



## BIROn - Birkbeck Institutional Research Online

Rangelov, D. and Muller, Hermann J. and Zehetleitner, M. (2017) Failure to pop out: feature singletons do not capture attention under low signal-to-noise ratio conditions. *Journal of Experimental Psychology: General* 146 (5), pp. 651-671. ISSN 0096-3445.

Downloaded from: <https://eprints.bbk.ac.uk/id/eprint/21565/>

*Usage Guidelines:*

Please refer to usage guidelines at <https://eprints.bbk.ac.uk/policies.html> or alternatively contact [lib-eprints@bbk.ac.uk](mailto:lib-eprints@bbk.ac.uk).

Short title: Feature singletons fail to capture attention

## **Failure to pop out: feature singletons do not capture attention under low signal-to-noise ratio conditions**

Dragan Rangelov<sup>1</sup>, Hermann J. Müller<sup>2,3</sup>, and Michael Zehetleitner<sup>4</sup>

<sup>1</sup>Queensland Brain Institute, University of Queensland Brisbane

<sup>2</sup>Ludwig-Maximilians-Universität München

<sup>3</sup>Birkbeck College, University of London

<sup>4</sup>Catholic University of Eichstätt-Ingolstadt

Contact:

Dragan Rangelov

Queensland Brain Institute

The University of Queensland

QBI Building 79

St Lucia, QLD 4072, Australia

Email: [d.rangelov@uq.edu.au](mailto:d.rangelov@uq.edu.au)

## **Abstract**

Pop-out search implies that the target is always the first item selected, no matter how many distractors are presented. Increasing evidence, however, indicates that search is not entirely independent of display density even for pop-out targets: search is slower with sparse (few distractors) than with dense displays (many distractors). Despite its significance, the cause of this anomaly remains unclear. We investigated several mechanisms that could slow down search for pop-out targets. Consistent with the assumption that pop-out targets frequently fail to pop out in sparse displays, we observed greater variability of search duration for sparse displays relative to dense. Computational modeling of the response time distributions also supported the view that pop-out targets fail to pop out in sparse displays. Our findings strongly question the classical assumption that early processing of pop-out targets is independent of the distractors. Rather, the density of distractors critically influences whether or not a stimulus pops out. These results call for new, more reliable measures of pop-out search and potentially a re-interpretation of studies that used relatively sparse displays.

## **Keywords**

Attentional capture, efficient visual search, distribution analyses, mixture models

## Introduction

To paraphrase William James' (1891) definition of attention: "Everyone knows what pop-out is". A pop-out stimulus is, by definition, always the first item to be selected, as if it captured attention automatically (White, Lunau, & Carrasco, 2014). Investigating which stimuli do and do not pop out reveals mechanisms of early, pre-attentive vision which, together with effortful, top-down mechanisms, form a cornerstone of most contemporary accounts of visual attention (e.g., Duncan & Humphreys, 1989; Treisman & Gelade, 1980; Wolfe, 2007; for more specific studies of pop-out, see, e.g., (e.g., Eimer & Kiss, 2008; Folk & Remington, 2006; Vera Maljkovic & Nakayama, 1994; Müller, Heller, & Ziegler, 1995; Treisman & Gelade, 1980; Wolfe & Horowitz, 2004). Yet, despite the fact that there is a large and growing body of work on 'pop-out', we will argue here that many paradigms that were designed to elucidate the pop-out mechanisms may actually not have investigated pop-out at all.

Whether or not a target pops out is typically investigated by inspecting how reaction times (RTs) change with adding more distractors to the search display. Failure to pop out predicts RTs to become slower as more distractors are added, which is consistent with a positive slope of the RT/display density function. If the target, on the other hand, does pop out, then adding more distractors would matter little, predicting RTs to remain constant across varying numbers of items in the display (e.g., Wolfe, 2007). Put differently, 'pop-out' is consistent with flat RT/display density functions (though see Buetti, Cronin, Madison, Wang, & Lleras, 2016, who reported positive search slopes for, presumably, pop-out search). A stimulus property typically considered to generate pop-out is an item's uniqueness. An item that differs from the other, homogeneous items in the display by one or multiple features – a 'feature singleton' – is thought to attract attention automatically (Theeuwes, 1992, 2010; Treisman & Gelade, 1980; Wolfe & Horowitz, 2004). A prominent notion, common to a host of visual search theories, is that early, pre-attentive stimulus coding mechanisms compute and integrate local 'feature contrasts' between an item and its surround; the item achieving the highest total contrast would win the competition for selection (e.g., Itti & Koch, 2000, 2001). Since feature singletons typically produce strong contrasts, such targets are always selected first no matter how many distractors are present, yielding flat search slopes.

A growing number of studies, however, show that RTs to singleton targets are not entirely independent of the number of distractors. Rather, many studies reported *negative search slopes*: an increase in display density actually decreases RTs (Bravo & Nakayama, 1992; Buetti et al., 2016; Geyer, Zehetleitner, & Müller, 2010; Kristjánsson, 2006; Lamy, Zivony, & Yashar, 2011; Meeter & Olivers, 2006; Rangelov, Müller, & Zehetleitner, 2013). Despite its significance for theories of early vision, the precise cause of this anomaly remains unclear. The aim of the present study was to evaluate several mechanisms that can, in principle, generate negative search slopes (Fig. 1). Two of these, henceforth referred to as 'slow pop-out' and 'different pop-out' models, maintain that finding a feature singleton target is simply slower in sparse than in dense displays, with the negative search slopes being attributable to *quantitative differences* in finding the target in displays of varying item density. An alternative, 'no pop-out' model maintains that finding a singleton target in sparse displays is *qualitatively different* from finding one in dense displays: the target frequently fails to pop out in sparse displays.

### *Models of negative search slopes*

Here, we present four models that can account for negative search slopes (Fig. 1). We assume that RT on any trial reflects the durations of two processes: (i) *the search processes*, or the time it takes to select an item from the display and determine whether or not it is the target, and (ii) *the post-search processes*, or the time it takes to process target features, choose a response, and execute it. All four models assume that the negative search slopes originate from the prolonged search duration in sparse relative to dense displays. Searching for the target is assumed to be an iterative process in which the visual system computes feature contrasts between each element in the display and its surround. The location with the strongest overall feature contrast signal is then selected for further inspection. If the item at the selected location is not the target, this location is discarded and a new location is selected (Koch & Ullman, 1985; Wolfe, 2007). In this framework, the pop-out stimuli would be items that are always selected in the first selection round.

*Figure 1 about here*

The simplest, '*slow pop-out*' model of negative search slopes assumes that it takes longer to generate a feature contrast signal of sufficient strength in sparse displays than in dense. Critically, the output of the contrast computations yields reliably stronger signals for the target than the distractors, thus permitting pop-out search. Since contrast computation is simply slower for sparse displays relative to dense whereas the output remains unaltered, the slow pop-out model predicts a simple shift in mean search duration between dense and sparse displays.

The '*different pop-out*' model, by contrast, assumes that the contrast computation itself differs between sparse and dense displays. These differences would alter the relation between the target and distractor signals. Due to iso-feature suppression mechanisms, by which detectors coding similar stimuli inhibit each other (e.g., Li, 1999, 2002), the many, closely spaced distractors in dense displays would each yield weaker signals relative to the few and widely spaced distractors in sparse displays. Thus, in sparse displays, the singleton target would compete with strong distractor signals, whereas, in dense displays, the same target would compete with relatively weak distractor signals. The difference among the contrast signals generated by the target and distractors, respectively, can be conceived of as a signal-to-noise ratio (SNR) – which would be higher for dense relative to sparse displays. The high SNR in dense displays would speed up search processes relative to sparse displays, yielding negative search slopes. Additionally, because of the low SNR, finding the target in sparse displays would be more influenced by random noise inherent in the contrast computation processes (Itti & Koch, 2000, 2001), predicting greater variability of search duration for sparse relative to dense displays. A variant of the "different pop-out" model has recently been proposed by Buetti et al. (2016). They assumed that, when the target and distractor colors are not known in advance, the early, pre-attentive processes are devoted to determining distractor colors. The distractor locations can be summarily rejected as soon as the distractor color has been determined (by finding two items of the same color), leaving only the target location for focal-attentional inspection. Simple statistical facilitation predicts that this target/distractor color assignment would operate faster for dense displays than for sparse, yielding negative search

slopes. Thus, while early processing differs between sparse and dense displays, the output of this stage would permit pop-out independently of display density.

The ‘no pop-out’ model assumes that display items are selected one after the other at a constant rate (e.g., one item every 100 ms) until the target is found. If the target does pop out, then it will always be the first item to be selected. Critically, the differences between sparse and dense displays would influence the probability of finding the target in the first run. The ‘no pop-out’ model assumes that, similar to the ‘different pop-out’ model, the SNR would be lower for sparse than for dense displays. Due to the low SNR for sparse displays, finding the target in the first round of selection would frequently fail, making it necessary to search for the target several times until it is found. By contrast, the high SNR for dense displays would permit finding the target invariably in the first round of selection. Consequently, the probability of finding the target in the first run ( $p_1$ ) would be lower for sparse ( $p_1 \ll 1$ ) than for dense displays ( $p_1 \approx 1$ ). This would prolong the mean search duration and increase its variability for sparse relative to dense displays, yielding negative search slopes<sup>1</sup>. Note that the ‘no pop-out’ model predicts a multi-modal distribution of search duration when the target frequently fails to pop out, and a unimodal distribution for targets that pop out reliably (Fig. 1). In contrast to the other two models that postulate quantitative differences in search duration between sparse and dense displays (e.g., location shifts, increase in variance), the ‘no pop-out’ model predicts the distributions of search duration to differ qualitatively between sparse and dense displays, respectively.

Finally, since the assumptions of the ‘different pop-out’ and the ‘no pop-out’ model are not mutually exclusive, it is possible that both the contrast computation and the probability of selecting the target in the first selection round differ between sparse and dense displays. This hybrid model (not shown on Figure 1) is a generalization of the two more restricted models (the ‘different’ and ‘no pop-out’ models) and would be compatible with any data that either model can account for. Henceforth, this model will be referred to as the ‘*hybrid model*’ of negative search slopes.

Of note, all discussed models can account for differences in the *mean search duration* between sparse and dense displays. The slow pop-out model, however, predicts no differences between sparse and dense displays in search duration *variability*. By contrast, the other models predict variability to be larger for sparse displays than for dense. Thus, comparing search variability between sparse and dense displays would help discriminate between the slow pop-out and the other models. Because the other three models all predict both a higher mean and greater variability of the search duration for sparse displays relative to dense, simple inspection of mean search duration and search variability would not suffice to distinguish among these models. Instead, the different pop-out, no-pop out, and hybrid models need to be compared in terms of their quantitative fits to the observed data to permit the models to be discriminated.

---

<sup>1</sup> While the ‘no pop-out’ model predicts a multi-modal distribution of search duration (as illustrated in Figure 1, top-right panel), this does not necessarily predict a multimodal distribution of the total RTs (Figure 1, bottom-right). The probability of finding the target on any single selection round decreases the longer the search takes, as a result of which multiple modi in the search time distribution would become progressively smaller (Figure 1). This might render the modi undetectable in the RT distribution.

As shown in Figure 1, total RTs are a combination of at least two distinct, search and post-search, processes. The fact that total RTs reflect a mixture of different (e.g., search vs. post-search) processes makes attributing RT effects to any single process equivocal. Larger RT means and variability for sparse displays than for dense may reflect differences in the search duration, the post-search duration, or both. To independently assess the contributions of search and post-search durations to RTs, we minimized the search duration by pre-cuing the target location with 100% cue validity. Since, on cued trials, there was no need to search for the target, the RTs would primarily reflect the post-search duration. Consequently, any RT effects present on uncued trials and absent on cued trials would most likely stem from differences in search duration. With this experimental design, it became possible to assess whether negative search slopes stem from search or post-search processes.

### *The role of task requirements and stimulus features*

Whether or not negative search slopes for feature singleton targets are observed seems to depend heavily on the task participants perform. Studies using *feature discrimination tasks*, in which participants had to discriminate the precise target features, reliably reported substantial negative search slopes (e.g., Becker & Ansorge, 2013, Experiment 1; Bravo & Nakayama, 1992; Lamy et al., 2011; Meeter & Olivers, 2006; Rangelov et al., 2013). By contrast, the few studies that used *detection tasks*, in which participants had to indicate whether or not the target was present, typically failed to observe negative search slopes (Bravo & Nakayama, 1992; Wan & Lleras, 2010). One notable difference between discrimination and detection tasks concerns the role of spatial attention. While attending to a narrow area occupied by target is necessary for discrimination tasks, the narrow focus of spatial attention may not be necessary to simply detect the target (Chan & Hayward, 2009; Müller & Krummenacher, 2006a, 2006b). Consequently, comparing performance between discrimination and detection tasks may reveal the role of focal spatial attention in generating negative search slopes.

Furthermore, there is ample evidence that stimulus features, both those of the target and those of the distractors, influence how quickly the target is found. For example, repeating the exact target and distractor features across consecutive trials (e.g., a red target and green distractors both on trials  $n-1$  and  $n$ ) reliably speeds up mean RTs on trial  $n$  relative to swapping the target and distractor features – an effect referred to as ‘priming of pop-out’ (PoP; Lamy, Antebi, Aviani, & Carmel, 2008; Lamy & Yashar, 2011; Maljkovic & Martini, 2005; Maljkovic & Nakayama, 1994). Other studies showed that repeating distractor features yields similar intertrial effects: manual responses and eye movements to the target are faster and more accurate when the distractor feature is known relative to when it varies randomly, suggesting that memory for distractor features may influence visual search (Lamy et al., 2008; Lamy & Yashar, 2011; Lleras, Kawahara, Wan, & Ariga, 2008).

Recent studies have shown that the effect of stimulus features co-varies with display density. In particular, the precise target features matter much more with sparse displays than with dense (Lamy et al., 2011; Olivers & Meeter, 2006; Rangelov et al., 2013). Importantly, this difference cannot be explained by simple shifts in overall response speed. Additionally, knowing distractor features in advance also seems to matter more for sparse displays than for dense. For sparse displays, fixed

distractor color produces weaker PoP relative to variable distractor colors, whereas, for dense displays, distractor variability matters little for the PoP magnitude (Rangelov et al., 2013). In the present study, we investigated whether differential effects of stimulus features between sparse and dense displays reflect overall shifts in RT distributions or changes in the shape of the distributions. Strong changes of the RT distributions would support the notion that finding the target differs qualitatively between sparse and dense displays.

