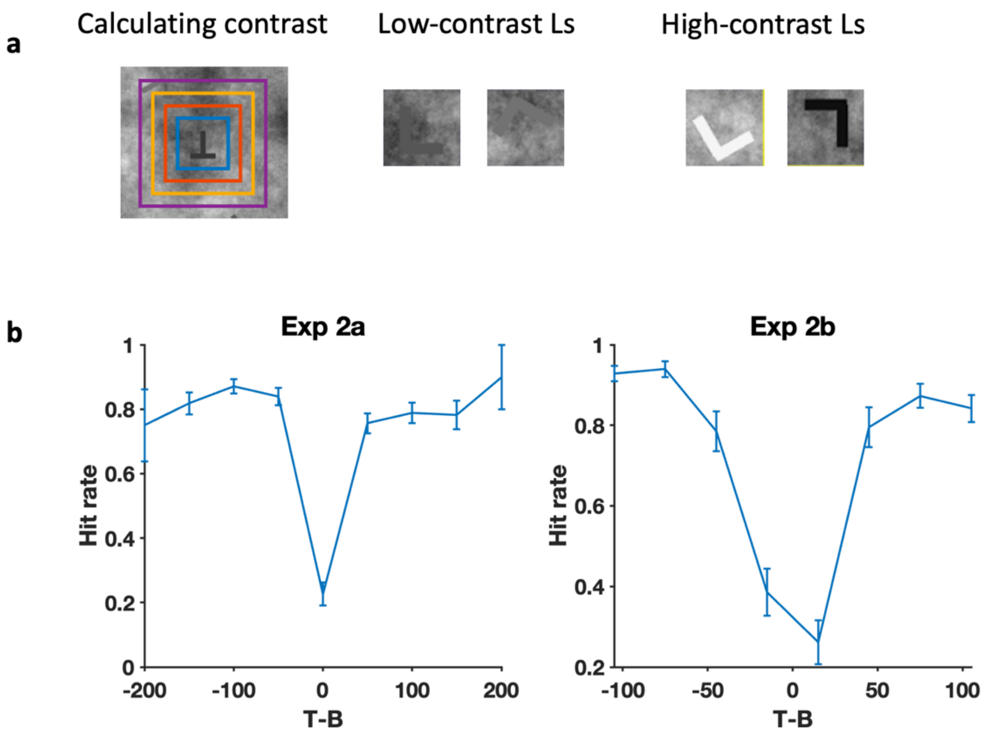


Figure 8: The effects of contrast on hit rate in Experiments 2a and 2b. (a) Example of contrast calculation. (b) Hit rate as a function of T-B. Data in Experiment 2a were binned (bin width: 50) to calculate the hit rate for each bin.

The impact of contrast on error rate is clearly illustrated in the graphs of Figure 8b. Unsurprisingly, observers were far more likely to miss targets if those targets were of low contrast. Of more interest, if we separately analyse low contrast and higher contrast stimuli, we see that the low contrast errors are more likely to be deterministic while high contrast errors are largely stochastic. This is shown in Figure 9.

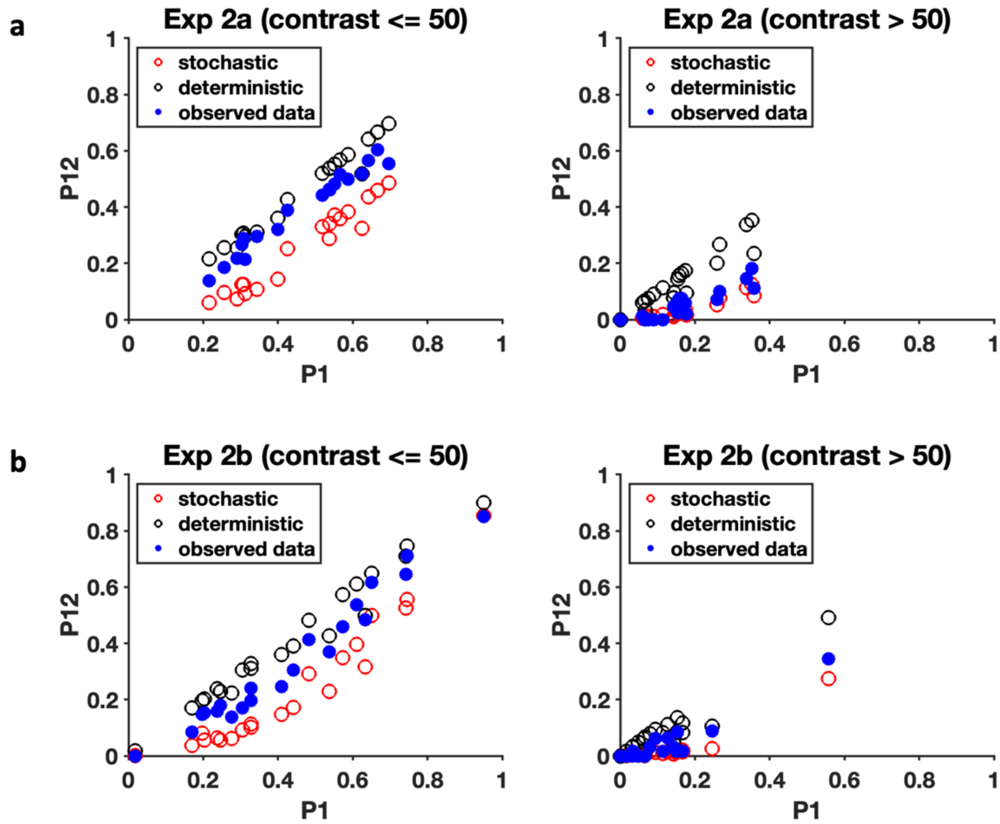


Figure 9: Error data for each observer (blue dots) for low contrast and higher contrast targets. Contrast was calculated by the absolute value of T – B. The red dots show the stochastic prediction and the black dots show the deterministic prediction.

3.6.3 Experiment 2 discussion

Compared to Experiment 1, Experiments 2a and 2b involved a more difficult T-vs-L task where the target could be of very low contrast. RT data in Experiments 2a and 2b showed that observers still responded faster in round 2 than in round 1, as they did in Experiment 1. However, different patterns were observed in the miss rate data. Miss rate analyses demonstrated that while errors were purely stochastic in Experiment 1 they were a mix of deterministic and stochastic errors in the Experiment 2. When analysed separately, the low-contrast targets in Experiments 2a and 2b appear to yield more deterministic errors. In addition, although there seemed to be some learning effect on RTs, no such effect was observed on search accuracy in these two experiments. It would not be terribly interesting to discover that observers do not find targets that they cannot see. However, in this case, these are targets that are *harder* to see, not impossible. This raises the possibility that these low contrast targets might be found if observers could be induced to pay attention to those items more effectively. In Experiment 3, we tried several methods for moving attention around the scene in a possibly useful manner. We call these methods "mindless AI" because our goal here is to direct attention with an intervention that does not need to know where the target actually resides or even where it is likely to reside. Of course, if a real AI system can solve a search problem, that changes the LBFTS issue. However, as that is not the case in general, it could be useful to have a generic intervention that reduces errors and that can be easily implemented in real-world tasks.

4. Experiments 3a, 3b and 3c: Mindless AI

In Experiments 3a, 3b and 3c, different forms of a mindless (or almost mindless) AI intervention were introduced to reduce the errors. Experiments 3a and 3b moved attention around the field in an effort to decrease the chance of overlooking an item of interest. In Experiment 3a, the mindless AI intervention was a yellow dot jumping to random places in the search display, summoning attention or the eyes to follow. In

Experiment 3b, a transparent, outline square moved in a spiral path from center to periphery in the hope of inducing observers to search regularly. In Experiment 3c the intervention was less completely "mindless". Each of the letters in the search display was highlighted by a yellow square around it in an effort to reduce the chance of missing a low contrast target. Highlighting each item is akin to having an AI that figures out where all the interesting information might be but that cannot discriminate targets from distractors.

4.1 Participants

Experiments 3a, 3b and 3c were run online on the Pavlovia platform (https://pavlovia.org). All the participants were recruited via Prolific. In Experiment 3a, we tested 20 participants (6 males, 9 females, mean = 32.1, SD = 11.3, min = 22, max = 59, five participants did not provide their gender and age information). In Experiment 3b, we tested 20 participants (7 males, 11 females, mean = 27.9, SD = 6.2, min = 20, max = 40, two participants did not provide the gender and age information). In Experiment 3c, we tested 20 participants (9 males, 5 females, mean = 24.9, SD = 5.5, min = 19, max = 41, six participants did not provide the gender and age information). All participants reported normal or corrected-to-normal vision and gave their informed consent before they began the experiment. Participants received 8 GBP for their participation. Ethics approval came from University of Manchester (2023-16117-29230 for Exp 3a, 2023-16117-30373 for Exp 3b, 2023-16117-30584 for Exp 3c).

4.2 Stimuli & apparatus

The exact same set of stimuli used in Experiment 2b were also used in Experiment 3a, 3b and 3c (Figure 5b), but in the AI trials there was either a randomly moving yellow dot (Experiment 3a), a transparent yellow square that moved in a spiral fashion (Experiment 3b), or a set of static yellow squares that highlighted the positions of all items (Experiment 3c).

4.3 Design & procedure

Participants were instructed to press ‘j’ if they found the target T and ‘f’ if they did not. The search time was limited to 20 seconds for Experiments 3a, 3b and 3c for the purpose of controlling the online experiment time. Trial by trial feedback was not given, but the percentage correct was displayed at the end of each block. The stimuli for Experiments 3a, 3b and 3c were the same as that of Experiment 2b, except for the introduction of the mindless AI on half of the repetition trials. In Experiment 3a (Figure 10, left, the yellow dot was enlarged here for visualization), a yellow dot (size = 0.01 screen height) jumped at random places in the search display as the mindless AI intervention, remaining at each location for 500 ms. In Experiment 3b (Figure 10, middle), a transparent square (size = 1/3 screen height) with yellow borders moved around the stimuli area on AI intervention trials, following a spiral path. Participants were instructed to follow the square when it appeared. In Experiment 3c (Figure 10, right), all the letters were highlighted by yellow squares around them (size = 0.06 screen height) as the AI intervention. In Experiments 3a and 3b, the AI intervention was not related to the presence or the location of the target. In Experiment 3c, the presence of the AI intervention was not related to the presence of the target either, but it did point out the positions of all the letters and thus also possibly the target. For half of the stimuli, we had no AI on the first copy of the stimulus but AI on the second copy (noAI – AI condition). As comparison, for the other half of the stimuli, we had AI on the neither the first or the second copy (noAI – noAI condition). All versions of Experiment 3 had therefore a design with four factors, each with two levels: repetition, AI, set size, and target. Participants were required to finish a 12-trial practice session before the experiment and would only be able to begin the experiment with an accuracy higher than 0.75, otherwise, they had to repeat the practice.

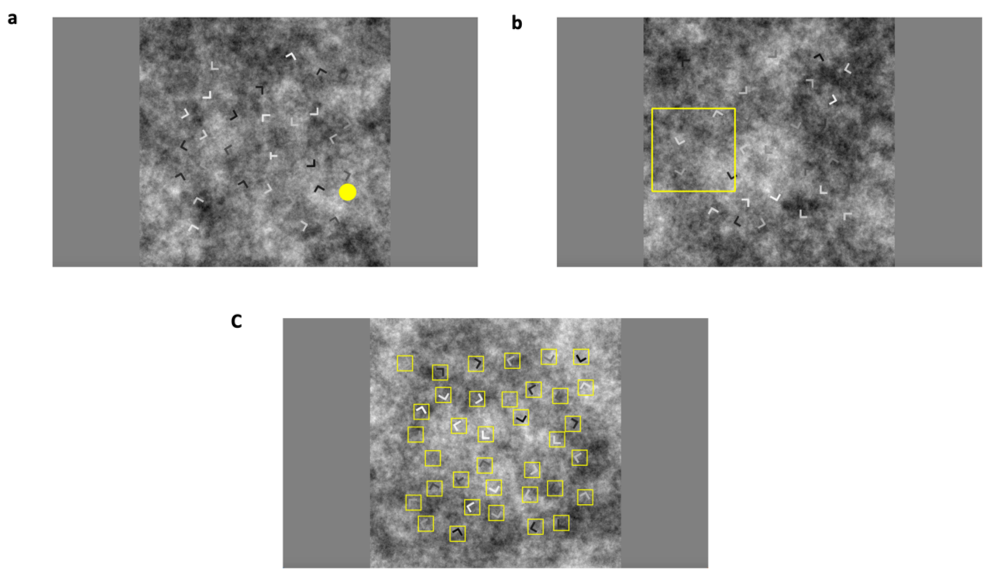


Figure 10: Illustration of the AI intervention in Experiments 3a, 3b and 3c.

4.4 Analysis method

We focused on the RT data and miss rate data with our primary interest in the error data as in the previous experiments. The four-way repeated measure ANOVA with target presence, set size, repetition and AI intervention as within-subject factors was conducted for Experiments 3a, 3b and 3c. Since the three experiments here also involved a typical visual search task, it was expected to observe main effects of target presence (longer reaction times for absent trials), set size (longer reaction times for larger set sizes) and their two-way interaction (steeper reaction time slopes for absent trials). These effects will not be reported in detail in the following Results section. The results of the full ANOVA can be found in supplementary tables.

To analyse miss rate, we also calculated $P1$, $P2$ and $P12$ to estimate $d1$, $d2$, s1 and $s2$. For the noAI – noAI trials in Experiments 3a, 3b, 3c, there was no AI intervention. A fixed deterministic rate was therefore assumed for round 1 and round 2, i.e., $d = d1 = d2$. The solution of d, $s1$ and $s2$ was the same as in the previous experiments. For noAI – AI trials in Experiments 3a, 3b and 3c, the AI intervention was introduced on the second copy of trials. If AI reduces deterministic errors, the assumption $d1 = d2$ becomes $d1 \geq d2$. This means that the assumption $d1 = d2$ can no longer be used to arrive at a unique solution. $s2$ and $d2$ are still uniquely determined by solving the equations, but there are multiple solutions for $s1$ and $d1$. However, considering that deterministic errors should be persistent when no additional intervention is implemented, the assumption $d1$ (noAI - AI) = d (noAI - noAI) should hold, thus leading to a unique solution for $s1$ in the noAI – AI condition.

$$s2 = \frac{P2 - P12}{1 - P1}$$
$$d2 = \frac{P12 - P1 * P2}{1 - P1 - P2 + P12}$$
$$s1 = \frac{P1 - d1}{1 - d1}$$

It also should be noted that the split of stimuli into the noAI - AI group and the noAI – noAI group was random for each participant. Therefore, it is possible that one group contained more deterministic error prone stimuli and the other group contains fewer such stimuli, but on average we should have $d1$ (noAI - AI) = d (noAI - noAI). If the noAI – noAI group contains much fewer deterministic error prone stimuli than the other group, $d1$ (noAI - AI) = d (noAI - noAI) will lead to an overestimate of the actual $d1$, which probably results in a negative $s1$ when the overestimate of $d1$ is larger than $P1$. To avoid such cases, any participant with a negative estimate of $s1$ will be excluded when the analysis concerns the estimate of the deterministic error proportion $d1/d2$ and the stochastic error rate $s1/s2$.