### *Aims of the study*

In summary, the present study was designed to investigate which cognitive mechanism yields the negative search slopes for feature singleton targets by examining the RT distributions for sparse and dense displays. To investigate the role of task requirements for negative search slopes, a target discrimination task was used in Experiment 1, and a target detection task in Experiment 2. Additionally, the role of stimulus features was assessed by varying whether or not these properties were known in advance. Experiments 3 and 4 went on to investigate the extent to which the findings of Experiments 1 and 2 can be generalized to: (i) a greater variety of display densities (sparse, medium, and dense) and (ii) stimulus properties other than surface color, in particular: contour color and orientation. Data analyses focused on comparing the observed RTs with the predictions of the various models of negative search slopes (i.e., the ‘different pop-out’, ‘no pop-out’, and ‘hybrid’ models; Fig. 1).

## **Method**

Participants. 16 observers (5 male, mean age 25 years) participated in Experiment 1, another 16 (6 male, mean age 25) in Experiment 2, 17 observers (8 males, mean age 26 years) in Experiment 3, and 16 observers (8 male, mean age 29) in Experiment 4 for either monetary compensation or course credits. All had normal or corrected-to-normal visual acuity and normal color vision (confirmed using the Ishihara color plates). All were naïve as to the purpose of the study, yet all had extensive experience with psychophysical experiments. The study was approved by the LMU Psychology ethics committee. All participants provided informed consent in accordance with the WMA Declaration of Helsinki.

Apparatus. The experiment was controlled by a Dell PC running under Windows XP. Stimuli were presented on a 19” CRT monitor (screen resolution: 1024x768 pixels; refresh rate: 85 Hz). The experimental software was custom-written in PsychoPy (Peirce, 2007, 2009). Head-to-monitor distance was 56 cm, controlled by means of a chin rest. Participants responded by pressing the left or right mouse button with their left- or right-hand index finger, respectively.

Procedure. The experiments were run in a dim, sound-attenuated cabin. Figure 2A illustrates the experimental procedure. Each trial started with a fixation display for 1 s, followed by, in half the trials (randomly selected), a spatial cue (30 ms) marking the location of the upcoming target. In the other half, no cue was presented. During the cue-stimulus interval (150 ms), only the fixation marker was shown. Next, search arrays were presented until either response or for maximally 1500 ms. Both response speed and accuracy were stressed. In case of a response error, the word “Error!” was presented for 1000 ms.

### *Experiments 1 and 2*

In Experiment 1 (discrimination task), the search arrays always comprised several distractors and one color singleton target. Individual items were diamond-shaped elements (Fig. 2A, lower-right) with a small part on either the left or the right side of every diamond missing. Participants had to find the target and report the missing side (left or right) by pressing the corresponding response key (left or right). Participants completed 1440 trials, in four sessions separated by short breaks, which took about 2 h to complete. In Experiment 2 (detection task), search arrays comparable to those of Experiment 1 were used on two-thirds of the trials. On the remaining trials, the search displays consisted of distractors only. Participants had to report whether or not a target was present. Participants completed 2160 trials (720 target-absent and 1440 target-present trials) which, by virtue of the target detection task being easier than the discrimination task, also took some two hours to complete.

Stimuli. A white square contour (1-pixel line thickness, CIE xy .005, -.010, luminance 100 cd/m<sup>2</sup>), subtending 1.5° x 1.5° of visual angle, served as the cue stimulus. The target was always either a single red (CIE xy .255, .071) or a single green (.358, .021) item (Fig. 2). The target color was chosen randomly on each trial. In one experimental session, the distractors' color could vary: if the target was red, distractors were green, and vice versa. In a different session, the distractor color was fixed across trials: it was always blue (.387, .133). Stimulus luminance was 51 cd/m<sup>2</sup> for all three colors. Individual items subtended 1.3° x 1.3° of visual angle. Search arrays consisted of individual items arranged around three (imaginary) concentric circles with radii of 2.5°, 5.0°, and 7.5°, respectively (maximum number of items presented on the inner, middle, and outer circles: 6, 12, and 18, respectively). The target always appeared at one, randomly chosen position on the middle circle. Two different display densities were used, fixed per session: sparse (3 items) and dense (36). In sparse displays, the two distractors were also located on the middle circle, at positions equidistant from the target.

*Figure 2 about here*

Design. Factorial combination of two levels of display density (sparse vs. dense) and distractor variability (variable vs. fixed) yielded four experimental conditions, which were performed in separate sessions. Within each session, trials were either uncued or cued, yielding a 2x2x2 design (variability, density, and cue), with 180 trials per experimental condition in Experiment 1. In Experiment 2 there were 180 trials per experimental condition for target present trials and 90 trials for target absent trials.

### *Experiments 3 and 4*

Experiments 3 and 4 were designed to test whether the findings of Experiments 1 would generalize to a greater variety of display densities and stimulus properties. Individual items were either circle contours (Exp. 3) or rectangle contours (Exp. 4) with a letter 'E' inside (Fig. 2B). Color singletons were used as targets in Experiment 3, and orientation singletons in Experiment 4. In both experiments, participants had to find the target singleton and report the orientation of the letter 'E' inside the target: E versus mirror E, by pressing the corresponding response key (discrimination task). Participants performed 1440

trials per experiment, in three sessions separated by short breaks; each experiment took about 2 h to complete.

**Stimuli.** In Experiment 3, the target was always either a single red (CIE xy .255, .071) or a single green (.358, .021) circle contour (Fig. 2B). The target color was chosen randomly on each trial. In Experiment 4, the target was always either a single left- or a single right-tilted (20 degrees from vertical) blue rectangle (.387, .133), with target orientation varying randomly across trials. In both Experiments 3 and 4, only the variable-distractors condition was administered. Similar to Experiments 1 and 2, search arrays comprised items arranged around three (imaginary) concentric circles. Three different display densities were used, fixed per experimental session: sparse (6 items), medium (12), and dense (36). In sparse and medium density displays, the stimuli were arranged around the middle circle; for dense displays, all locations were used (Fig. 2B). All other details were the same as in Experiments 1 and 2.

**Design.** Factorial combination of three levels of display density (sparse, medium, and dense) and cue (cued vs. uncued) yielded six experimental conditions, with 240 trials per condition.

## Results

### Experiment 1

Inspection of response accuracy across experimental conditions in Experiment 1 (Table 1) indicated that responses were quite accurate overall (4% of errors). A log-linear model of the error rates with display density, distractor variability, and cue as fixed factors and participant as a random factor revealed that participants made significantly more errors on uncued trials with sparse displays and variable distractors (6.5%,  $\beta = .73$ ,  $SE_{\beta} = .28$ ,  $z = 2.60$ ,  $p < .01$ ) relative to all other conditions.

Table 1. Response accuracy (mean error percentages and 95% CI) for Experiments 1–4, across the different experimental conditions.

Distractors	Density	Exp. 1	Exp. 2		Exp. 3	Exp. 4
			Target present	Target absent		
<i>Cued trials</i>						
Variable	Sparse	2.52(0.74)	2.22(0.93)	10.43(2.70)	1.71(0.74)	3.98(2.56)
	Medium				1.63(0.70)	2.85(1.14)
	Dense	3.43(1.22)	1.91(1.51)	8.62(3.88)	1.50(0.56)	2.95(1.04)
Fixed	Sparse	3.66(1.15)	2.06(0.89)	8.10(2.54)		
	Medium					
	Dense	3.85(1.25)	2.13(0.86)	7.87(2.16)		
<i>Uncued trials</i>						
Variable	Sparse	6.48(1.50)	2.83(1.25)	6.70(2.84)	3.95(1.34)	11.88(2.01)
	Medium				3.16(0.99)	9.55(2.32)
	Dense	3.66(0.96)	3.64(1.72)	5.51(2.77)	2.96(0.95)	5.67(0.88)
Fixed	Sparse	3.91(1.06)	1.85(0.69)	4.54(1.50)		

Medium			
Dense	3.53(0.83)	3.39(1.36)	5.02(2.24)

Correct trials with RTs beyond 1.5\*inter-quartile range per participant and condition ('outliers') were excluded from further analyses ( $\approx$  5% of the correct trials). Next, RT distributions were assessed by computing participant-specific RTs across different cumulative probabilities (.1–.9 in steps of .1), separately for each experimental condition (cue, display density, distractor variability, and target sequence). Figure 3 shows the distribution of RTs for a representative participant as well as average RT distributions for the different conditions of Experiment 1 (target discrimination).

*Figure 3 about here*

Analyses of median RTs. Median RTs were overall faster on cued (418 ms) than on uncued trials (524 ms). On *uncued trials*, RTs were slower to sparse than to dense displays (552 ms vs. 496 ms), that is, negative search slopes were observed. On *cued trials*, by contrast, sparse displays were responded to faster than dense displays (405 ms vs. 431 ms). While RTs were overall longer in the variable-distractors relative to the fixed-distractors condition, this difference was more prominent on uncued trials (549 ms vs. 499 ms) than on cued trials (425 ms vs. 411 ms). Priming of pop-out (PoP) was substantial only for sparse displays with variable distractors (24 ms), relative to all other conditions ( $\approx$  2 ms).

To test these observations, two separate general linear model (GLM) analyses were ran with RTs and PoP as dependent variables, display density, distractor variability, and cue as categorical predictors, and participant as a random factor. Inspection of the model coefficients for RTs revealed significant effects of cueing ( $\beta = 63.71$ ,  $SE_{\beta} = 8.20$ ,  $t = 7.76$ ,  $p < .001$ ) and display density ( $\beta = -31.38$ ,  $SE_{\beta} = 8.20$ ,  $t = -3.82$ ,  $p < .001$ ). The difference between dense and sparse displays was substantial and *negative* on uncued trials ( $\beta = -50.94$ ,  $SE_{\beta} = 11.60$ ,  $t = -4.39$ ,  $p < .001$ ); by contrast, a somewhat weaker and *positive* difference was observed on cued trials ( $\beta = 31.38$ ,  $SE_{\beta} = 8.20$ ,  $t = 3.82$ ,  $p < .001$ ). Finally, distractor variability mattered only for uncued trials with sparse displays ( $\beta = 64.40$ ,  $SE_{\beta} = 16.41$ ,  $t = 3.99$ ,  $p < .001$ ). Inspection of model coefficients for PoP showed significant priming effect only for uncued trials with sparse displays and variable distractors ( $\beta = 38$ ,  $SE_{\beta} = 12.80$ ,  $t = 2.97$ ,  $p < .01$ ). Taken together, analyses of the median RTs in Experiment 1 showed substantial negative search slopes on uncued trials, which were stronger with variable than with fixed distractors. Furthermore, PoP in terms of median RTs was specific to sparse displays with variable distractors.

Analyses of RT distributions. As is suggested by the broader RT distributions for sparse than for dense displays (Fig. 3), not just the median RTs differed between display densities but the whole RT distributions varied. To examine whether the negative search slopes for the median RTs were due to simple distribution shifts or changes in distribution shapes (or both), the whole RT distributions were analyzed using a GLM with display density, distractor variability, and cue as categorical predictors, cumulative probability as a continuous predictor, and participant as a random factor. Main effects of categorical predictors would reflect differences in the *fastest* RTs, indicative of *shifts* of the whole RT distribution. An effect of cumulative probability would correspond to changes in RT variability, whereas interactions of probability with categorical predictors would be indicative of different experimental

conditions yielding different RT variabilities. For *cued trials*, the fastest RTs were faster for sparse than for dense displays ( $\beta = -25.18$ ,  $SE_{\beta} = 8.69$ ,  $t = -2.90$ ,  $p < .01$ ). RT variability did not change substantially across experimental conditions, as indicated by the absence of significant interactions between cumulative probability and categorical predictors (all  $t < 1.35$ , all  $p$ 's  $> .18$ ). For *uncued trials*, there was no difference between sparse and dense displays in the fastest RTs for fixed distractors ( $t < 1$ ,  $p = .64$ ); for variable distractors, by contrast, sparse displays proved significantly slower than dense displays ( $\beta = 28.49$ ,  $SE_{\beta} = 8.66$ ,  $t = 3.29$ ,  $p < .01$ ). RT variability on uncued trials was larger for sparse than for dense displays, with both fixed distractors ( $\beta = 5.91$ ,  $SE_{\beta} = 1.09$ ,  $t = 5.43$ ,  $p < .01$ ) and variable distractors ( $\beta = 11.09$ ,  $SE_{\beta} = 1.54$ ,  $t = 7.21$ ,  $p < .01$ ). Thus, the analyses of the whole RT distributions showed that differences between sparse and dense displays on *cued trials* originated primarily from distribution shifts, whereas the differences on *uncued trials* were primarily due to increased RT variability for sparse displays.

To investigate how PoP magnitude varied with response speed, a GLM of the PoP effect was run with cue, display density, and distractor variability as categorical predictors and the cumulative probability as a continuous predictor. Similar to the PoP analyses for median RTs, this GLM revealed substantial PoP only for uncued trials with sparse displays and variable distractors: the PoP effect was significant for the fastest RTs ( $\beta = 42.95$ ,  $SE_{\beta} = 11.52$ ,  $t = 3.72$ ,  $p < .01$ ) and stronger for slower RTs ( $\beta = 4.04$ ,  $SE_{\beta} = 1.45$ ,  $t = 2.79$ ,  $p < .01$ ).

#### *Modeling negative search slopes*

As discussed earlier, a general increase in search duration (Fig. 1) for sparse displays cannot explain the differences in search variability between sparse and dense displays. Experiment 1 showed that display density effects on RT variability were specific to uncued trials, indicating that search variability varied substantially between sparse and dense displays. Thus, the results of Experiment 1 falsify the 'slow pop-out' account of negative search slopes. To investigate how well the other models – the 'different pop-out', the 'no pop-out', and the 'hybrid model' – can account for the effects on RT variability observed in Experiment 1, a model of response speed was developed (Eq. 1–4).