4.5 Data exclusion

The same exclusion criteria as in the previous experiments were applied for Experiments 3a, 3b and 3c. We removed trials with RTs smaller or greater than 2.5 SD from the mean RT in each condition for each observer (3.21% in Exp 3a, 3.22% in Exp 3b, 3.87% in Exp 3c). Then trials where participants corrected their motor responses were removed (0.70% in Exp 3a, 0.76% in Exp 3b, 0.69% in Exp 3c). When one trial was removed, the other copy of the trial was also removed (remaining trials: 93.77% in Exp 3a, 93.71% in Exp 3b, 92.42% in Exp 3c). After the removal of the above trials, we further checked the d' of all the participants. Participants with d' beyond 2.5 SD from the group mean for each individual experiment were excluded. One participant with d' = 1.39 was removed from Exp 3a (for remaining participants, min d' = 1.98, max d' = 3.86). One participant with d' = 0.69 was removed from Exp 3b (for remaining participants, min d' = 2.53, max d' = 3.59). One participant with d' = 1.68 was removed from Exp 3c (for remaining participants, min d' = 2.38, max d' = 4.11).

4.6 Results

4.6.1 RTs

Figure 11 shows RTs on correct response trials from Experiments 3a, 3b and 3c. It appears that the AI had very little qualitative effect in Experiments 3a and 3b. RTs were faster on the second copy of the stimuli, especially for absent trials, but the presence or absence of the mindless AI made little difference. In Experiment 3c, by contrast, the presence of the AI slowed the RT on the second appearance. Note that RT2 is faster when the AI is absent and slower when the RT is present. We can presume that the boxes marking all the items induced the observers to attend to more of the items or spend more time checking the highlighted areas.

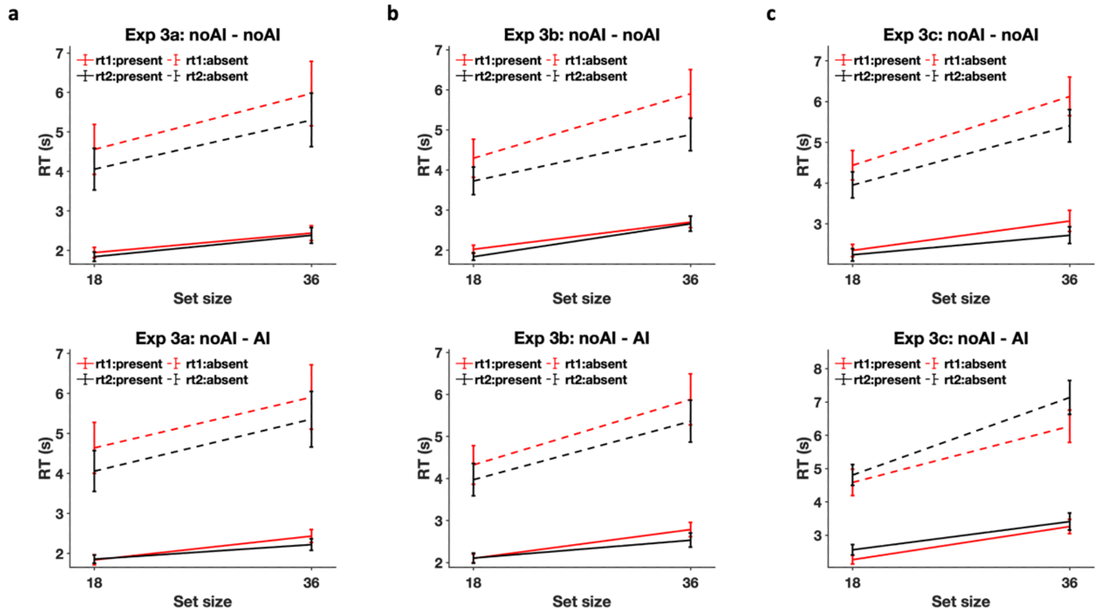


Figure 11: RTs from Experiments 3a, 3b and 3c as a function of set size, target presence, and repetition. Red lines: First presentation, Black lines: Second presentation. Full lines: present trials, Dotted lines: absent trials..

To evaluate this statistically, four-way repeated measure ANOVAs with target presence, set size, repetition and AI intervention as within-subject factors were conducted for Experiments 3a, 3b and 3c.

In Experiment 3a, the four-way interaction was significant [$F(1, 18) = 6.61$, $p < 0.05$, $\eta_p^2 = 0.27$], so two three-way ANOVAs with set size, repetition and AI intervention as within-subjects factors were conducted for target present and target absent trials separately. On target present trials, the main effect of repetition was not significant [$F(1, 18) = 3.25$, $p = 0.088$, $\eta_p^2 = 0.15$], but on target absent trials, it was [$F(1, 18) = 14.30$, $p = 0.001$, $\eta_p^2 = 0.44$], showing that observers responded faster in round 2 than in round 1 for absent trials. The two-way interaction between repetition and AI was not significant for either target present [F(1, 18) = 0.028, p = 0.87, η_p^2 = 0.002] or target absent trials [F(1, 18) = 0.046, p = 0.83, η_p^2 = 0.003], suggesting that the AI intervention did not have any effect on reaction time in Experiment 3a. The full results of the four-way ANOVA and three-way ANOVAs are presented in Figures S4-1, S4-2 and S4-3 in the appendix.

In Experiment 3b, the four-way interaction was also significant [F(1, 18) = 4.67, p < 0.044, η_p^2 = 0.21], so again two three-way ANOVAs were conducted for target present and target absent trials separately. On target present trials, the main effect of repetition was almost significant [F(1, 18) = 4.17, p = 0.056, η_p^2 = 0.19]. The two-way interaction between repetition and AI was not significant [F(1, 18) = 0.008, p = 0.93, η_p^2 = 0.00], suggesting that the AI intervention did not influence RTs on target present trials. On target absent trials, the main effect of repetition was significant [F(1, 18) = 8.32, p = 0.01, η_p^2 = 0.32], showing that observers made faster responses in round 2 than in round 1 on target absent trials. The two-way interaction between repetition and AI was significant as well, [F(1, 18) = 4.95, p = 0.039, η_p^2 = 0.22]. The AI intervention appears to have made observers search longer on target absent trials. The full results of the four-way ANOVA and three-way ANOVAs are presented in Figures S5-1, S5-2 and S5-3 in the appendix.

In Experiment 3c, the four-way interaction was again significant [F(1, 18) = 11.53, p = 0.003, η_p^2 = 0.39], so two three-way ANOVAs were conducted for target present and target absent trials separately. On target present trials, the main effect of repetition was not significant [F(1, 18) = 0.002, p = 0.96, η_p^2 = 0.00]. On target absent trials, the main effect of repetition was not significant either [F(1, 18) = 0.019, p = 0.89, η_p^2 = 0.001]. However, there was a strong two-way interaction between repetition and AI for both target present [F(1, 18) = 7.80, p = 0.012, η_p^2 = 0.30] and target absent trials [F(1, 18) = 45.79, p < 0.001, η_p^2 = 0.72], demonstrating that the AI intervention in Experiment 3c slowed the search regardless of target presence. The effect of this no-quite-mindless AI was larger on target absent trials (noAI – noAI: RT2 – RT1 = -622 ms, noAI – AI: RT2 – RT1 = 529 ms) than on target present trials (noAI – noAI: RT2 – RT1 = -222 ms, noAI – AI: RT2 – RT1 = 202 ms).

4.6.2 Miss Rates

Figure 12 shows the miss rate difference (P2 – P1) for the noAI – noAI stimuli and the noAI – AI stimuli. The critical comparison in all cases is between the miss rate on the second appearance compared to the first appearance, since the AI was only implemented on the second copy in the noAI – AI group. Did the AI intervention lower the miss error rate on the second appearance compared to when there was no AI intervention? In Experiments 3a and 3b, the AI did not reduce the error rate [paired t-tests, 3a: t(18) = 0.19, p = 0.85, Cohen's d = 0.04; 3b: t(18) = 0.63, p = 0.54, Cohen's d = 0.14]. In Experiment 3c, the drop in misses on second presentation was larger in the presence of the AI than in the absence of AI [t(18) = 3.38, p = 0.0033, Cohen d = 0.78], suggesting that the AI intervention in Experiment 3c effectively reduced errors.