Model of RT distributions. The model presented here is an adaptation of the Competitive Guided Search model (Moran, Zehetleitner, Müller, & Usher, 2013) modified to account for RTs in discrimination tasks in which a target is present on every trial. Response time on any single trial (i.e.,  $RT_i$  for trial  $i$ ) can be coarsely divided into two components: (i) the time it takes to select an item and verify that the selected item is indeed the target ( $ST_i$ ), and (ii) processing of the response-relevant stimulus attribute and response selection ( $PT_i$ ). For brevity, the former component will be referred to as the search duration and the latter as the post-search duration. In contrast to Moran et al. (2013) who used the ex-Wald distribution to model search duration, in the present study, the search and post-search durations were modeled as gamma distributions ( $\Gamma$ ) defined by two, *shape* ( $k$ ) and *scale* ( $\theta$ ), parameters. Beyond being more tractable computationally, the gamma distribution was chosen because it describes the distribution of waiting times until an event occurs, that is, until a target or a response is selected. Convolving two gamma distributions (as in the present model where search and post-search durations

are combined to produce RTs) yields a gamma distribution, which is consistent with the typical shape of empirical RT distributions (Palmer, Horowitz, Torralba, & Wolfe, 2011; Ratcliff, 1979).

$$RT_i = ST_i + PT_i \quad \text{Eq. 1}$$

$$\Gamma(k_{RT}, \theta_{RT}) = \left[ \sum_{j=1}^N p_j \Gamma(k_{ST}, \theta_{ST}) \right] * \Gamma(k_{PT}, \theta_{PT}) \quad \text{Eq. 2}$$

$$p_j = \left(1 - \sum_{m=1}^{j-1} p_m\right) \frac{\omega_T}{\omega_T + \sum_{n=1}^{N-j} \omega_D} \quad \text{Eq. 3}$$

$$\omega_D = 1 \xrightarrow{\text{yields}} p_j = \left(1 - \sum_{m=1}^{j-1} p_m\right) \frac{\omega_T}{\omega_T + N - j} \quad \text{Eq. 4}$$

Search duration was modeled as a sample from the  $\Gamma(k_{ST}, \theta_{ST})$  distribution (Eq. 2). Since the target may be found in the first, the second, or any subsequent sample, the total search time was expressed as the sum of all selection rounds prior to finding the target. Importantly, each summand  $j$  was weighted by the probability of finding the target in the  $j$ -th round ( $p_j$ ). Differences in  $p_j$  can capture the difference between pop-out and no-pop-out searches. Assuming pop-out search implies that the target is found immediately, from which it follows that  $p_1 \approx 1$ , whereas  $p_{j>1} \approx 0$ . If the target does not pop out, it is found only after several selection rounds, from which it follows that  $p_1 \ll 1$ , whereas  $p_{j>1} \gg 0$ .

The  $p_j$  values were modeled (Eq. 3) as a function of target conspicuity (or target weight  $\omega_T$ ) relative to distractor conspicuity ( $\omega_D$ ), using a modified version of Luce's choice rule (Luce, 1977). The probability of selecting the target was the ratio between the target weight and the sum of the target and distractor weights (i.e., the total weight of all items in display). We assumed perfect memory for previously selected locations, so that the probability of selecting the target in the  $j$ -th round was weighed by the probability that the target was *not* selected in the previous rounds. This assumption has no bearing on computations if the target is found immediately (i.e.,  $p_1 \gg p_{j>1}$ ), but it was useful in case the target does not pop out. We set out to assess whether or not, even when  $p_1 < 1$ , the target would still be preferentially selected in the first couple of rounds, that is,  $p_{1,2,3} > p_{4,5,6}$ , which was possible to compute on the assumption that a specific location was selected only once. Assuming no memory for past selections in case of pop-out failure would result in identical  $p_j$  for all  $j$ 's, making it impossible to assess the probability of selecting the target in the first couple of rounds. To simplify computations, the distractor weight ( $\omega_D$ ) was set to 1 for all distractors, which reduced Eq. 3 to Eq. 4. Thus, the probability of selecting the target in  $j$ -th selection round was the function of only one parameter: the target weight ( $\omega_T$ ). Finally, because the search component of the model captured the time until the target was correctly selected and verified, the post-search times were modeled as a single sample of the  $\Gamma(k_{PT}, \theta_{PT})$ .

For model fitting, the RT distribution was described as quantile RTs ( $p = \{.1, .3, .5, .7, .9\}$ ), which were calculated for each participant and then averaged across participants (vincentizing, Ratcliff, 1979). The fitting was performed in two steps. In the first step, the cued trials were used in order to assess the post-search distribution  $\Gamma(k_{PT}, \theta_{PT})$ . Using the  $k_{PT}$  and  $\theta_{PT}$  parameters computed in the first step, the uncued

trials were used to find the best-fitting parameters of the search duration ( $\omega_T$ ,  $k_{ST}$ ,  $\theta_{ST}$ ) in the second step. The best fitting parameters were identified using the simplex algorithm (Nelder & Mead, 1965), which was iteratively run for 1000 times, with the resulting parameters of one iteration providing the starting parameters for the next iteration. The goodness-of-fit was expressed as Bayesian Information Criterion (BIC), which summarily expresses how closely the model predictions match the observed values while penalizing for model complexity. The exact algorithm of BIC computation was as described in an earlier study (Moran et al., 2013). When BICs across two consecutive iterations improved by less than .05, which was taken as evidence that a minimum was reached, the fitting was interrupted, and a new run with a different set of starting parameters was initiated. Overall, it took some 100-200 iterations per run to reach the minimum. To avoid local minima, runs were repeated for 1000 times with different, randomly selected, starting shape and scale parameters.

Fitting post-search duration (cued trials). Experiment 1 varied display density (sparse vs. dense) and distractor variability (fixed vs. variable), yielding four experimental conditions. Depending on which conditions are assumed to differ from the rest, accounting for the RT distributions on cued trials of these four conditions could require different numbers of parameters. *The full model*, assuming that all four conditions yield different RT distributions, would require fitting eight parameters, that is, finding the best shape and scale ( $k_{PT}$  and  $\theta_{PT}$ ) parameters for each of the four conditions. The full model may be restricted by assuming that some of the experimental conditions did not differ. Three such sub-models were fitted to the data: (i) *Sub1* model was motivated by the observed RTs: it assumed differences between sparse and dense displays, whereas the variable and fixed distractor conditions would differ only for sparse displays; (ii) *Sub2* model assumed that only sparse and dense distractors differed, and (iii) *Sub3* model assumed that only variable and fixed distractors differed. Finally, *the null model* assumed no differences between conditions, that is, a single set of  $k_{PT}$  and  $\theta_{PT}$  parameters were used to fit all conditions. Note that, for tasks like the discrimination task used in Experiment 1, a consensus in the literature appears to be that the null model of post-search duration should hold true (Duncan, 1985; Theeuwes, 1992, 2010). Table 2 shows the goodness-of-fit for different models of post-search duration, and Figure 4 depicts the observed values and the predictions of the best-fitting Sub1 model.

Table 2. The goodness-of-fit of different models to the data from Experiment 1, both without penalization for model complexity ( $\log L - \log$ -likelihood) and with penalization using different penalization methods ( $AIC$  - Akaike Information Criterion, and  $BIC$  – Bayesian Information Criterion), together with an index of model complexity ( $df$  - number of free parameters). Indices  $w_{AIC}$  and  $w_{BIC}$ , which were computed on the basis of  $AIC$  and  $BIC$  values (Wagenmakers & Farrell, 2004), denote the conditional probability that the respective model fits the data better than the other tested models. The best-fitting models are marked in bold script. The subscripts “Gamma”, “Gauss”, and “Wald” indicate which probability density function was used to model search durations.

Exp. 1 – Color targets								
Model	df	logL	AIC	$w_{AIC}$	$BIC_{\text{Gamma}}$	$w_{BIC}$	$BIC_{\text{Gauss}}$	$BIC_{\text{Wald}}$
<i>Cued trials, N = 10,408</i>								
Full	8	36,903	36,919	.187	36,977	.000		
<b>Sub1</b>	<b>6</b>	<b>36,904</b>	<b>36,916</b>	<b>.813</b>	<b>36,959</b>	<b>1.00</b>		
Sub2	4	37,084	37,092	.000	37,121	.000		
Sub3	4	37,203	37,211	.000	37,240	.000		

Null	2	37,336	37,340	.000	37,355	.000		
<i>Uncued trials, N = 9,866</i>								
<b>Hybrid model</b>	<b>9</b>	<b>34,710</b>	<b>34,728</b>	<b>1.00</b>	<b>34,793</b>	<b>1.00</b>	<b>34,793</b>	<b>34,792</b>
Diff. pop-out	6	34,854	34,866	.000	34,909	.000	58,806	34,901
No pop-out	5	35,088	35,098	.000	35,134	.000	35,180	35,117

As Table 2 indicates, for cued trials, the Sub1 model fitted the observed data best, as evidenced by the lowest BIC values. Inspection of the best-fitting parameters (Table 3) revealed that the mean post-search duration was shorter for sparse relative to dense displays.

*Figure 4 about here*

*Table 3.* Parameters of the best-fitting models in Experiments 1, 3, and 4. While model fitting used the shape and scale parameters ( $k$  and  $\theta$ , respectively) of gamma distributions, the table shows the more conventional means,  $M = k \cdot \theta$ , and standard deviations,  $SD = \theta \sqrt{k}$ , of the respective distributions. The  $M$  and  $SD$  parameters better illustrate changes in location and variability of gamma distributions across conditions. Further, the scale of target weights depended on the display density (Eq. 4). To ease comparisons across different display densities, the table shows the probability of the target being the first item selected,  $p_1$  (Eq. 3), which has the same scale (0-1) for all densities.

		Exp. 1			Exp. 3			Exp. 4		
Distractors	Density	M	SD	$p_1$	M	SD	$p_1$	M	SD	$p_1$
<i>Cued trials</i>										
Variable	Sparse	433	77		526	101		663	131	
	Medium				522	93		557	115	
	Dense	440	70		527	91		539	100	
Fixed	Sparse	404	66							
	Medium									
	Dense	440	70							
<i>Uncued trials</i>										
Variable	Sparse	123	23	.61	87	9	.76	271	10	.31
	Medium				87	9	.80	198	28	.54
	Dense	65	9	1.0	87	9	1.0	159	17	.95
Fixed	Sparse	96	16	.80						
	Medium									
	Dense	65	9	1.0						

Fitting search duration (uncued trials). In the second step, the parameters defining search duration ( $\omega_{ST}$ ,  $k_{ST}$ ,  $\theta_{ST}$ ) were fitted to the observed data. Because the observed RTs differed between sparse and dense displays and between variable and fixed distractors for sparse displays only, three sets of parameters were fitted: one for dense displays independently of distractor variability, one for sparse displays with variable distractors, and one for sparse displays with fixed distractors. Three different models were tested: (i) the *hybrid model* permitted both the target weight and the parameters of the  $\Gamma_{ST}$  distribution to vary between experimental conditions; (ii) the *different pop-out* model permitted the  $\Gamma_{ST}$  parameters to vary between conditions, while the probability of the target being selected as the first item was fixed

to 1; and (iii) the *no pop-out* model permitted target weights to vary across conditions, with fixed  $\Gamma_{ST}$  parameters for all experimental conditions.

As Table 2 shows, the hybrid model fitted the observed data best, despite its relatively high complexity. Comparing the RTs predicted by the best-fitting hybrid model (Fig. 4) and the observed RTs shows a close fit between the two. Inspection of the best-fitting parameters for the hybrid model (Table 3) revealed that the mean search duration was approximately twice as long for sparse displays relative to dense displays. Furthermore, the search variability was approximately twice as large for sparse relative displays to dense. For dense displays, the probability of selecting the target in the first round ( $p_1$ ) was 1, indicative of pop-out search. By contrast, the respective parameters for sparse displays with fixed and variable displays were .80 and .61, respectively, indicating that the target was *not* the first item selected in 20-40% of the trials with sparse displays.

Model robustness. Inspection of the RT distributions had revealed strong differences between sparse and dense displays. Computational modeling of the observed RTs demonstrated that simply changing distribution parameters of the search duration (as in the ‘different pop-out’ model) yielded relatively good fits to the data. Using the Gamma distribution was particularly convenient, since changes in parameters of the Gamma distribution simultaneously change both the mean and the variance. Critically, however, model fitting showed that the addition of another parameter – namely, target weight – was necessary to produce the best fits. In other words, the variability increases inherent in changing parameters of the Gamma distribution failed to account fully for the observed RTs. The ‘hybrid’ model, by contrast, captured the RT variability differences very well.

It is possible, though, that the Gamma distribution was suboptimal for the distribution of search durations. It is plausible that a different distribution might have yielded overall better fits to the observed data – and, critically, better fits of the simpler models (e.g., different pop-out) relative to the ‘hybrid’ model. To test whether the results of the model fitting were specific to the choice of the distribution, two other, commonly used distributions were examined as well: (i) the normal (N), or Gaussian distribution, and (ii) the inverse-Gaussian ( $N^{-1}$ ), or Wald distribution. Similar to the Gamma distribution, both the Gaussian and the Wald distributions are defined by two parameters: (i) the mean ( $\mu$ ) and the standard deviation ( $\sigma$ ) for the Gaussian, and, respectively, (ii) the mean ( $\mu$ ) and the shape ( $\lambda$ ) parameters for the Wald. To assess the effects of the distribution shape independently of the model parameters, the first and the second moment (i.e., the mean and the variance) were matched between the best-fitting Gamma distributions on the one hand and the Gaussian and the Wald distributions on the other (Eq. 5–7). This made it possible to compute parameters of the Gaussian and the Wald distributions on the basis of the best-fitting parameters of the Gamma distributions, separately for different models of negative search slopes (the ‘no pop-out’, the ‘different pop-out’, and the ‘hybrid’ model).

$$E(\Gamma; k, \theta) = k\theta; V(\Gamma; k, \theta) = k\theta^2 \quad \text{Eq. 5}$$

$$E(N; \mu, \sigma) = \mu; V(N; \mu, \sigma) = \sigma^2 \xrightarrow{\text{yields}} \mu = k\theta; \sigma = \theta\sqrt{k} \quad \text{Eq. 6}$$

$$E(N^{-1}; \mu, \lambda) = \mu; V(N^{-1}; \mu, \lambda) = \frac{\mu^3 \text{ yields}}{\lambda} \rightarrow \mu = k\theta; \lambda = k^2\theta \quad \text{Eq. 7}$$

Table 2 shows the goodness-of-fit (BIC values) for the different models on the assumption that the search durations followed the Gaussian distribution ( $BIC_{\text{Gauss}}$ ) or, respectively, the Wald distribution ( $BIC_{\text{Wald}}$ ). The ‘hybrid’ model yielded the best fits to the observed RTs independently of the chosen distribution (Gamma, Gauss, or Wald), and the predicted RTs were virtually identical across different distributions (Fig. 4). These analyses demonstrate that the results of the initial model fitting were not a mere consequence of having chosen the Gamma distribution.

Another concern is that the hybrid model yielded the best fits because it was able to capture a source of RT variability that the other two models failed to capture. Importantly, this source need not be a difference in target weights (as postulated in the ‘hybrid’ model). Recent advances in computational modeling of other cognitive functions such as working memory suggest that the parameters describing a cognitive function may vary randomly from trial to trial (Fougnie, Suchow, & Alvarez, 2012; van den Berg, Shin, Chou, George, & Ma, 2012). In the context of the present study, this finding implies that using a fixed set of parameters to describe the search and the post-search durations in all trials may underestimate the real variability and decrease the goodness-of-fit to the observed data.

Since the *post-search duration* was modeled using the same set of parameters for all models, the difference in the goodness-of-fit among the various models would originate primarily from the inter-trial variability in the parameters of the *search duration*. To assess the role of inter-trial variability, we expanded the original model of search durations by adding an inter-trial variability parameter ( $\varepsilon$ , Eq. 8). The distribution of the search durations with added inter-trial variability was parameterized by dividing the shape parameter ( $k_{ST}$ ) and multiplying the scale parameter ( $\theta_{ST}$ ) with the inter-trial variability ( $\varepsilon$ ). Accordingly, the mean search duration was the same with and without added inter-trial variability (Eq. 9), whereas the search duration variability increased with higher  $\varepsilon$  (Eq. 10).

$$\Gamma(\varepsilon, k_{ST}, \theta_{ST}) = \Gamma\left(\frac{k_{ST}}{\varepsilon}, \varepsilon\theta_{ST}\right) \quad \text{Eq. 8}$$

$$E(ST) = k_{ST}\theta_{ST}; E(\varepsilon, ST) = \frac{k_{ST}}{\varepsilon} \varepsilon\theta_{ST} = \frac{\varepsilon}{\varepsilon} k_{ST}\theta_{ST} = k_{ST}\theta_{ST} \quad \text{Eq. 9}$$

$$V(ST) = k_{ST}^2\theta_{ST}^2; V(\varepsilon, ST) = \frac{k_{ST}^2}{\varepsilon^2} \varepsilon^2\theta_{ST}^2 = \frac{\varepsilon^2}{\varepsilon^2} k_{ST}^2\theta_{ST}^2 = \varepsilon k_{ST}^2\theta_{ST}^2 \quad \text{Eq. 10}$$

$$PDF(\varepsilon) = N(\mu = 1, \sigma, a = 0, b = \infty); \sigma = \{1, 2, 4\} \quad \text{Eq. 11}$$

The inter-trial variability ( $\varepsilon$ ) was considered a continuous random variable whose probability density function was modeled as a truncated normal distribution with a mean of 1 and a range from 0 to  $\infty$  (Eq. 11). Different degrees of inter-trial variability were modeled by setting the standard deviation ( $\sigma$ ) of  $\varepsilon$  to 1, 2, or 4, representing low, intermediate, and high degrees of inter-trial variability, respectively. Note that setting  $\sigma$  to 0 (i.e., no inter-trial variability) reduces this extended model to the original model

presented earlier. As such, this extended model is a generalization of the original model taking into account inter-trial variability of search duration parameters.

The extended model (Eq. 8–11) was used to simulate search durations on the basis of the ‘hybrid’, the ‘different pop-out’, and the ‘no pop-out’ models, respectively. Random samples of  $\varepsilon$  were drawn separately per different experimental condition for the same number of times as there were trials in that condition in Experiment 1 (total  $N = 9,866$ ). Search durations were simulated separately for different degrees of inter-trial variability ( $\sigma = 1, 2, \text{ or } 4$ ). Nine data sets were simulated in total (3 models  $\times$  3 degrees of inter-trial variability). Figure 5 shows the probability density distributions of the simulated search durations for different data sets.

*Figure 5 about here*

The three different models of negative search slopes (the ‘hybrid’, the ‘different pop-out’, and the ‘no pop-out’ models) without inter-trial variability parameter (Eq. 1–4) were then fitted to the nine simulated data sets which contained inter-trial variability (Eq. 8–11). The details of these 27 model-fitting routines (3 models  $\times$  9 simulated data sets) were the same as in the original model fitting. Table 4 shows the goodness-of-fits (BICs) for the various search slope models separately for each simulated data set. The most striking result of these analyses is that the models used to simulate different data sets were also the models that fitted the simulated data best. In other words, the models without an inter-trial variability parameter are able to explain data with added inter-trial variability. This finding held true even when very large inter-trial variability ( $\sigma = 4$ ) was added to the simulations.

Since plotting the predictions for nine simulated data sets on the basis of three different search slope models is, arguably, impractical, the correspondence between the predicted and the simulated search durations is presented summarily as mean squared error (MSE) in Table 4. The MSE reflects the difference between the predicted and the simulated distributions for the 10<sup>th</sup>, 30<sup>th</sup>, 50<sup>th</sup>, 70<sup>th</sup> and 90<sup>th</sup> percentile across all experimental conditions. More precisely, every MSE is the mean of 20 data points (5 percentiles  $\times$  4 conditions). Inspection of MSEs for the best-fitting models revealed deviations between the observed and the predicted values to be small overall (up to 40 ms), indicating excellent fits. Consistent with the BICs, the models used to simulate the data yielded the smallest MSE<sup>2</sup>.

*Table 4.* Goodness-of-fit (BICs) and mean squared-errors (MSE) for different search slope models (the ‘hybrid’, ‘different pop-out’, and ‘no pop-out’ models) separately for different simulated data set. The models used for fitting the data did not include inter-trial variability, whereas the simulated data sets did include inter-trial variability of different degrees (1, 2, and 4). The best fitting models per data set (i.e., smallest BICs) are marked in bold script. All other details (e.g., degrees of freedom per model) are the same as in Table 2. The MSEs represent the mean squared difference between the simulated and the predicted data for the percentiles 10–90 across all experimental conditions. Larger MSEs indicate worse fits.

---

<sup>2</sup> There was one exception: for the ‘different pop-out’ simulation with large inter-trial variability, the smallest MSE was produced by the ‘hybrid’ model. Nevertheless, this exception was most likely due to increased complexity of the ‘hybrid’ model relative to the ‘different pop-out’ model: the goodness-of-fits after penalization for model complexity (BIC) still indicated that the best-fitting model was the ‘different pop-out’ model.

Fitted model	Simulated data set								
	Hybrid			Different pop-out			No pop-out		
	$\sigma = 1$	$\sigma = 2$	$\sigma = 4$	$\sigma = 1$	$\sigma = 2$	$\sigma = 4$	$\sigma = 1$	$\sigma = 2$	$\sigma = 4$
	<i>BIC (a.u.)</i>								
Hybrid	<b>34,913</b>	<b>34,92</b>	<b>34,797</b>	34,697	34,780	34,843	34,791	34,851	34,823
Diff. pop-out	36,175	35,592	35,137	<b>34,668</b>	<b>34,752</b>	<b>34,816</b>	34,879	34,824	34,795
No pop-out	41,390	39,663	37,836	42,082	40,900	41,851	<b>34,647</b>	<b>34,655</b>	<b>34,701</b>
	<i>MSE (ms)</i>								
Hybrid	37	27	10	5	17	15	15	48	50
Diff. pop-out	122	84	47	4	15	25	54	51	47
No pop-out	419	401	380	282	253	618	5	5	12

Taken together, testing the model robustness showed that the negative search slope models described in Eq. 1–4 are not sensitive to the particular distribution shape used. Moreover, the superior fits of the ‘hybrid’ model cannot be explained by the fact that, unlike the other two models, this model was potentially able to capture inter-trial variability. Thus, the results of Experiment 1 strongly suggest that the performance difference between sparse and dense displays in feature singleton search tasks originates from the differences in target conspicuity (i.e., target weights) between these displays. In fact, the target may be so inconspicuous in sparse displays that it fails to pop out in 40% of trials.

### Experiment 2

The stimuli and experimental design in Experiment 2 were comparable to those of Experiment 1, with one exception: rather than having to discriminate the target singleton, participants simply detected whether or not the target was present. Inspection of response accuracy in Experiment 2 (Table 1) revealed only a few misses (2% of errors on target-present trials) and substantially more false alarms (7% of errors on target-absent trials), indicative of a liberal response criterion. Inspection of response accuracy across experimental conditions, however, revealed no signs that the liberal criterion selectively influenced performance across conditions. These observations were confirmed by a log-linear model of the error rates with target type (present vs. absent), cue, display density, and distractor variability as fixed factors and participant as a random factor. Inspection of the model coefficients revealed significant effects of target type ( $\beta = -.62$ ,  $SE_{\beta} = .16$ ,  $z = -3.87$ ,  $p < .001$ ) and cueing ( $\beta = -.45$ ,  $SE_{\beta} = .15$ ,  $z = -2.96$ ,  $p < .01$ ), as well as a significant interaction between target type and cueing ( $\beta = .93$ ,  $SE_{\beta} = .22$ ,  $z = 4.13$ ,  $p < .001$ ), indicating that cueing influenced false alarms (8.75% false alarms on cued trials vs. 5.33% on uncued) more strongly than misses (2.08% on cued trials vs. 2.93% on uncued). No other main effect or interaction turned out significant (all  $z = 1.30$ , all  $p$ 's  $> .19$ ), indicating that the response bias had little effect on the experimental manipulations of critical interest (display density and distractor variability).

Analyses of median RTs. Outlier detection for correct trials was analogous to Experiment 1. RTs (averaged across participants) on correct trials across different cumulative probabilities for the various experimental conditions of Experiment 2 are shown in Figure 6. Regarding the median RTs, the difference between target-present (filled symbols) and target-absent trials (empty symbols) was

substantial in all experimental conditions (overall 378 ms vs. 424 ms). Cueing had little influence on the median response speed (overall 392 ms for cued vs. 410 for uncued trials). Also, there was virtually no difference in median RTs between sparse and dense displays (overall, 402 ms for sparse and 400 ms for dense displays). The PoP effect was overall very weak (8 ms) and did not vary substantially across conditions. These observations were confirmed by GLM analyses of the median RTs and PoP with target type (present vs. absent), cue, display density, and distractor variability as categorical predictors and participant as a random factor. Inspection of the model coefficients for median RTs revealed the effect of target type to be significant ( $\beta = -51.13$ ,  $SE_{\beta} = 7.01$ ,  $t = -7.29$ ,  $p < .001$ ); no other main effect or interaction reached significance (all  $t$ 's  $< 1.55$ , all  $p$ 's  $> .12$ ). Inspection of the model coefficients for PoP showed only the intercept term to be significant ( $\beta = 9.28$ ,  $SE_{\beta} = 3.65$ ,  $t = 2.54$ ,  $p < .05$ ), indicative of an overall weak but significant PoP effect. No other effects or interactions reached significance (all  $t$ 's  $< 1.06$ , all  $p$ 's  $> .29$ ), indicating that the PoP magnitude varied little across experimental conditions.

*Figure 6 about here*

Analyses of RT variability. As shown in Figure 6, the RT variability differed little across the experimental conditions. Inspection of the PoP magnitude showed stronger PoP for slower RTs, suggesting that changes in RT variability, rather than distribution shifts, were the source of the significant PoP effect on the median RTs. These observations were confirmed by two GLMs for RTs and PoP, respectively, with cumulative probability as a continuous predictor, target type (present vs. absent), cue, display density, and distractor variability as categorical predictors, and participant as a random factor. Inspection of the model coefficients for RTs revealed a significant main effect of target type ( $\beta = -58.30$ ,  $SE_{\beta} = 5.69$ ,  $t = -10.24$ ,  $p < .001$ ), as well as a marginally significant interaction with cumulative probability ( $\beta = 1.97$ ,  $SE_{\beta} = 1.01$ ,  $t = 1.95$ ,  $p = .051$ ). This pattern indicates that the target-present RTs were both 'faster' and somewhat less variable than the RTs on target-absent trials. Finally, the interaction between target type and cue also turned out significant ( $\beta = 24.01$ ,  $SE_{\beta} = 8.04$ ,  $t = 2.98$ ,  $p < .01$ ), indicating that, on *target-present* trials, the fastest RTs were significantly faster on cued trials than on uncued. No other effects or interactions proved significant (all  $t$ 's  $< 1.64$ , all  $p$ 's  $> .10$ ). Inspection of the model coefficients for PoP revealed only a main effect of cumulative probability ( $\beta = 1.62$ ,  $SE_{\beta} = .59$ ,  $t = 2.72$ ,  $p < .01$ ), indicating that PoP magnitude increased with slower RTs. This increase was particularly marked on uncued trials with variable distractors (Figure 6, upper-left), as evidenced by a significant interaction among cumulative probability, cue, and distractor variability ( $\beta = 3.26$ ,  $SE_{\beta} = 1.19$ ,  $t = 2.74$ ,  $p < .01$ ).

Taken together, the analyses of Experiment 2 (detection task) revealed no negative search slopes in either overall RT speed or RT variability. The cueing effects were rather weak ( $\approx 20$  ms), as compared to the substantial cueing effects observed in Experiment 1 ( $\approx 100$  ms). Also at variance with Experiment 1, overall RT speed and RT variability were little influenced by distractor variability. The PoP effect was very weak (8 ms) and primarily driven by the changes in RT variability.

### Experiment 3

In Experiment 3, similar to Experiment 1, participants performed a discrimination task. In contrast to Experiment 1, Experiment 3 (Fig. 2B) investigated negative search slopes for a greater variety of display densities: (i) sparse (6 items), (ii) medium (12 items), and (iii) dense displays (36 items). Inspection of the error rates (Table 1) showed that participants responded very accurately overall (2.5% of errors), with somewhat worse performance for uncued (3.5%) relative to cued trials (1.6%). The effect of display density was very small (2.8%, 2.4%, and 2.3% of errors for sparse, medium, and dense displays, respectively) and it did not co-vary with cue type. These observations were confirmed by a log-linear model, which revealed the effect of cueing to be significant ( $\beta = .83$ ,  $SE_{\beta} = .19$ ,  $z = 6.18$ ,  $p < .001$ ); all other effects were non-significant (all  $z$ 's  $< .9$ , all  $p$ 's  $> .418$ ).

Analyses of median RTs. Median RTs were overall faster on cued (510 ms) than on uncued trials (615 ms). On *uncued trials*, moderately negative search slopes were evidenced by RTs decreasing slightly with increasing display density (624 ms, 618 ms, and 601 ms for sparse, medium, and dense displays, respectively). On *cued trials*, median RTs were only little influenced by increasing display density (507 ms, 508 ms, and 514 ms for sparse, medium, and dense displays). Priming of pop-out (PoP) decreased substantially with increasing display density (53 ms, 29 ms, and 7 ms for sparse, medium and dense displays, respectively) on uncued trials, and to a lesser degree on cued trials (20 ms, 15 ms, and 8 ms, respectively).

*Figure 7 about here*

These observations were confirmed by two separate GLM analyses with RTs and PoP as dependent variables, display density and cue as categorical predictors, and participant as a random factor. Inspection of the model coefficients for the median RTs revealed significant effects of cueing ( $\beta = 116.82$ ,  $SE_{\beta} = 11.64$ ,  $t = 10.04$ ,  $p < .001$ ). For *cued trials*, no effects of display density were observed (all  $t$ 's  $< .6$ , all  $p$ 's  $> .56$ ). For *uncued trials*, only the contrast between sparse and dense displays turned out marginally significant ( $\beta = -29.92$ ,  $t = -1.82$ ,  $p = .073$ ). Inspection of the model coefficients for PoP revealed a significant priming effect for cued trials ( $\beta = 20$ ,  $SE_{\beta} = 5$ ,  $t = 4.05$ ,  $p < .001$ ), which was little influenced by display density (all  $t$ 's  $< -1.84$ , all  $p$ 's  $> .06$ ). An even stronger PoP effect was observed for uncued trials ( $\beta = 33.22$ ,  $SE_{\beta} = 6.74$ ,  $t = 4.93$ ,  $p < .001$ ), which decreased substantially with an increase in display density (sparse vs. medium,  $\beta = -19.72$ ,  $SE_{\beta} = 9.53$ ,  $t = -2.07$ ,  $p < .05$ ; sparse vs. dense,  $\beta = -34.48$ ,  $SE_{\beta} = 9.53$ ,  $t = -3.62$ ,  $p < .001$ ). Taken together, analyses of the median RTs in Experiment 3 showed relatively weak effects of display density, which were, however, stronger for uncued than for cued trials. Furthermore, the PoP effect varied substantially across display densities for uncued trials.