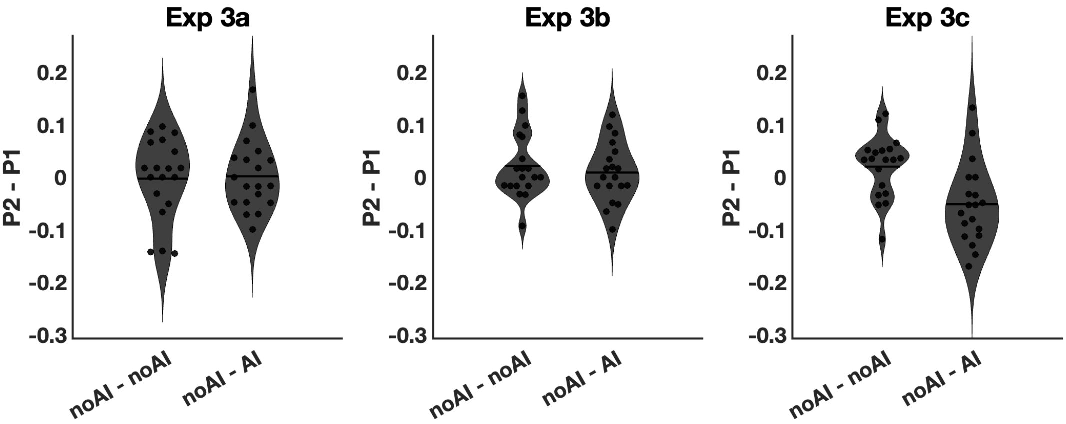


Figure 12: Miss rate difference (P2 – P1) for the noAI – noAI stimuli and noAI – AI stimuli.

As can be seen in the scatter plots of Figure 13, all versions of Experiment 3 replicated the main result of Experiment 2 in producing a mix of stochastic and deterministic. The observed data lie between the stochastic and deterministic predictions regardless of AI presence in Experiments 3a, 3b and 3c, indicating that the errors were a mix of deterministic and stochastic errors in all the conditions. As discussed in the *Analysis method* section, the proportion of deterministic errors and stochastic errors can be calculated by solving the relevant equations. For noAI – noAI trials in Experiments 3a, 3b, 3c, the proportion of deterministic errors was fixed for round 1 and round 2 as in the previous experiments ($d = d1 = d2$), resulting in three parameters d, $s1$ and $s2$. For noAI – AI trials in Experiments 3a, 3b, 3c, the presence of the AI intervention could influence the proportion of either deterministic or stochastic errors (or both) in round 2. Thus there were four parameters $d1$, $d2$, $s1$ and $s2$. Unique solutions could still be obtained for $d2$ and $s2$ by solving the equations, but there could be multiple solutions for $d1$ and $s1$. Under the assumption that the deterministic errors should be persistent when no interference was added, the $d1$ parameter in the noAI – AI condition was taken to be identical with d from the noAI – noAI condition. This could be used to derive a unique solution for $s1$.

In Experiment 3a, the noAI – noAI trials replicated the results from Experiments 2a and 2b. The deterministic error rate was significantly different from 0 [$t(18) = 6.56$, $p < 0.001$, Cohen's $d = 1.51$] and no learning effect was found [$t(18) = 0.32$, $p = 0.76$, Cohen's $d = 0.07$]. For the noAI – AI trials, the critical comparisons are between d1 and d2 and/or s1 and s2. Did the AI reduce the error rate? One participant got a negative $s1$ after we replaced $d1$ (noAI - AI) with d (noAI - noAI) and was therefore excluded from the following analysis. Neither the deterministic proportions nor the stochastic rates were significantly different between round 1 and round 2 [$d1$ vs. $d2$: $t(17) = 0.18$, $p = 0.86$, Cohen's $d = 0.04$; $s1$ vs. $s2$: $t(17) = 0.04$, $p = 0.97$, Cohen's $d = 0.01$], suggesting that the AI intervention in Experiment 3a failed to reduce either type of errors.

Results from Experiment 3b were similar to the results from Experiment 3a. For the noAI – noAI trials, the deterministic error rate was significantly different from 0 [$t(18) = 14.39$, $p < 0.001$, Cohen's $d = 3.30$] and no learning effect was found [$t(18) = 1.42$, $p = 0.17$, Cohen's $d = 0.33$]. For the noAI – AI trials, two participants were excluded due to negative $s1$. The AI intervention also failed to reduce either type of errors [$d1$ vs. $d2$: $t(16) = 0.80$, $p = 0.43$, Cohen's $d = 0.19$; $s1$ vs. $s2$: $t(16) = 0.72$, $p = 0.48$, Cohen's $d = 0.17$].

In Experiment 3c, the results for the noAI – noAI trials were, again, the same as in the previous experiments. The deterministic error rate was significantly different from 0 [t(18) = 9.76, p < 0.001, Cohen's d = 2.24] and no learning effect was found [t(18) = 1.64, p = 0.12, Cohen's d = 0.38]. The noAI – AI trials produced a different, more interesting result in Experiment 3c. Three participants were excluded due to negative $s1$. The comparison between $d1$ and $d2$ for the remaining participants did show a significant difference. $d2$ was smaller than $d1$ [t(15) = 2.69, p = 0.017, Cohen's d = 0.67], suggesting that the AI intervention in Experiment 3c effectively reduced deterministic errors. No significant effect was found on stochastic errors [t(15) = 0.86, p = 0.40, Cohen's d = 0.22].

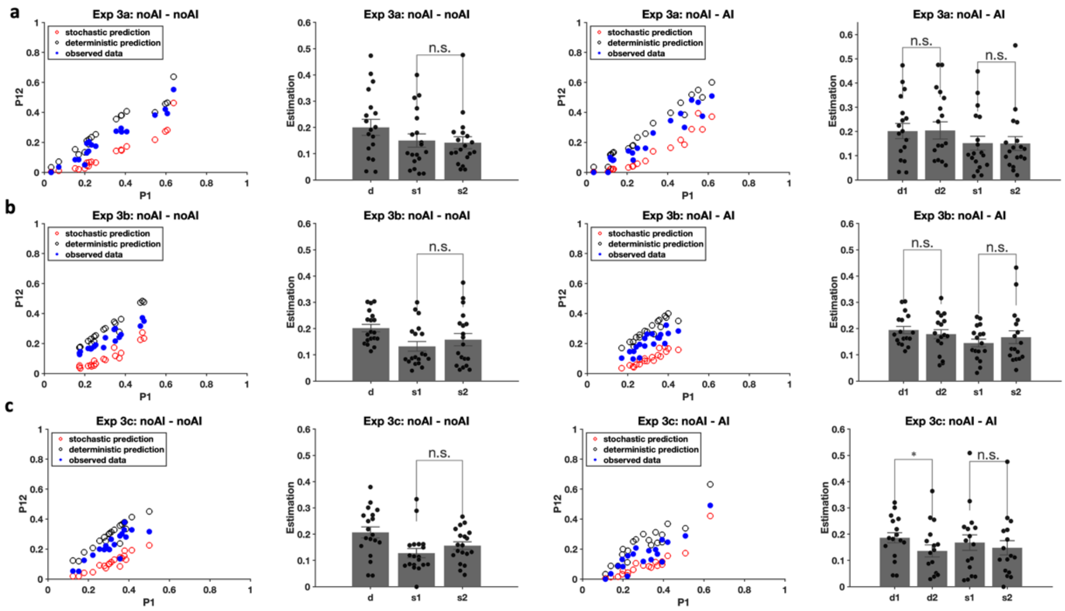


Figure 13: Miss rate analyses. (a) data from Experiment 3a. (b) data from Experiment 3b. (c) data from Experiment 3c.

3.3.3 Experiment 3 discussion

Experiments 3a, 3b and 3c used the same set of stimuli as in Experiment 2b but introduced mindless AI interventions in an attempt to reduce error rates. Experiments 3a and 3b were efforts to spread attention around the display without needing to know anything about the contents of the display. In Experiment 3a, this was implemented as a dot that jumped to random locations. In Experiment 3b, an outline square moved systematically. Neither of these interventions had an impact on the errors

However, the somewhat less mindless AI intervention in Experiment 3c did have some effects. In 3c, when all of the locations of items were outlined on the screen participants slowed down, compared to the noAI - noAI condition. More importantly, miss error rates were reduced. The analysis of those errors indicates that the intervention in 3c had its biggest effect on deterministic errors. It seems likely that the outline boxes directed attention to some lower contrast items that might have otherwise been overlooked. Paired t-tests suggest that for the noAI – AI stimuli in Experiment 3c, the target contrast in correct response trials was significantly lower in round 2 than in round 1[$t(18) = 6.13$, $p < 0.001$, Cohen's $d = 1.41$] while in all other situations the difference was not significant.

General Discussion

Search errors are ubiquitous in tasks from the lab and real-life. Although it is unlikely that such errors could ever be completely eliminated (Brady, 2017), efforts to reduce errors are still worthwhile and hold significant potential to improve performance on socially important search tasks. In this paper, we were interested in the nature of miss errors in a simple laboratory-based search task. We chose a typical T-vs-L task but, we presume, that choice is not critical. Even in such a simple task, errors still occur at a steady rate. Those errors could be purely stochastic, purely deterministic or a mix of both types of errors. Our approach to distinguishing stochastic from deterministic contributions to errors was to show each display twice in the experiment. The straight-forward logic is that a stochastic error, made on one appearance of a display, tells you nothing about whether it will be missed on the second appearance of the displays. On the other hand, if that target is missed for some deterministic cause, it would definitely be missed again at the next opportunity. Six experiments with repeated displays were conducted. In Experiment 1, all the letters were white and presented against a uniform gray background. The target letter was always clearly visible when present in the search array. Our analysis showed that the errors in this experiment were almost purely stochastic. In Experiments 2a and 2b, the letters were of different grayscale values and were presented against a noisy background. The target letters varied from clearly visible to low contrast. Our results suggested that the errors in Experiments 2a and 2b were a mix of both types of errors with lower contrast targets accounting for more of the deterministic errors. Experiments 3a, 3b and 3c used the same stimuli as in Experiment 2 and attempted to reduce the errors with different forms of "mindless" AI intervention. In Experiment 3a, a yellow dot jumped at random places in the search display on some trials, remaining at each location for 500 ms. This was to enhance observers' attention at those locations. In Experiment 3b, a transparent square with yellow borders appeared on some trials, following a spiral path. This intervention was intended to guide observers to search the entire display. In Experiment 3c, all the letters were highlighted by yellow squares around them on the AI related trials. Our results suggest that only the AI intervention that had knowledge of item locations could effectively reduce the errors and the reduced errors were mainly deterministic errors. These results make it less likely that a truly 'mindless' intervention would be helpful. That said, one could try more forceful efforts to get participants to look at "everything" and, thus, not overlook targets like low contrast items. For instance, military surveillance officers used to divide large aerial photographs into a grid of smaller regions and systematically mark each region to indicate that it had been examined. This could reduce errors caused by simply overlooking some region. Of course, such a protocol greatly increases the time per image. In 3a and 3b, we attempted to get a similar benefit at less of a cost. Sadly, we did not succeed. In situations where it is worth paying the cost in time, a more mandatory style of intervention could be tried.

Our experiments focused on the miss errors in the simple T-vs-L visual search task. However, when it comes to some real-life tasks, the target might not be as specific as the letter T in our task. The target definition might be broader (e.g. find animals) and/or the target might be more ambiguous (Is that really a cancerous skin lesion?). There have been other attempts to reduce the errors in these more complex situations. For example, Nartker et al. (2020) tested three different methods to reduce categorical errors in a "mixed hybrid search" task where participants searched for a list of targets; some specific (this hammer) and some categorical (any animal). In mixed hybrid search, participants tended to miss more categorical targets. To reduce such errors, several strategies were tried: (1) boosting categorical targets in memory; (2) separating the responses for specific and categorical targets; (3) full check list procedure that required participants to make an explicit response to the presence or absence of each type of target. Of all these measures, only the full checklist procedure effectively reduced categorical target errors. As with dividing an aerial image into little squares, this improvement comes at the expense of substantially longer reaction time.

Low-prevalence targets are also more frequently missed. Horowitz (2017) summarized some of the experimental manipulations that have be tried to reduce those errors. These include introducing a regime of brief retraining periods with high prevalence and full feedback (Wolfe et al., 2007), reducing the uncertainty of examined area by eye movement feedback (Drew & Williams, 2017) and providing an opportunity to correct motor errors (Fleck & Mitroff, 2007). The success of such methods is mixed. In contrast to the manipulations of the task, other researchers have focused more on individual differences to identify those who are likely to perform better on a low prevalence search task (Peltier & Becker, 2017, 2020). Such individual approaches could also provide some insights about how to improve real-world visual search performance.

In summary, errors in visual search are ubiquitous and stubborn. Our results suggest that errors may be almost completely stochastic when targets are clearly visible. Such errors may be hard to be reduced by any method that does not come down to spending more time and “paying” more attention. When targets are harder, but not impossible, to see, more of those hard to see targets appear to be missed in a deterministic manner. In the present experiments, deterministic errors due to low target contrast could be reduced to some extent by appropriate, rather simple interventions. Drawing attention to targets that might otherwise be reliably overlooked seems like a potentially promising approach to reducing LBFTS errors.

Acknowledgements

JMW was supported by NIH-NEI: EY017001, NSF: 2146617, and NIH-NCI: CA207490. AL and JH were supported by UKRI grant ES/X000443/1. The authors also thank Wentao (Taylor) Si for his valuable inputs and constructive suggestions on this work.

References

Adamo, S. H., Cain, M. S., & Mitroff, S. R. (2013). Self-induced attentional blink: A cause of errors in multiple-target search. Psychological science, 24(12), 2569-2574.

Adamo, S. H., Cox, P. H., Kravitz, D. J., & Mitroff, S. R. (2019). How to correctly put the "subsequent" in subsequent search miss errors. Attention, Perception, & Psychophysics, 81, 2648-2657.

Berbaum, K. S., Franken Jr, E. A., Dorfman, D. D., Rooholamini, S. A., Coffman, C. E., Cornell, S. H., ... & Kao, S. C. (1991). Time course of satisfaction of search. Investigative radiology, 26(7), 640-648.

Berbaum, K. S., Franken Jr, E. A., Dorfman, D. D., Rooholamini, S. A., Kathol, M. H., Barloon, T. J., ... & MONTGOMERY, W. J. (1990). Satisfaction of search in diagnostic radiology. Investigative radiology, 25(2), 133-140.

Brady, A. P. (2017). Error and discrepancy in radiology: inevitable or avoidable?. Insights into imaging, 8, 171-182.

Drew, T., Võ, M. L. H., & Wolfe, J. M. (2013). The invisible gorilla strikes again: Sustained inattentional blindness in expert observers. Psychological science, 24(9), 1848-1853.

Drew, T., & Williams, L. H. (2017). Simple eye-movement feedback during visual search is not helpful. Cognitive research: principles and implications, 2(1), 1-8.

Ekroll, V., Svalebjørg, M., Pirrone, A., Böhm, G., Jentschke, S., van Lier, R., ... & Høye, A. (2021). The illusion of absence: how a common feature of magic shows can explain a class of road accidents. Cognitive research: principles and implications, 6(1), 1-16.

Fleck, M. S., & Mitroff, S. R. (2007). Rare targets are rarely missed in correctable search. Psychological science, 18(11), 943-947.

Gandomkar, Z., & Mello-Thoms, C. (2019). Visual search in breast imaging. The British journal of radiology, 92(1102), 20190057.