Analyses of RT distributions. Similar to Experiment 1, the RT variability increased with an increase in display density in Experiment 3 (Fig. 7): the variability was largest for sparse displays, intermediate for medium-density, and smallest for dense displays. This trend was more prominent for uncued than for cued trials, indicating that the increase in RT variability was due to variability in search duration. Inspection of PoP magnitude revealed stronger PoP for slower RTs; this trend was most prominent for

sparse displays, followed by medium-density and dense displays. The RT distributions were analyzed by a GLM with RTs as dependent variable, cue type and display density as categorical predictors, and cumulative probability as a continuous predictor. Inspection of the model coefficients revealed the main effect of cueing ( $\beta = 92.72$ ,  $SE_{\beta} = 9.01$ ,  $t = 10.29$ ,  $p < .001$ ) and the interaction between cue and cumulative probability ( $\beta = 3.53$ ,  $SE_{\beta} = 1.60$ ,  $t = 2.20$ ,  $p < .05$ ) to be significant. These findings indicate that cueing influenced both the fastest RTs and the RT variability. For *cued trials*, display density had little influence on the fastest RTs (all  $t$ 's  $< 1.8$ , all  $p$ 's  $> .08$ ) and on RT variability (all  $t$ 's  $< 1.8$ , all  $p$ 's  $> .07$ ). For *uncued trials*, the effect of display density on the fastest RTs was relatively weak (contrast sparse vs. medium,  $\beta = 12.98$ ,  $SE_{\beta} = 6.51$ ,  $t = 1.99$ ,  $p = .047$ ; sparse vs. dense,  $\beta = 11.25$ ,  $SE_{\beta} = 6.51$ ,  $t = 1.72$ ,  $p = .085$ ). By contrast, display density had a substantial influence on RT variability (sparse vs. medium,  $\beta = -3.48$ ,  $SE_{\beta} = 1.16$ ,  $t = -3.00$ ,  $p < .01$ ; sparse vs. dense,  $\beta = -7.23$ ,  $SE_{\beta} = 1.16$ ,  $t = -6.25$ ,  $p < .001$ ). Overall, these analyses show that, on cued trials, the RT distributions differed little across display densities. On uncued trials, by contrast, the RT variability was much stronger for sparse relative to dense displays.

Analysis of the PoP magnitude revealed a significant effect of cueing, with stronger PoP on uncued relative to cued trials ( $\beta = 16.95$ ,  $SE_{\beta} = 5.46$ ,  $t = 3.10$ ,  $p < .05$ ). Further, the PoP effect was stronger for slower RTs ( $\beta = 4.53$ ,  $SE_{\beta} = .69$ ,  $t = 6.59$ ,  $p < .001$ ). Finally, the increase in the PoP magnitude for slower RTs was the strongest for sparse relative to both medium-density displays ( $\beta = -2.86$ ,  $SE_{\beta} = .97$ ,  $t = -2.95$ ,  $p < .05$ ) and dense displays ( $\beta = -5.27$ ,  $SE_{\beta} = .97$ ,  $t = -5.43$ ,  $p < .001$ ).

Modelling negative search slopes. Since analyses of the RT distributions revealed larger RT variability for sparse relative to dense displays, the RT distributions observed in Experiment 3 across the different display densities were fitted to the same theoretical models as in Experiment 1. Table 4 shows the goodness-of-fits for the different models with the best-fitting models marked in bold script.

The post-search duration was assessed by fitting the various models to the RT distributions in the cued conditions. Table 5 shows that the Sub1 model yielded better fits to the observed data, as indicated by the lowest BIC and AIC values. Figure 8 shows that the values predicted by the Sub1 model were very close to the observed values in Experiment 3. Inspection of the best-fitting parameters (Table 3) revealed that, while the mean post-search times differed little across display densities, the main difference derived from smaller variability of the post-search durations with an increase in display density.

Table 5. The goodness-of-fit to the data from Experiments 3–4 . Conventions as in Table 2. Best-fitting models are marked in bold script.

Model	df	logL	Exp. 3 – Color targets				Exp. 4 – Orientation targets				
			AIC	w <sub>AIC</sub>	BIC	w <sub>BIC</sub>	logL	AIC	w <sub>AIC</sub>	BIC	w <sub>BIC</sub>
<i>Cued trials</i>											
<i>N = 10,160</i>						<i>N = 9,437</i>					
<b>Sub1</b>	<b>6</b>	<b>39,964</b>	<b>36,976</b>	<b>1.00</b>	<b>40,020</b>	<b>.936</b>	<b>38,617</b>	<b>38,629</b>	<b>1.00</b>	<b>38,671</b>	<b>1.00</b>
Null	2	40,007	40,010	.000	40,025	.064	40,495	40,499	.000	40,514	.000
<i>Uncued trials</i>											

		<i>N</i> = 9,793					<i>N</i> = 8,056				
Hybrid model	9	38,248	38,266	.001	38,330	.000	<b>31,491</b>	<b>31,509</b>	<b>1.00</b>	<b>31,572</b>	<b>1.00</b>
Diff. pop-out	6	38,273	38,285	.000	38,328	.000	37,134	37,146	.000	37,188	.000
No pop-out	<b>5</b>	<b>38,241</b>	<b>38,251</b>	<b>.999</b>	<b>38,287</b>	<b>1.00</b>	31,937	31,902	.000	31,937	.000

The RT distributions on uncued trials were used for fitting of the search duration parameters across display densities. As can be seen from Table 5, the ‘no pop-out’ model yielded the closest fits to the observed data. As Figure 8 shows, the values predicted by the ‘no pop-out’ model were very close to the observed values. This model assumes that only the target weight varied across display densities, while a single gamma distribution described the search duration across all densities. Inspection of the model parameters (Table 3) revealed that the probability of selecting the target in the first selection round decreased from 1.0 to .76 with a decrease in display density. In other words, in 20–25% of the trials with sparse and medium-density displays, the target was not the first item selected.

*Figure 8 about here*

#### *Experiment 4*

Experiment 4 used the same design as Experiment 3, the only difference being that, instead of color singleton targets, participants had to find orientation singletons (Fig. 2B). Inspection of the error rates (Table 1) revealed that the orientation singletons used in Experiment 4 were rather difficult to find, as indicated by relatively high error rates with uncued targets (9%). For cued trials, the error rates were comparable to those in Experiments 1–3 (3%). Further, the error rates decreased with an increase in display density (8%, 6%, and 4% for sparse, medium, and dense displays, respectively). These observations were confirmed by a log-linear model of the error rates, which revealed a substantial effect of cueing ( $\beta = 1.10$ ,  $SE_{\beta} = .10$ ,  $z = 11.51$ ,  $p < .001$ ), as well as significant effects of display density for both cued and uncued trials (all  $z$ 's  $> -2.39$ , all  $p$ 's  $< .05$ ).

Analyses of median RTs. Inspection of the median RTs revealed very strong cueing effects, with some 400 ms faster RTs for cued (558 ms) relative to uncued trials (946 ms). Further, median RTs decreased with an increase in display density for cued trials (620, 533, and 522 ms for sparse, medium, and dense displays, respectively), and even more prominently for uncued trials (1291, 858, and 691 ms, respectively). A GLM of the median RTs revealed a significant effect of cueing ( $\beta = 670.77$ ,  $SE_{\beta} = 66.07$ ,  $t = 10.15$ ,  $p < .001$ ). Despite numerical trends, the effect of display density on cued trials did not reach significance (all  $t$ 's  $< -1.5$ , all  $p$ 's  $> .14$ ). The effect of display density on uncued trials, however, was significant and very marked (sparse vs. medium contrast,  $\beta = -345.86$ ,  $SE_{\beta} = 93.44$ ,  $t = -3.70$ ,  $p < .001$ ; sparse vs. dense,  $\beta = -502.42$ ,  $SE_{\beta} = 93.44$ ,  $t = -5.38$ ,  $p < .001$ ). Inspection of the PoP magnitude suggested a pattern of stronger PoP on uncued relative to cued trials (22 vs. 4 ms), and stronger PoP for sparse relative to dense displays (23, 14, and 1 ms for sparse, medium, and dense displays, respectively). These observations, however, were not substantiated by a GLM of the PoP magnitude, which failed to reveal any significant effects (all  $t$ 's  $< -1.5$ , all  $p$ 's  $> .15$ ). As shown in Figure 7, the standard errors of the PoP effects were rather large, indicating strong individual differences in PoP magnitude. This was the

most likely reason why effects of display density on the PoP magnitude failed to reach significance despite the numerical differences.

Analyses of RT distributions. Inspection of the RT distributions (Fig. 7) showed marked distributions shifts and even more prominent increases in the RT variability with a decrease in display density. The density effect was much stronger for uncued than for cued trials. A GLM of the RT distributions with cue and display density as categorical predictors and cumulative probability as continuous predictor revealed strong cueing effects for both the fastest RTs ( $\beta = 323.45$ ,  $SE_{\beta} = 46.28$ ,  $t = 6.99$ ,  $p < .001$ ) and the RT variability ( $\beta = 59.26$ ,  $SE_{\beta} = 5.81$ ,  $t = 10.19$ ,  $p < .001$ ). For *cued trials*, display density influenced primarily the RT variability (contrast sparse vs. medium,  $\beta = -23.10$ ,  $SE_{\beta} = 8.22$ ,  $t = -2.81$ ,  $p < .01$ ; sparse vs. dense,  $\beta = -28.83$ ,  $SE_{\beta} = 8.22$ ,  $t = -3.50$ ,  $p < .001$ ), while the effects of display density for the fastest RTs did not reach significance (all  $t$ 's  $< .45$ , all  $p$ 's  $> .67$ ). For *uncued trials*, display density influenced both the RT variability and the fastest RTs (all  $t$ 's  $> -3.90$ , all  $p$ 's  $< .001$ ).

Inspection of the PoP magnitude (Fig. 7) across different RTs showed a pattern similar to that observed in Experiments 1 and 3: PoP was stronger on uncued than on cued trials; further, PoP increased for slower RTs, and this increase was most prominent with sparse displays. A notable exception from this pattern were the PoP effects on uncued trials with sparse displays, where PoP magnitude did not co-vary systematically with an increase in RTs. This condition was also characterized by very strong individual differences (see Fig. 7, error bars for PoP on uncued trials with sparse displays). These observations were confirmed by a GLM, which revealed a strong increase of PoP magnitude with slower RTs on *cued trials* ( $\beta = 6.96$ ,  $SE_{\beta} = 1.44$ ,  $t = 4.84$ ,  $p < .001$ ). As suggested by the inspection of the numerical trends (Fig. 7), PoP magnitude did not co-vary with response speed on *uncued trials* ( $t = .39$ ,  $p = .69$ ).

Taken together, analyses of the RTs observed in Experiment 4 revealed a pattern similar to that of Experiments 1 and 3: the median RTs increased with a decrease in display density, that is, negative search slopes were observed. Analyses of the RT distributions revealed that the negative search slopes were due to both distributions shifts and an increase in RT variability. The increase in RT variability with a decrease in display density was much stronger on uncued than on cued trials, indicating that display density influenced the variability of search durations. Finally, PoP magnitude increased with slower RTs, at least on cued trials. The respective findings on uncued trials were equivocal, primarily due to large individual differences observed with sparse displays.

Modeling negative search slopes. Similar to Experiments 1 and 3, the RT distributions observed in Experiment 4 were fitted to the various models of the search and post-search durations. Fitting of the data for cued trials revealed the Sub1 model to yield the best fits to the data, as evidenced by the smallest BIC and AIC values for this model (Table 5). Figure 8 shows that, in Experiment 4, the predicted values closely matched the observed values on cued trials for medium-density and dense displays. The model captured less well the RT distribution for sparse displays, presumably due to strongly skewed RT distribution observed in this condition (Fig. 8). Inspection of the model parameters (Table 3) showed

that an increase in display density increased both the mean duration and the variability of the post-search processes.

Fitting of the data for uncued trials revealed the ‘hybrid’ model to yield the closest fits to the observed data (Table 5). Figure 8 shows a very close correspondence between the observed and the predicted values. Inspection of the model parameters indicated that finding the target took longest in sparse displays (271, 198, and 159 ms for sparse, medium, and dense displays, respectively). The probability of selecting the target in the first selection round was smallest for sparse displays (.31, .54, and .95 for sparse, medium and dense displays, respectively): the target was not the first item selected in 70% of trials with sparse displays, 50% with medium-density displays, and 5% with dense displays.

## Discussion

The four experiments reported in this study characterized some of the boundary conditions for negative search slopes to manifest. First, negative search slopes were evident in discrimination tasks only (Experiments 1, 3, and 4), independently of the precise target-defining stimulus features (surface color, contour color, or contour orientation, respectively). Second, the median RTs scaled linearly with an increase in display density from very sparse (3 items, Experiment 1) through sparse and medium-density (6 and 12 items, Experiments 3 and 4, respectively) to very dense displays (36 items, Experiments 1, 3, and 4). Third, the negative search slopes primarily reflected changes in RT variability across display densities; distribution shifts, by contrast, contributed only sporadically to the negative slopes. Finally, the density effects were more pronounced for uncued relative to cued trials, indicating that changes in display density influenced primarily the search duration, rather than the post-search duration.

### *Search dynamics across display densities*

Regarding the alternative models of negative search slopes (Fig. 1), the observed RT distributions clearly disqualify the ‘slow pop-out’ model, which predicted only distribution shifts across the different densities. The three remaining models (the ‘different pop-out’, the ‘no pop-out’, and the ‘hybrid’ model) were all consistent with the observed RT distributions. Fitting these models to the observed data revealed that a decrease in display density prolonged the mean search duration and increased the search variability. Critically, to achieve the best fits to the observed data, in all three experiments that used a discrimination task, it was necessary to assume strong differences in target conspicuity (i.e., the target weight) across the different display densities. Model fitting indicated that the target was so inconspicuous in sparse displays that it failed to be the first item selected in 20–70% of the trials. This outcome is consistent with the results of eye-tracking studies which typically found that, with sparse displays, participants first directed their gaze to a distractor in some 20% of the trials (Becker, 2008; Caddigan & Lleras, 2010)<sup>3</sup>. By contrast, the target was the first item selected on virtually all trials with

---

<sup>3</sup> Furthermore, even when the target was immediately fixated, the saccadic latencies were relatively long ( $\approx 350$  ms) compared to typical saccade latencies in selective-attention tasks (about 250 ms; Jonikaitis & Theeuwes,

dense displays. Going beyond previous studies, the present study strongly supports the idea that feature singleton targets, in fact, may fail to pop out in (as much as) two out of three trials with sparse displays, whereas the same targets reliably pop out in dense displays (see, e.g., Rangelov et al., 2013, and Tseng, Glaser, Caddigan, & Lleras, 2014, for similar ideas).

#### *Post-search dynamics across display densities*

Analyses of the RT distributions for cued trials helped characterize the post-search dynamics across different display densities. On cued trials, a cue stimulus was briefly presented prior to the onset of the search display. When presented, the location of the cue always coincided with that of the target. Consequently, attending to the cue eliminated the need to search for the target and left post-search processes as the only determinant of response speed. Analyses of the RTs on cued trials in Experiments 1, 3, and 4 consistently revealed differences between display densities. Additionally, model fitting for Experiments 1, 3, and 4 consistently showed that, in order to accurately predict the observed RTs, it is necessary to assume that post-search processes differ across display densities. The precise pattern of results was less consistent across experiments. In Experiment 1, the *mean* post-search duration was shorter for sparse relative to dense displays. In Experiment 3, differences in display density primarily influenced the post-search time *variability*, which was larger for sparse than for dense displays. Finally, in Experiment 4, both the *mean* post-search duration and the post-search time *variability* were influenced by display density. The experiments reported here do not permit any conclusive interpretations as to the cause of the equivocal data patterns across the different experiments. As the experiments were not designed to address this issue in detail, further studies are necessary to comprehensively characterize the effects of display density on post-search duration. Importantly, though, the available findings do highlight the fact that distractors, and their density in particular, not only influence the search dynamics, but also the post-search processes. This aspect of our findings is in stark contrast with conventional opinion, which holds that, after the target item has been successfully selected for focal inspection, the rest of the visual display would matter little. Critically, the fact that both search and post-search dynamics can be sensitive to the same experimental manipulation (e.g., display density) emphasizes the importance of quantitative, computational models of RTs, which can separate different sources of RT variability, for theories of early visual processes.

#### *The role of task and stimulus features for negative search slopes*

The results of the present study replicated a strong dissociation between detection and discrimination tasks (Bravo & Nakayama, 1992; Song & Nakayama, 2006). In discrimination tasks (Experiments 1, 3, and 4), selecting the precise location of the target is critical for successful performance because if, by chance, a distractor location is selected instead, this would severely affect response accuracy. In detection tasks (Experiment 2), by contrast, it suffices to simply detect a feature discontinuity *anywhere* in display to respond correctly (Chan & Hayward, 2009, 2014; Treisman, 1988; Treisman & Gelade, 1980;

---

2013). Thus, it is possible that the portion of ‘failure-to-pop-out’ trials was even higher in the eye tracking studies, but that a (fast) erroneous saccade to a distractor was not actually triggered on some of these trials.

Treisman & Sato, 1990). Consistent with the idea that the negative search slopes reflect a failure to correctly select the precise target location, there was a strong effect of display density on uncued trials in Experiments 1, 3, and 4. The display density effect was primarily driven by increased RT variability for sparse relative to dense displays, rather than by overall shifts in RT distributions. By contrast, in Experiment 2, which used a detection task, no differences between sparse and dense displays were observed, for either cued or uncued trials. The absence of negative search slopes in detection tasks and their presence in discrimination tasks indicate that success in selecting the target location is the primary determinant of the negative search slopes.

Analyses of the effects of stimulus features also suggested that search processes differ between sparse and dense displays. The effects of both the intertrial sequence of target features (PoP) and distractor variability were much stronger for sparse than for dense displays. Additionally, analyses of PoP magnitude across different, slow and fast RTs showed that, for comparable RTs, the PoP effect was still greater for sparse displays than for dense. This suggests that display density effects on PoP magnitude are not simply due to overall faster RTs to dense relative to sparse displays. Together with the computational modeling results, these findings suggest that stimulus features matter most when the target frequently fails to pop out. This result has important implications for many studies that investigated effects of stimulus features on early visual processing (e.g., Becker, 2008; Becker & Ansorge, 2013; Hickey, Olivers, Meeter, & Theeuwes, 2011; Maljkovic & Nakayama, 1994, 1996; Olivers & Hickey, 2010; Theeuwes, 1992). Such studies typically used search arrays comprising a singleton target and relatively few distractors (from 3 to 12 items). The present finding that singleton targets in this range are frequently not the first item selected (i.e.,  $p_1 \ll 1$ ) raises a concern as to whether the target can actually be said to have popped out in those studies.

### *Theoretical implications*

In the literature, there are several broad theoretical frameworks that may serve as general theories of visual attention. Here, we discuss the ability of the four most successful frameworks to account for the present findings. From the earliest to the latest, these frameworks are: (i) Feature Integration Theory (Treisman, 1988; Treisman & Gelade, 1980; Treisman & Sato, 1990), (ii) Guided Search (Wolfe, 1994, 2007; Wolfe, Cave, & Franzel, 1989), (iii) Theory of Visual Attention (Bundesen, 1990, 1998), and (iv) Boolean Map theory (Huang, 2010; Huang & Pashler, 2007). While these frameworks differ in numerous respects, the present discussion will focus on the differences in how they conceptualize early vision. In our understanding, the processes constituting early vision would involve (i) segmenting the visual input into potentially relevant locations (e.g., the target location) and irrelevant locations (e.g., distractor locations) and (ii) selecting potentially relevant locations for focal-attentional inspection.

According to *Feature Integration Theory* (FIT), early vision represents visual input in two distinct ways: (i) a map of locations representing which locations in the scene are occupied by an item, and (ii) feature maps representing which features (e.g., red, green, etc.) are present at these locations. Importantly, the signals from these maps indicate either the presence (i.e., 1) or absence (0) of a particular feature. Similar to the FIT, a key assumption of *Boolean Map* (BM) theory is that visual input is represented in a

categorical manner (1 or 0) at the various map locations. For example, a red singleton would be represented as 1 in the 'red' independently of display density, that is, a feature singleton would yield the same signal in all displays. At variance with this prediction, the negative search slopes observed in this and earlier studies imply that, for the same stimulus, the output of early processes differs between sparse and dense displays. Thus, the processes of early vision as conceived by FIT and BM theory fail to account for the findings of the present study. Of note, BM theory allows for the existence of a bottom-up salience map which represents locations that are inherently salient on the basis of feature contrast computations (Huang & Pashler, 2007, p. 615). By virtue of inheriting salience computations from earlier work (e.g., Itti & Koch, 2001), the BM could, in principle, account for the density effects observed in the present study. Importantly, however, the BM assumes these processes to be fully automatic, that is, impenetrable to top-down influences of the task set. This aspect of the BM theory is difficult to reconcile with the evidence that negative search slopes are specific to discrimination tasks.

While the *Theory of Visual Attention* (TVA) focuses on explaining relatively late attentional processes, that is, processes following visual segmentation, the CTVA (an extension of TVA) attempted to model segmentation processes (Logan, 1996; the 'C' in 'CTVA' stands for 'CODE' or 'COntour DETector'). According to the CTVA, the visual scene is represented as a map of basic stimulus features (e.g., different colors, orientations, etc.). An important difference to FIT and BM theory is that these representations are continuous, with higher values indicating stronger signals. Thus, in principle, CTVA could accommodate the fact that the same stimulus is less conspicuous in sparse than in dense displays and so explain the negative search slopes. However, CTVA, as originally conceived, does not postulate interactions between distractors, such as in terms of iso-feature suppression mechanisms. So, in order to explain the negative search slopes, CTVA would require substantial modifications and extensions.

The *Guided Search* (GS) model postulates that the visual scene is represented as a map of feature difference (or feature contrast) signals that can vary continuously. The outputs of feature-specific maps are then pooled across all features at the level of the overall-saliency map – a 'featureless' representation that signals *how different* (or salient) a given location is relative to its surround. Unlike the previously discussed models, GS postulates the existence of iso-feature suppression mechanisms at the level of feature-maps that could account for the display density effects and negative search slopes. However, GS also postulates that performance in all tasks relies on the overall-saliency map, which signals candidate target locations for focal-attentional inspection – in fact, this was the earliest, and most critical point of distinction between GS and FIT. Given this, the GS framework cannot readily accommodate the finding that negative search slopes are present in discrimination tasks and absent in detection tasks. Thus, taken together, the results of the present study are difficult to accommodate by any of the dominant theories of early vision. – In what follows, we present our account of the present findings.

Similar to the GS account of early vision, we believe that strong and reliably negative search slopes indicate that early vision represents visual input as a map of potentially interesting locations. Constructing this map is characterized by strong suppressive interactions between similar stimuli (see Itti & Koch, 2000, for computational modeling of distractor salience under conditions of varying display

density). In this framework, the relation between the target signal and distractor signals could be conceptualized in terms of a signal-to-noise ratio (SNR). The density of distractors would influence the strength of the distractor signals and, by implication, the noise component of the SNR: strong distractor signals in sparse displays would result in strong noise. In fact, as the present findings indicate, strong noise may effectively prevent singleton targets to pop out. In contrast to the classical notions of pop-out, which define pop-out by its independence of (the number and density of) distractors, our findings suggest that pop-out does critically depend on distractors.

When the SNR does not permit reliable selection of the target based on its overall-saliency signal, as with sparse displays, additional adaptive processes may come into play to help find the target. Since the present study demonstrated that feature singletons do not invariably (if at all) pop out in sparse displays, any selected item may be the target or a distractor. Consequently, participants have to adapt search processes so as to ensure that they respond to the target rather than to a distractor. These adaptive mechanisms may influence pre-selective processing of the visual scene. For example, the past target features may be selectively enhanced, and/or the past distractor features may be suppressed (Treisman & Sato, 1990). Further, participants may adopt an altogether different ‘search mode’ in which they selectively scan the visual scene for the past target feature (i.e., a ‘feature mode’), rather than looking for a singleton element (i.e., a ‘singleton mode’; Bacon & Egeth, 1994). Alternatively, or in addition, repeating precise stimulus features may speed up deciding whether or not an already selected item is the target (Huang, Holcombe, & Pashler, 2004). Finally, the response-selection processes may be influenced by the feature sequence across trials (Yashar & Lamy, 2011). In contrast to sparse displays, the SNR in dense displays would permit reliable target selection on the basis of ‘featureless’ saliency signals alone, that is, engaging feature-specific processes would not be necessary. The present study cannot discriminate between the various mechanisms proposed to account for feature-specific effects in singleton search tasks (Becker, 2008; Caddigan & Lleras, 2010; Eimer, Kiss, & Cheung, 2010; Hillstrom, 2000; Shin, Wan, Fabiani, Gratton, & Lleras, 2008; Tseng et al., 2014). In fact, recent findings suggest that multiple, pre- and post-selective mechanisms can give rise to feature-specific effects (Krummenacher, Grubert, & Müller, 2010; Lamy et al., 2008; Lamy, Yashar, & Ruderman, 2010; Rangelov, Müller, & Zehetleitner, 2011, 2012). However, the present study highlights the fact that the feature-specific mechanisms are adaptive, exerting an influence only when search processes cannot rely on target saliency. Put differently: sensitivity to stimulus features is not a general governing principle of early vision, but rather specific for certain types of visual input.

Concerning the differential density effects between detection and discrimination tasks, a possible explanation may be that responses in the two types of task rely on *different properties* of the overall-saliency map. As argued earlier, accurate performance in discrimination tasks requires selecting a narrow region around the target. By contrast, the target’s presence can be detected on the basis of summary statistics of the overall-saliency map (Rosenholtz, Huang, & Ehinger, 2012, expressed a similar idea in a more elaborate way). One such statistic may be *the variability of saliency signals* across different locations. Since, in target-absent displays, all stimuli are identical, the signals from individual locations will be comparable – yielding relatively low variability across different locations. By contrast, in target-present trials, the signal from one location (the target) should be larger than the signals from the

other locations, thus increasing the variability. On trials with sparse displays, this variability in target-present trials will still be fractionally stronger compared to target-absent trials, permitting relatively fast target detection even in sparse displays. When targets are frequently present (such as in Experiment 2), even weak variability differences between target-present and target-absent displays may permit accurate target detection. In more general terms, we propose that *in detection tasks, participants rely on the summary statistic of the overall-saliency map. In discrimination tasks, by contrast, participants use the spatial layout of the master map* to select a narrow region for focal-attentional inspection. Concerning the differential feature-specific effects in detection versus discrimination tasks, assuming that responses in detection tasks rely on the featureless summary statistic would be sufficient to account for the weak PoP and distractor variability effects in such tasks.

Finally, the results of the modeling analyses consistently demonstrated that, even though the  $p_1$  parameters for sparse displays were substantially lower than 1, the search for feature singletons in sparse displays was not fully random. Instead, our findings show that the saliency of feature singleton targets was still somewhat higher than that of distractors. This demonstrates that the search efficiency is a continuously distributed variable bound by two extremes: (i) fully efficient search (i.e., pop-out search,  $p_1 \approx 1$ ) and (ii) fully random search (i.e.,  $p_1 \approx \text{number of items}^{-1}$ ). The present finding, together with recent work (Buetti et al., 2016; Moran et al., 2013), highlight the need to account for the full spectrum of search efficiency – that is, including the range between fully random and fully efficient, pop-out search – rather than focusing on explaining the dichotomy between these extremes.

#### *Methodological implications*

The evidence that searching for feature singletons in sparse and dense displays is qualitatively different raises concerns about the reliability of flat search slopes as a criterion of pop-out. To illustrate, even if the target fails to pop out in sparse displays, finding the target will still be relatively fast given the small number of items to search through. On the other hand, finding the same target in dense displays would also be relatively fast, however because the target does pop out. Thus, sparse and dense displays can yield comparably fast RTs (and flat search slopes), but for different reasons. The results of the present study suggest that a more reliable criterion of search efficiency might be the variability of search duration: computational modeling indicated that the variability of search duration was twice as large when the target failed to pop out relative to when it did pop out. The small variability of search duration when the target pops out implies that the total RT variability on uncued trials is primarily driven by the variability of the post-search duration. Consequently, there should be no differences in RT variability between cued and uncued trials because both these trial types would primarily reflect the same post-search processes. With this in mind, we propose usage of the cueing procedure as a diagnostic measure of pop-out search, instead of the set size manipulation (e.g., Horowitz & Wolfe, 1998; Rauschenberger & Yantis, 2001; Wolfe & Horowitz, 2004). In contrast to the set size method, the cueing procedure relies on the *same* stimulus display, which avoids comparing RTs across different stimuli and potential risks that search processes differ qualitatively between different displays. From a practical perspective, since the search variability criterion compares RTs between only two conditions, its implementation might prove more cost-effective compared to implementing the search slope criterion.