Horowitz, T. S. (2017). Prevalence in visual search: From the clinic to the lab and back again. Japanese Psychological Research, 59(2), 65-108.

Hout, M. C., Walenchok, S. C., Goldinger, S. D., & Wolfe, J. M. (2015). Failures of perception in the low-prevalence effect: Evidence from active and passive visual search. Journal of Experimental Psychology: Human Perception and Performance, 41(4), 977-994.

Hovda, T., Hoff, S. R., Larsen, M., Romundstad, L., Sahlberg, K. K., & Hofvind, S. (2022). True and missed interval cancer in organized mammographic screening: a retrospective review study of diagnostic and prior screening mammograms. Academic Radiology, 29, S180-S191.

Hovda, T., Larsen, M., Romundstad, L., Sahlberg, K. K., & Hofvind, S. (2023). Breast cancer missed at screening; hindsight or mistakes? European Journal of Radiology, 110913.

Koivisto, M., Hyönä, J., & Revonsuo, A. (2004). The effects of eye movements, spatial attention, and stimulus features on inattentional blindness. Vision research, 44(27), 3211-3221.

Kuhn, G., & Tatler, B. W. (2011). Misdirected by the gap: The relationship between inattentional blindness and attentional misdirection. Consciousness and Cognition, 20(2), 432-436.

Lumbreras, B., Donat, L., & Hernández-Aguado, I. (2010). Incidental findings in imaging diagnostic tests: a systematic review. The British journal of radiology, 83(988), 276-289.

Mack, A., & Rock, I. (1998). Inattentional Blindness. Cambridge, MA: MIT Press.

Macknik, S. L., King, M., Randi, J., Robbins, A., Teller, Thompson, J., & Martinez-Conde, S. (2008). Attention and awareness in stage magic: turning tricks into research. Nature Reviews Neuroscience, 9(11), 871-879.

Nartker, M. S., Alaoui-Soce, A., & Wolfe, J. M. (2020). Visual search errors are persistent in a laboratory analog of the incidental finding problem. Cognitive Research: Principles and Implications, 5(1), 1-14.

Pammer, K., Sabadas, S., & Lentern, S. (2018). Allocating attention to detect motorcycles: The role of inattentional blindness. Human factors, 60(1), 5-19.

Peli, E. (1990). Contrast in complex images. Journal of the Optical Society of America A, 7(10), 2032-2040.

Peltier, C., & Becker, M. W. (2016). Decision processes in visual search as a function of target prevalence. Journal of Experimental Psychology: Human Perception and Performance, 42(9), 1466 - 1476.

Peltier, C., & Becker, M. W. (2017). Individual differences predict low prevalence visual search performance. Cognitive research: principles and implications, 2, 1-11.

Peltier, C., & Becker, M. W. (2020). Individual differences predict low prevalence visual search performance and sources of errors: An eye-tracking study. Journal of experimental psychology: applied, 26(4), 646 - 658.

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. (2019). PsychoPy2: Experiments in behavior made easy. Behavior research methods, 51, 195-203.

Simons, D. J. (2000). Attentional capture and inattentional blindness. Trends in cognitive sciences, 4(4), 147-155.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. perception, 28(9), 1059-1074.

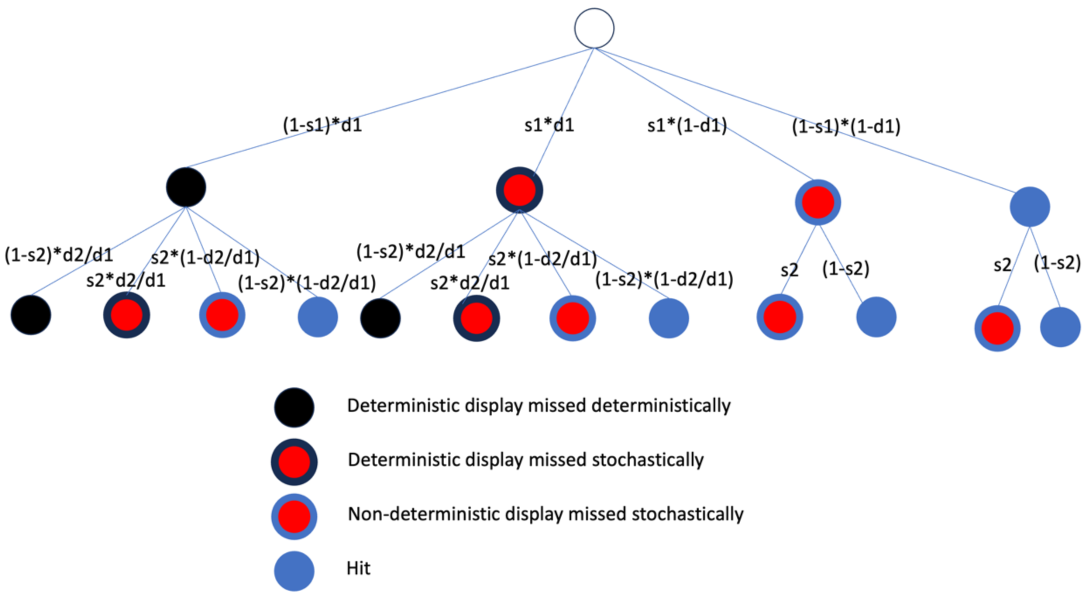
Wolfe, J. M., Alaoui Soce, A., & Schill, H. M. (2017). How did I miss that? Developing mixed hybrid visual search as a 'model system'for incidental finding errors in radiology. Cognitive Research: Principles and Implications, 2, 1-10.

Wolfe, J. M., Kosovicheva, A., & Wolfe, B. (2022). Normal blindness: when we Look But Fail To See. Trends in Cognitive Sciences, 26(9), 809-819.

Wolfe, J. M., Horowitz, T. S., Van Wert, M. J., Kenner, N. M., Place, S. S., & Kibbi, N. (2007). Low target prevalence is a stubborn source of errors in visual search tasks.

Journal of experimental psychology: General, 136(4), 623 - 638.

Appendix A:



- $P1$ is the sum of the probability of the black circle, the red circle with the black border and the red circle with blue border in row 1.

$$P1 = (1 - s1) * d1 + s1 * d1 + s1 * (1 - d1) = (1 - s1) * d1 + s1$$

$$= d1 - s1 * d1 + s1 = d1 * (1 - s1) + s1$$

- $P2$ is the sum of probability of the black circles, the red circles with the black border and the red circles with blue border in row 2.

$$P2 = [(1 - s1) * d1 + s1 * d1] * \left[(1 - s2) * \frac{d2}{d1} + s2 * \frac{d2}{d1} + s2 * \left(1 - \frac{d2}{d1}\right)\right] + s1 * (1 - d1) * s2 + (1 - s1) * (1 - d1) * s2$$

$$= d1 * \left(\frac{d2}{d1} + s2 - s2 * \frac{d2}{d1}\right) + (1 - d1) * s2$$

$$= d2 + s2 * d1 - s2 * d2 + (1 - d1) * s2$$

$$= d2 + s2 * (d1 - d2) + (1 - d1) * s2$$

- $P12$ is the sum of is the sum of probability of the black circles, the red circles with the black border and the red circles with blue border in row 2 excluding the last red circle with blue border stemmed from a hit in row 1.

$$P12 = [(1 - s1) * d1 + s1 * d1] * \left[(1 - s2) * \frac{d2}{d1} + s2 * \frac{d2}{d1} + s2 * \left(1 - \frac{d2}{d1}\right)\right] + s1 * (1 - d1) * s2 = d2 + s2 * (d1 - d2) + s1 * (1 - d1) * s2$$

Appendix B: Analyses on RTs

Repeated Measures ANOVA

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2_p
presence	77.0234	1	77.0234	169.08926	< .001	0.904
Residual	8.1993	18	0.4555			
set size	15.6649	1	15.6649	163.33818	< .001	0.901
Residual	1.7263	18	0.0959			
repetition	0.4940	1	0.4940	6.99613	0.016	0.280
Residual	1.2709	18	0.0706			
presence ✻ set size	4.8117	1	4.8117	115.51955	< .001	0.865
Residual	0.7497	18	0.0417			
presence ✻ repetition	0.0155	1	0.0155	0.37044	0.550	0.020
Residual	0.7534	18	0.0419			
set size ✻ repetition	0.0119	1	0.0119	0.49796	0.489	0.027
Residual	0.4303	18	0.0239			
presence ✻ set size ✻ repetition	6.51e–5	1	6.51e–5	0.00404	0.950	0.000
Residual	0.2901	18	0.0161			

Note. Type 3 Sums of Squares

[4]

Between Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2_p
Residual	24.0	18	1.33			

Note. Type 3 Sums of Squares

Figure S1. RT analyses for Experiment 1.

Repeated Measures ANOVA

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2_p
presence	180.413	1	180.4130	38.50	<.001	0.681
Residual	84.344	18	4.6858			
set size	35.488	1	35.4882	42.27	<.001	0.701
Residual	15.111	18	0.8395			
repetition	4.688	1	4.6880	11.84	0.003	0.397
Residual	7.127	18	0.3959			
presence ✻ set size	9.707	1	9.7068	29.95	<.001	0.625
Residual	5.835	18	0.3242			
presence ✻ repetition	1.569	1	1.5691	4.80	0.042	0.211
Residual	5.880	18	0.3267			
set size ✻ repetition	0.164	1	0.1640	3.49	0.078	0.162
Residual	0.847	18	0.0471			
presence ✻ set size ✻ repetition	0.281	1	0.2808	3.16	0.092	0.150
Residual	1.597	18	0.0887			

Note. Type 3 Sums of Squares

[4]

Between Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2_p
Residual	218	18	12.1			

Note. Type 3 Sums of Squares

Figure S2. RT analyses for Experiment 2a.

Repeated Measures ANOVA

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2_p
presence	804.1269	1	804.1269	29.50641	<.001	0.596
Residual	545.0524	20	27.2526			
set size	79.7114	1	79.7114	104.57098	<.001	0.839
Residual	15.2454	20	0.7623			
repetition	15.4763	1	15.4763	18.05848	<.001	0.474
Residual	17.1402	20	0.8570			
presence ✻ set size	17.2527	1	17.2527	49.43557	<.001	0.712
Residual	6.9799	20	0.3490			
presence ✻ repetition	5.5312	1	5.5312	9.78441	0.005	0.329
Residual	11.3062	20	0.5653			
set size ✻ repetition	0.0722	1	0.0722	0.47683	0.498	0.023
Residual	3.0292	20	0.1515			
presence ✻ set size ✻ repetition	3.66e–4	1	3.66e–4	0.00385	0.951	0.000
Residual	1.9004	20	0.0950			

Note. Type 3 Sums of Squares

[4]

Between Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2_p
Residual	1035	20	51.8			

Note. Type 3 Sums of Squares

Figure S3. RT analyses for Experiment 2b.

Repeated Measures ANOVA

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2_p
presence	621.88040	1	621.88040	27.76751	<.001	0.607
Residual	403.12746	18	22.39597			
set size	61.92813	1	61.92813	37.81937	<.001	0.678
Residual	29.47448	18	1.63747			
repetition	8.35286	1	8.35286	14.66580	0.001	0.449
Residual	10.25184	18	0.56955			
AI	0.04235	1	0.04235	0.35249	0.560	0.019
Residual	2.16262	18	0.12015			
presence ✻ set size	12.41813	1	12.41813	23.90512	<.001	0.570
Residual	9.35057	18	0.51948			
presence ✻ repetition	4.46105	1	4.46105	11.19134	0.004	0.383
Residual	7.17510	18	0.39862			
set size ✻ repetition	0.12624	1	0.12624	2.34861	0.143	0.115
Residual	0.96752	18	0.05375			
presence ✻ AI	0.12072	1	0.12072	1.61346	0.220	0.082
Residual	1.34674	18	0.07482			
set size ✻ AI	0.03316	1	0.03316	0.33405	0.570	0.018
Residual	1.78679	18	0.09927			
repetition ✻ AI	1.17e−4	1	1.17e−4	0.00185	0.966	0.000
Residual	1.13103	18	0.06284			
presence ✻ set size ✻ repetition	0.00245	1	0.00245	0.05846	0.812	0.003
Residual	0.75459	18	0.04192			
presence ✻ set size ✻ AI	3.50e−4	1	3.50e−4	0.00414	0.949	0.000
Residual	1.51862	18	0.08437			
presence ✻ repetition ✻ AI	0.00544	1	0.00544	0.06387	0.803	0.004
Residual	1.53359	18	0.08520			
set size ✻ repetition ✻ AI	0.00731	1	0.00731	0.11153	0.742	0.006
Residual	1.17957	18	0.06553			
presence ✻ set size ✻ repetition ✻ AI	0.26660	1	0.26660	6.61475	0.019	0.269
Residual	0.72547	18	0.04030			

Note. Type 3 Sums of Squares

Figure S4-1. RT analyses for Experiment 3a (Four-way ANOVA).

Repeated Measures ANOVA

Target present trials

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2_p
Set size	9.44171	1	9.44171	33.7885	< .001	0.652
Residual	5.02983	18	0.27944			
repetition	0.30265	1	0.30265	3.2492	0.088	0.153
Residual	1.67665	18	0.09315			
AI	0.15303	1	0.15303	0.9851	0.334	0.052
Residual	2.79620	18	0.15534			
Set size ✻ repetition	0.08193	1	0.08193	2.3925	0.139	0.117
Residual	0.61644	18	0.03425			
Set size ✻ AI	0.01335	1	0.01335	0.3313	0.572	0.018
Residual	0.72533	18	0.04030			
repetition ✻ AI	0.00198	1	0.00198	0.0281	0.869	0.002
Residual	1.27194	18	0.07066			
Set size ✻ repetition ✻ AI	0.18110	1	0.18110	2.7499	0.115	0.133
Residual	1.18542	18	0.06586			

Note. Type 3 Sums of Squares

[4]

Figure S4-2. RT analyses for Experiment 3a (Three-way ANOVA on target present trials).

Repeated Measures ANOVA

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2_p
set size	64.90455	1	64.90455	34.5695	< .001	0.658
Residual	33.79521	18	1.87751			
repetition	12.51126	1	12.51126	14.2983	0.001	0.443
Residual	15.75029	18	0.87502			
AI	0.01003	1	0.01003	0.2532	0.621	0.014
Residual	0.71315	18	0.03962			
set size ✻ repetition	0.04676	1	0.04676	0.7612	0.394	0.041
Residual	1.10567	18	0.06143			
set size ✻ AI	0.02016	1	0.02016	0.1406	0.712	0.008
Residual	2.58009	18	0.14334			
repetition ✻ AI	0.00358	1	0.00358	0.0462	0.832	0.003
Residual	1.39268	18	0.07737			
set size ✻ repetition ✻ AI	0.09281	1	0.09281	2.3215	0.145	0.114
Residual	0.71962	18	0.03998			

Note. Type 3 Sums of Squares

Figure S4-3. RT analyses for Experiment 3a (Three-way ANOVA on target absent trials).

Repeated Measures ANOVA

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2_p
presence	455.5345	1	455.5345	46.768	<.001	0.722
Residual	175.3243	18	9.7402			
set size	82.3179	1	82.3179	94.733	<.001	0.840
Residual	15.6411	18	0.8690			
repetition	10.1189	1	10.1189	11.527	0.003	0.390
Residual	15.8017	18	0.8779			
AI	1.2726	1	1.2726	11.000	0.004	0.379
Residual	2.0825	18	0.1157			
presence ✻ set size	11.4302	1	11.4302	30.458	<.001	0.629
Residual	6.7550	18	0.3753			
presence ✻ repetition	4.6794	1	4.6794	4.861	0.041	0.213
Residual	17.3273	18	0.9626			
set size ✻ repetition	0.6528	1	0.6528	5.466	0.031	0.233
Residual	2.1498	18	0.1194			
presence ✻ AI	0.2194	1	0.2194	1.291	0.271	0.067
Residual	3.0584	18	0.1699			
set size ✻ AI	0.0487	1	0.0487	0.727	0.405	0.039
Residual	1.2073	18	0.0671			
repetition ✻ AI	0.5656	1	0.5656	2.201	0.155	0.109
Residual	4.6262	18	0.2570			
presence ✻ set size ✻ repetition	0.3099	1	0.3099	3.446	0.080	0.161
Residual	1.6187	18	0.0899			
presence ✻ set size ✻ AI	0.3878	1	0.3878	8.135	0.011	0.311
Residual	0.8581	18	0.0477			
presence ✻ repetition ✻ AI	0.6327	1	0.6327	5.761	0.027	0.242
Residual	1.9771	18	0.1098			
set size ✻ repetition ✻ AI	0.0155	1	0.0155	0.209	0.653	0.012
Residual	1.3355	18	0.0742			
presence ✻ set size ✻ repetition ✻ AI	0.5667	1	0.5667	4.670	0.044	0.206
Residual	2.1845	18	0.1214			

Note. Type 3 Sums of Squares

Figure S5-1. RT analyses for Experiment 3b (Four-way ANOVA).

Repeated Measures ANOVA

target present trials

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2_p
set size	16.1998	1	16.1998	56.89834	<.001	0.760
Residual	5.1249	18	0.2847			
repetition	0.5180	1	0.5180	4.17142	0.056	0.188
Residual	2.2352	18	0.1242			
AI	0.2176	1	0.2176	1.73054	0.205	0.088
Residual	2.2633	18	0.1257			
set size ✻ repetition	0.0316	1	0.0316	0.34680	0.563	0.019
Residual	1.6386	18	0.0910			
set size ✻ AI	0.3557	1	0.3557	6.54527	0.020	0.267
Residual	0.9783	18	0.0544			
repetition ✻ AI	9.42e−4	1	9.42e−4	0.00754	0.932	0.000
Residual	2.2490	18	0.1249			
set size ✻ repetition ✻ AI	0.3850	1	0.3850	4.32802	0.052	0.194
Residual	1.6012	18	0.0890			

Note. Type 3 Sums of Squares

Figure S5-2. RT analyses for Experiment 3b (Three-way ANOVA on target present trials).

Repeated Measures ANOVA

target absent trials

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2_p
set size	77.5484	1	77.5484	80.82	<.001	0.818
Residual	17.2712	18	0.9595			
repetition	14.2804	1	14.2804	8.32	0.010	0.316
Residual	30.8937	18	1.7163			
AI	1.2744	1	1.2744	7.97	0.011	0.307
Residual	2.8776	18	0.1599			
set size ✻ repetition	0.9312	1	0.9312	7.87	0.012	0.304
Residual	2.1298	18	0.1183			
set size ✻ AI	0.0808	1	0.0808	1.34	0.263	0.069
Residual	1.0870	18	0.0604			
repetition ✻ AI	1.1974	1	1.1974	4.95	0.039	0.216
Residual	4.3543	18	0.2419			
set size ✻ repetition ✻ AI	0.1973	1	0.1973	1.85	0.190	0.093
Residual	1.9189	18	0.1066			

Note. Type 3 Sums of Squares

Figure S5-3. RT analyses for Experiment 3b (Three-way ANOVA on target absent trials).

Repeated Measures ANOVA

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2_p
presence	515.5392	1	515.5392	104.4692	<.001	0.853
Residual	88.8272	18	4.9348			
set size	123.4782	1	123.4782	116.4938	<.001	0.866
Residual	19.0792	18	1.0600			
repetition	0.0194	1	0.0194	0.0149	0.904	0.001
Residual	23.4583	18	1.3032			
AI	19.0156	1	19.0156	71.5293	<.001	0.799
Residual	4.7852	18	0.2658			
presence ✻ set size	19.9857	1	19.9857	75.2525	<.001	0.807
Residual	4.7805	18	0.2656			
presence ✻ repetition	0.0132	1	0.0132	0.0242	0.878	0.001
Residual	9.7930	18	0.5441			
set size ✻ repetition	1.04e−4	1	1.04e−4	6.88e−4	0.979	0.000
Residual	2.7347	18	0.1519			
presence ✻ AI	3.6547	1	3.6547	8.9853	0.008	0.333
Residual	7.3213	18	0.4067			
set size ✻ AI	2.6916	1	2.6916	13.6374	0.002	0.431
Residual	3.5526	18	0.1974			
repetition ✻ AI	12.1379	1	12.1379	31.3947	<.001	0.636
Residual	6.9592	18	0.3866			
presence ✻ set size ✻ repetition	0.7404	1	0.7404	4.1082	0.058	0.186
Residual	3.2439	18	0.1802			
presence ✻ set size ✻ AI	0.0648	1	0.0648	0.8539	0.368	0.045
Residual	1.3654	18	0.0759			
presence ✻ repetition ✻ AI	2.3038	1	2.3038	17.1272	<.001	0.488
Residual	2.4212	18	0.1345			
set size ✻ repetition ✻ AI	1.1236	1	1.1236	8.5038	0.009	0.321
Residual	2.3783	18	0.1321			
presence ✻ set size ✻ repetition ✻ AI	0.7341	1	0.7341	11.5278	0.003	0.390
Residual	1.1463	18	0.0637			

Note. Type 3 Sums of Squares

Figure S6-1. RT analyses for Experiment 3c (Four-way ANOVA).

Repeated Measures ANOVA

target present trials

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2_p
set size	22.0550	1	22.0550	52.77493	<.001	0.746
Residual	7.5223	18	0.4179			
repetition	3.00e−4	1	3.00e−4	0.00208	0.964	0.000
Residual	2.6047	18	0.1447			
AI	2.9987	1	2.9987	8.94384	0.008	0.332
Residual	6.0351	18	0.3353			
set size ✻ repetition	0.3614	1	0.3614	2.25583	0.150	0.111
Residual	2.8841	18	0.1602			
set size ✻ AI	0.9606	1	0.9606	5.65060	0.029	0.239
Residual	3.0601	18	0.1700			
repetition ✻ AI	1.9328	1	1.9328	7.79573	0.012	0.302
Residual	4.4628	18	0.2479			
set size ✻ repetition ✻ AI	0.0206	1	0.0206	0.27004	0.610	0.015
Residual	1.3759	18	0.0764			

Note. Type 3 Sums of Squares

Figure S6-2. RT analyses for Experiment 3c (Three-way ANOVA on target present trials).

Repeated Measures ANOVA

target absent trials

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2_p
set size	121.4089	1	121.4089	133.7647	<.001	0.881
Residual	16.3374	18	0.9076			
repetition	0.0322	1	0.0322	0.0189	0.892	0.001
Residual	30.6466	18	1.7026			
AI	19.6715	1	19.6715	58.3204	<.001	0.764
Residual	6.0714	18	0.3373			
set size ✻ repetition	0.3790	1	0.3790	2.2047	0.155	0.109
Residual	3.0946	18	0.1719			
set size ✻ AI	1.7957	1	1.7957	17.3977	<.001	0.491
Residual	1.8579	18	0.1032			
repetition ✻ AI	12.5088	1	12.5088	45.7868	<.001	0.718
Residual	4.9175	18	0.2732			
set size ✻ repetition ✻ AI	1.8371	1	1.8371	15.3892	<.001	0.461
Residual	2.1488	18	0.1194			

Note. Type 3 Sums of Squares

Figure S6-3. RT analyses for Experiment 3c (Three-way ANOVA on target absent trials).