The dissociation between the detection and discrimination tasks raises a concern about how sensitive the performance in detection task is to identify pop-out (which is defined as the target being the first item *selected*). We propose that using discrimination tasks, rather than detection tasks, would permit to reliably discern whether or not a stimulus pops out.

## **Conclusions**

The present findings suggest that pop-out search is not an all-or-nothing phenomenon. Rather, the probability of selecting the target in the first round varies continuously and is influenced by at least two properties of the search display. On the one hand, if the target is a feature singleton, it is likely that it will be selected in the first several rounds, though not necessarily the very first round. On the other hand, for a target to pop out on all trials, a sufficiently high display density is necessary. Due to iso-feature suppression mechanisms, by which detectors coding similar stimuli inhibit each other, the many, closely spaced distractors in dense displays would be less conspicuous relative to the few and widely spaced distractors in sparse displays. Thus, even though the target is the same, the relative difference between the target and distractors is much lower in sparse than in dense displays. In fact, the present study suggests that, for sparse displays, this difference is so small that the target very frequently does not pop out at all.

## **Author Note**

The mean RTs and intertrial effects for the mean RTs in Experiment 1 were presented as Experiment 1 in Rangelov, D., Müller, H. J., & Zehetleitner, M. (2013). *Journal of Vision*, 13(3), 22. <https://doi.org/10.1167/13.3.22>. The analyses of RT distributions and the computational models of Experiment 1 as well as the results of Experiments 2–4 have not been disseminated in any form prior to this publication. This research was supported by the German Research Foundation (DFG) grants RA 2191/1-1 to D.R. and MU773/14-1 and MU773/16-1 to H.M.

## References

- Bacon, W. F., & Egeth, H. E. (1994). Overriding stimulus-driven attentional capture. *Perception & Psychophysics*, *55*(5), 485–496. <https://doi.org/10.3758/BF03205306>
- Becker, S. I. (2008). The mechanism of priming: episodic retrieval or priming of pop-out? *Acta Psychologica*, *127*(2), 324–339.
- Becker, S. I., & Ansorge, U. (2013). Higher set sizes in pop-out search displays do not eliminate priming or enhance target selection. *Vision Research*, *81*, 18–28. <https://doi.org/10.1016/j.visres.2013.01.009>
- Bravo, M. J., & Nakayama, K. (1992). The role of attention in different visual-search tasks. *Perception & Psychophysics*, *51*(5), 465–72.
- Buetti, S., Cronin, D. A., Madison, A. M., Wang, Z., & Lleras, A. (2016). Towards a better understanding of parallel visual processing in human vision: Evidence for exhaustive analysis of visual information. *Journal of Experimental Psychology. General*, *145*(6), 672–707. <https://doi.org/10.1037/xge0000163>
- Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, *97*(4), 523–547.
- Bundesen, C. (1998). A computational theory of visual attention. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *353*(1373), 1271–1281.
- Caddigan, E., & Lleras, A. (2010). Saccadic repulsion in pop-out search: how a target's dodgy history can push the eyes away from it. *Journal of Vision*, *10*(14), 9. <https://doi.org/10.1167/10.14.9>
- Chan, L. K. H., & Hayward, W. G. (2009). Feature integration theory revisited: dissociating feature detection and attentional guidance in visual search. *Journal of Experimental Psychology. Human Perception and Performance*, *35*(1), 119–132. <https://doi.org/10.1037/0096-1523.35.1.119>
- Chan, L. K. H., & Hayward, W. G. (2014). No attentional capture for simple visual search: Evidence for a dual-route account. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(6), 2154–2166. <https://doi.org/10.1037/a0037897>
- Duncan, J. (1985). Visual search and visual attention. In M. I. Posner & O. S. M. Marin (Eds.), *Attention and Performance XI* (pp. 85–106). Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*(3), 433–458. <https://doi.org/10.1037/0033-295X.96.3.433>
- Eimer, M., & Kiss, M. (2008). Involuntary Attentional Capture is Determined by Task Set: Evidence from Event-related Brain Potentials. *Journal of Cognitive Neuroscience*, *20*(8), 1423–1433. <https://doi.org/10.1162/jocn.2008.20099>
- Eimer, M., Kiss, M., & Cheung, T. (2010). Priming of pop-out modulates attentional target selection in visual search: Behavioural and electrophysiological evidence. *Vision Research*, *50*(14), 1353–1361. <https://doi.org/10.1016/j.visres.2009.11.001>
- Folk, C. L., & Remington, R. (2006). Top-down modulation of preattentive processing: Testing the recovery account of contingent capture. *Visual Cognition*, *14*(4–8), 445–465. <https://doi.org/10.1080/13506280500193545>
- Fougnie, D., Suchow, J. W., & Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nature Communications*, *3*, 1229. <https://doi.org/10.1038/ncomms2237>
- Geyer, T., Zehetleitner, M., & Müller, H. J. (2010). Positional priming of pop-out: A relational-encoding account. *Journal of Vision*, *10*(2), 3. <https://doi.org/10.1167/10.2.3>
- Hickey, C., Olivers, C., Meeter, M., & Theeuwes, J. (2011). Feature priming and the capture of visual attention: Linking two ambiguity resolution hypotheses. *Brain Research*, *1370*(0), 175–184. <https://doi.org/10.1016/j.brainres.2010.11.025>
- Hillstrom, A. P. (2000). Repetition effects in visual search. *Perception & Psychophysics*, *62*(4), 800–17.
- Horowitz, T. S., & Wolfe, J. M. (1998). Visual search has no memory. *Nature*, *394*(6693), 575–577. <https://doi.org/10.1038/29068>

- Huang, L. (2010). What is the unit of visual attention? Object for selection, but Boolean map for access. *Journal of Experimental Psychology: General*, *139*(1), 162–179. <https://doi.org/10.1037/a0018034>
- Huang, L., Holcombe, A. O., & Pashler, H. (2004). Repetition priming in visual search: episodic retrieval, not feature priming. *Memory & Cognition*, *32*(1), 12–20.
- Huang, L., & Pashler, H. (2007). A Boolean map theory of visual attention. *Psychological Review*, *114*(3), 599.
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10–12), 1489–506.
- Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews. Neuroscience*, *2*(3), 194–203.
- James, W. (1891). *The principles of psychology*. London : Macmillan. Retrieved from <http://archive.org/details/principlesofpsyc02jameuoft>
- Jonikaitis, D., & Theeuwes, J. (2013). Dissociating oculomotor contributions to spatial and feature-based selection. *Journal of Neurophysiology*, *jn.00275.2013*. <https://doi.org/10.1152/jn.00275.2013>
- Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, *4*(4), 219–227.
- Kristjánsson, A. (2006). Simultaneous priming along multiple feature dimensions in a visual search task. *Vision Research*, *46*(16), 2554–70. <https://doi.org/10.1016/j.visres.2006.01.015>
- Krummenacher, J., Grubert, A., & Müller, H. J. (2010). Inter-trial and redundant-signals effects in visual search and discrimination tasks: Separable pre-attentive and post-selective effects. *Vision Research*. <https://doi.org/10.1016/j.visres.2010.04.006>
- Lamy, D., Antebi, C., Aviani, N., & Carmel, T. (2008). Priming of Pop-out provides reliable measures of target activation and distractor inhibition in selective attention. *Vision Research*, *48*(1), 30–41. <https://doi.org/10.1016/j.visres.2007.10.009>
- Lamy, D., & Yashar, A. (2011). Boundary conditions of the components of Priming of Pop-out. *Journal of Vision*, *11*(11), 1315–1315. <https://doi.org/10.1167/11.11.1315>
- Lamy, D., Yashar, A., & Ruderman, L. (2010). A dual-stage account of inter-trial priming effects. *Vision Research*, *50*(14), 1396–1401. <https://doi.org/10.1016/j.visres.2010.01.008>
- Lamy, D., Zivony, A., & Yashar, A. (2011). The role of search difficulty in intertrial feature priming. *Vision Research*, *51*(19), 2099–2109. <https://doi.org/10.1016/j.visres.2011.07.010>
- Li, Z. (1999). Contextual influences in V1 as a basis for pop out and asymmetry in visual search. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(18), 10530–10535.
- Li, Z. (2002). A saliency map in primary visual cortex. *Trends in Cognitive Sciences*, *6*(1), 9–16. [https://doi.org/10.1016/S1364-6613\(00\)01817-9](https://doi.org/10.1016/S1364-6613(00)01817-9)
- Lleras, A., Kawahara, J.-I., Wan, X. I., & Ariga, A. (2008). Intertrial inhibition of focused attention in pop-out search. *Perception & Psychophysics*, *70*(1), 114–131.
- Logan, G. D. (1996). The CODE theory of visual attention: an integration of space-based and object-based attention. *Psychological Review*, *103*(4), 603–49.
- Luce, R. D. (1977). The choice axiom after twenty years. *Journal of Mathematical Psychology*, 215–223.
- Maljkovic, V., & Martini, P. (2005). Implicit short-term memory and event frequency effects in visual search. *Vision Research*, *45*(21), 2831–2846. <https://doi.org/10.1016/j.visres.2005.05.019>
- Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: I. Role of features. *Memory & Cognition*, *22*(6), 657–672. <https://doi.org/10.3758/BF03209251>
- Maljkovic, V., & Nakayama, K. (1996). Priming of pop-out: II. The role of position. *Perception & Psychophysics*, *58*(7), 977–91.

- Meeter, M., & Olivers, C. N. L. (2006). Intertrial priming stemming from ambiguity: A new account of priming in visual search. *Visual Cognition*, *13*(2), 202–222.  
<https://doi.org/10.1080/13506280500277488>
- Moran, R., Zehetleitner, M., Müller, H. J., & Usher, M. (2013). Competitive guided search: Meeting the challenge of benchmark RT distributions. *Journal of Vision*, *13*(8), 24. <https://doi.org/10.1167/13.8.24>
- Müller, H. J., Heller, D., & Ziegler, J. (1995). Visual search for singleton feature targets within and across feature dimensions. *Perception & Psychophysics*, *57*(1), 1–17.
- Müller, H. J., & Krummenacher, J. (2006a). Locus of dimension weighting: Preattentive or postselective? *Visual Cognition*, *14*(4–8), 490–513.
- Müller, H. J., & Krummenacher, J. (2006b). Visual search and selective attention. *Visual Cognition*, *14*(4), 389–410.
- Nelder, J. A., & Mead, R. (1965). A Simplex Method for Function Minimization. *The Computer Journal*, *7*(4), 308–313. <https://doi.org/10.1093/comjnl/7.4.308>
- Olivers, C. N. L., & Hickey, C. (2010). Priming resolves perceptual ambiguity in visual search: Evidence from behaviour and electrophysiology. *Vision Research*, *50*(14), 1362–1371.  
<https://doi.org/10.1016/j.visres.2009.11.022>
- Olivers, C. N. L., & Meeter, M. (2006). On the dissociation between compound and present/absent tasks in visual search: Intertrial priming is ambiguity driven. *Visual Cognition*, *13*(1), 1–28.
- Palmer, E. M., Horowitz, T. S., Torralba, A., & Wolfe, J. M. (2011). What are the Shapes of Response Time Distributions in Visual Search? *Journal of Experimental Psychology: Human Perception and Performance*, *37*(1), 58–71. <https://doi.org/10.1037/a0020747>
- Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. <https://doi.org/10.1016/j.jneumeth.2006.11.017>
- Peirce, J. W. (2009). Generating Stimuli for Neuroscience Using PsychoPy. *Frontiers in Neuroinformatics*, *2*. <https://doi.org/10.3389/neuro.11.010.2008>
- Rangelov, D., Müller, H. J., & Zehetleitner, M. (2011). Independent dimension-weighting mechanisms for visual selection and stimulus identification. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(5), 1369–1382. <https://doi.org/10.1037/a0024265>
- Rangelov, D., Müller, H. J., & Zehetleitner, M. (2012). The multiple-weighting-systems hypothesis: Theory and empirical support. *Attention, Perception, & Psychophysics*, *74*(3), 540–552.  
<https://doi.org/10.3758/s13414-011-0251-2>
- Rangelov, D., Müller, H. J., & Zehetleitner, M. (2013). Visual search for feature singletons: Multiple mechanisms produce sequence effects in visual search. *Journal of Vision*, *13*(3), 22.  
<https://doi.org/10.1167/13.3.22>
- Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, *86*(3), 446–461.
- Rauschenberger, R., & Yantis, S. (2001). Masking unveils pre-amodal completion representation in visual search. *Nature*, *410*(6826), 369–372. <https://doi.org/10.1038/35066577>
- Rosenholtz, R., Huang, J., & Ehinger, K. A. (2012). Rethinking the Role of Top-Down Attention in Vision: Effects Attributable to a Lossy Representation in Peripheral Vision. *Frontiers in Psychology*, *3*.  
<https://doi.org/10.3389/fpsyg.2012.00013>
- Shin, E., Wan, X. I., Fabiani, M., Gratton, G., & Lleras, A. (2008). Electrophysiological evidence of feature-based inhibition of focused attention across consecutive trials. *Psychophysiology*, *45*(5), 804–811.  
<https://doi.org/10.1111/j.1469-8986.2008.00679.x>
- Song, J.-H., & Nakayama, K. (2006). Role of focal attention on latencies and trajectories of visually guided manual pointing. *Journal of Vision*, *6*(9), 982–995. <https://doi.org/10.1167/6.9.11>
- Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics*, *51*(6), 599–606.

- Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychologica*, 135(2), 77–99. <https://doi.org/10.1016/j.actpsy.2010.02.006>
- Treisman, A. M. (1988). Features and objects: the fourteenth Bartlett memorial lecture. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, 40(2), 201–37.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97–136.
- Treisman, A. M., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance*, 16(3), 459–478. <https://doi.org/10.1037/0096-1523.16.3.459>
- Tseng, Y.-C., Glaser, J. I., Caddigan, E., & Lleras, A. (2014). Modeling the effect of selection history on pop-out visual search. *PLoS One*, 9(3), e89996. <https://doi.org/10.1371/journal.pone.0089996>
- van den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences of the United States of America*, 109(22), 8780–8785. <https://doi.org/10.1073/pnas.1117465109>
- Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, 11(1), 192–196. <https://doi.org/10.3758/BF03206482>
- Wan, X., & Lleras, A. (2010). The effect of feature discriminability on the intertrial inhibition of focused attention. *Visual Cognition*, 18(6), 920–944. <https://doi.org/10.1080/13506280903507143>
- White, A. L., Lunau, R., & Carrasco, M. (2014). The attentional effects of single cues and color singletons on visual sensitivity. *Journal of Experimental Psychology: Human Perception and Performance*, 40(2), 639–652. <https://doi.org/10.1037/a0033775>
- Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, 1(2), 202–238.
- Wolfe, J. M. (2007). Guided Search 4.0: Current Progress with a model of visual search. In W. D. Gray (Ed.), *Integrated models of cognitive systems* (pp. 99–119). Oxford University Press.
- Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: an alternative to the feature integration model for visual search. *Journal of Experimental Psychology. Human Perception and Performance*, 15(3), 419–33.
- Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nat Rev Neurosci*, 5(6), 495–501. <https://doi.org/10.1038/nrn1411>
- Yashar, A., & Lamy, D. (2011). Refining the dual-stage account of intertrial feature priming: Does motor response or response feature matter? *Attention, Perception, & Psychophysics*. <https://doi.org/10.3758/s13414-011-0182-y>

## Figure captions

*Figure 1.* Probability density functions for search (top row) and post-search durations (middle row) as well as the total RTs (bottom row). Several mechanisms could yield negative search slopes: (1) *slow pop-out* – finding the target in sparse displays is, on average, slower, while the variability of the search duration remains comparable to that for dense displays; (2) *different pop-out* – finding the target in sparse displays is more variable compared to dense displays; and (3) *no pop-out* – the probability of finding the target in the first round of selection is lower for sparse (i.e.,  $p_1 < 1$ ) than for dense displays ( $p_1 \approx 1$ ). Total RTs are a combination of the time it takes to find the target (search duration) and post-search processes, i.e., the observed RT distribution is a convolution of the distributions of the search and post-search durations. See the online article for the color version of this figure.

*Figure 2.* (A) Illustration of the stimulus displays used in Experiments 1 and 2 along with the respective exposure durations. In the cued condition, the target appeared invariably at the pre-cued location. In the variable-distractors conditions, the target was always either red amongst green distractors or green amongst red distractors. In the fixed-distractors conditions, distractors were always blue. (B) Illustration of the stimulus displays used in Experiments 3 and 4. In Experiment 3, the target was either a red singleton amongst green distractors or a green singleton amongst red distractors. In Experiment 4, the target was either a left-tilted singleton amongst right-tilted distractors or right-tilted singleton amongst left-tilted distractors. Participants had to report the letter inside the target singleton (E vs. mirrored E). Unlike Experiments 1 and 2, only variable distractors were used. See the online article for the color version of this figure.

*Figure 3.* Distribution of RTs on uncued trials for a representative participant (top panel) in Experiment 1 (target discrimination). Mean correct RTs across different cumulative probabilities (.1–.9), with PoP magnitude (different target – same target), for sparse and dense displays, separately for variable and fixed distractors on uncued trials (central panel) and cued trials (lower panel). To better illustrate RT variability, probabilities above the median (.5) were plotted as the difference from 1 (e.g.,  $p = .9$  is plotted as  $1 - .9 = .1$ ). Whiskers denote 95% CI. See the online article for the color version of this figure.

*Figure 4.* Comparison of the observed RTs in Experiment 1 (filled symbols) with the predictions of the best-fitting models (empty symbols). Individual symbols denote RT percentiles (10<sup>th</sup>, 30<sup>th</sup>, 50<sup>th</sup>, 70<sup>th</sup>, and 90<sup>th</sup>) from the fastest (10<sup>th</sup> percentile) to the slowest RTs (90<sup>th</sup> percentile). For cued trials, predictions by the best-fitting Sub1 model are depicted. For uncued trials, predictions by the best-fitting ‘hybrid’ model are shown. Gray “+” and “x” denote predicted RTs for models that used the Gaussian and Wald distributions, respectively, to model search duration. The difference between the observed and predicted RT percentiles reflect goodness-of-fit, with smaller differences indicating better fits. See the online article for the color version of this figure.

*Figure 5.* Probability density distributions of simulated search durations for three pop-out models (‘hybrid’, ‘different pop-out’, and ‘no pop-out’) and varying degrees of inter-trial variability ( $\sigma$ ). Since the empirical RT distributions showed no differences between fixed- and variable-distractor conditions for

dense displays, only one probability density distribution is depicted for such displays. See the online article for the color version of this figure.

*Figure 6.* Distribution of RTs on uncued trials for a representative participant (top panel) in Experiment 2 (detection task). Mean RTs across different cumulative probabilities (.1–.9), together with the PoP magnitude (different target – same target), for correct responses to sparse and to dense displays, separately for variable and fixed distractors on uncued trials (central panel) and cued trials (lower panel). Conventions as in Figure 3. See the online article for the color version of this figure.

*Figure 7.* Distribution of RTs on uncued trials for a representative participant (top panel) in Experiments 3 and 4 (left and right panels, respectively). Mean correct RTs and PoP magnitude across different cumulative probabilities (.1–.9), separately for uncued trials (central panel) and cued trials (lower). Conventions as in Figure 3. See the online article for the color version of this figure.

*Figure 8.* Comparison of the observed RTs in Experiments 3–4 (filled symbols) with the predictions of the best-fitting models (empty symbols). For cued trials, predictions of the best-fitting Sub1 model are depicted. For uncued trials, predictions of the best-fitting ‘no pop-out’ model (Exp. 3) and, respectively, the best-fitting ‘hybrid’ model (Exp. 4) are shown. Other conventions as in Figure 4. See text for details. See the online article for the color version of this figure.

Figure 1

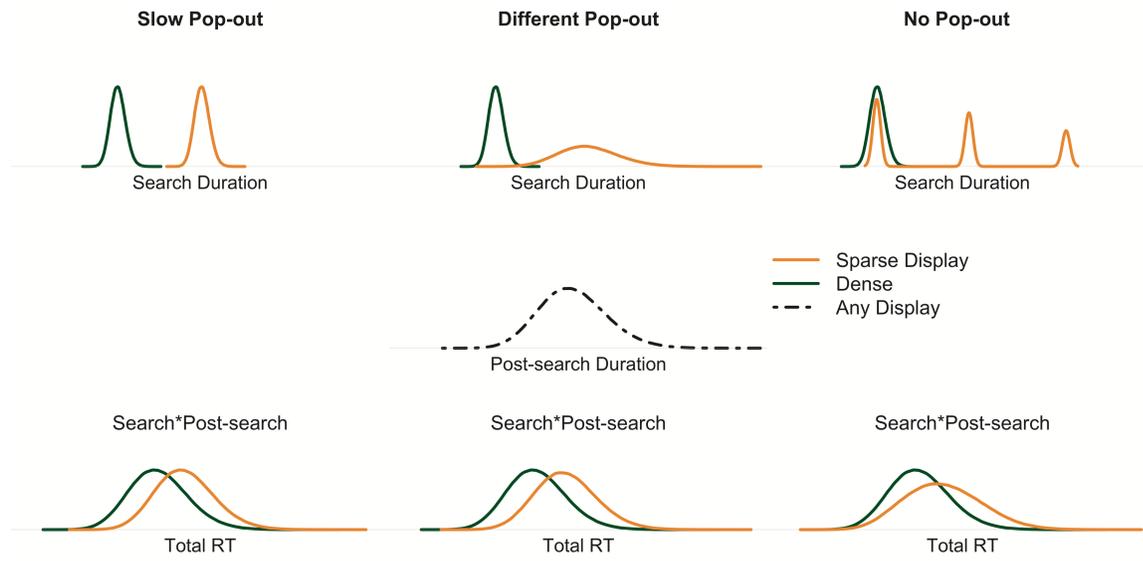


Figure 2

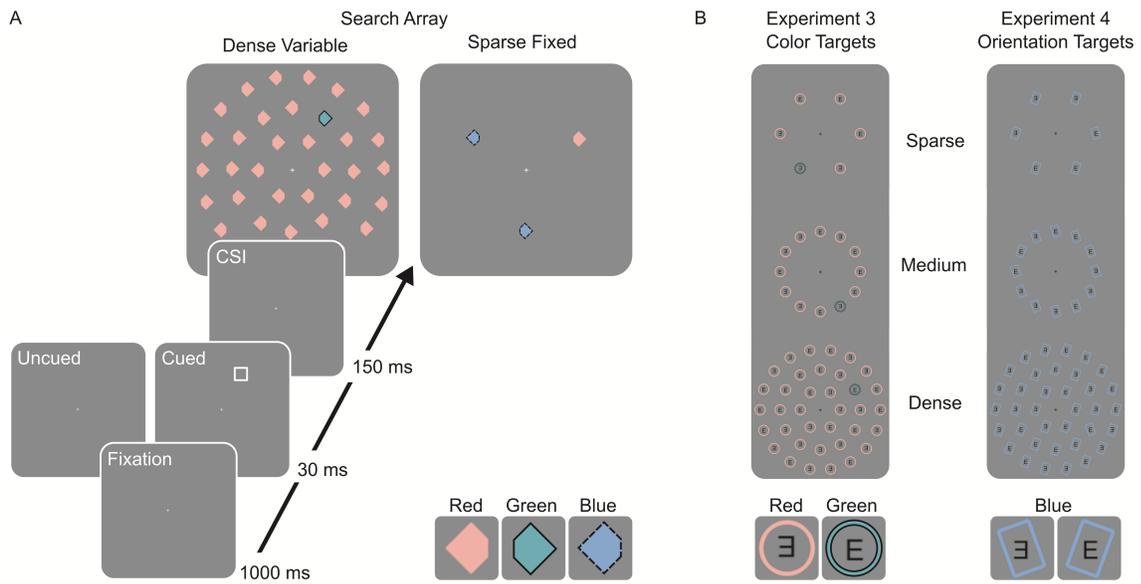


Figure 3

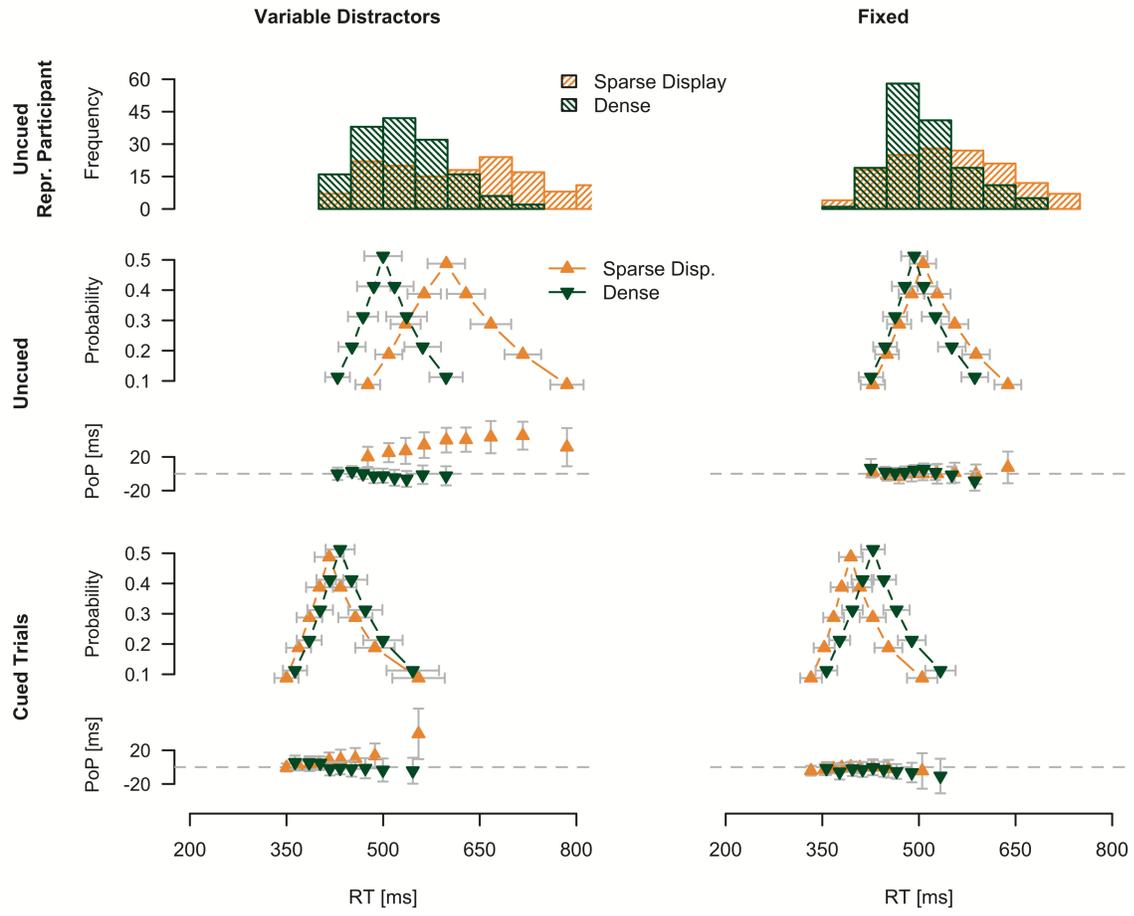


Figure 4

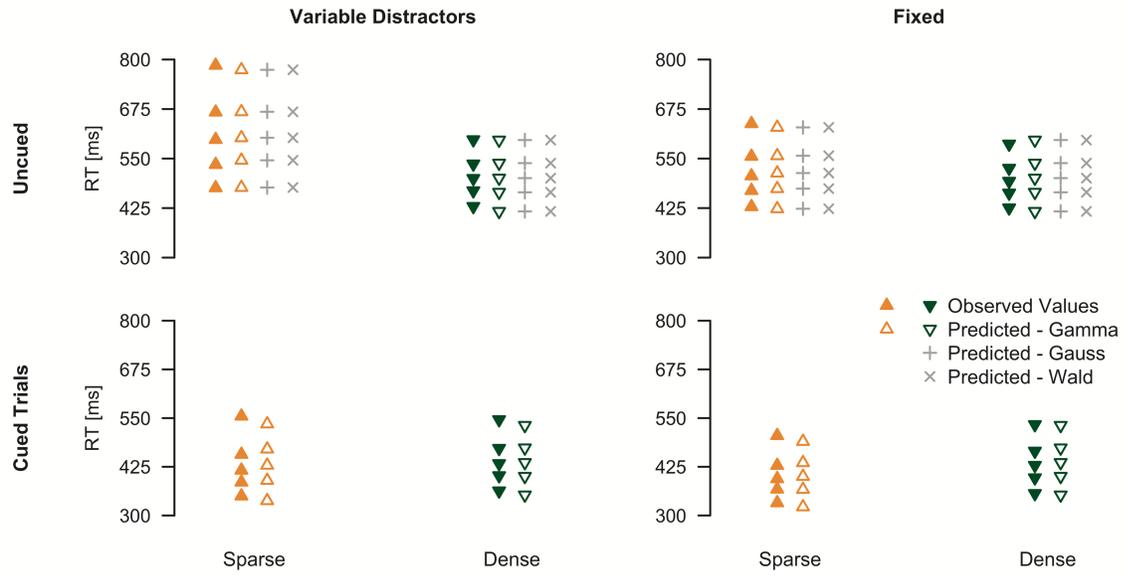


Figure 5

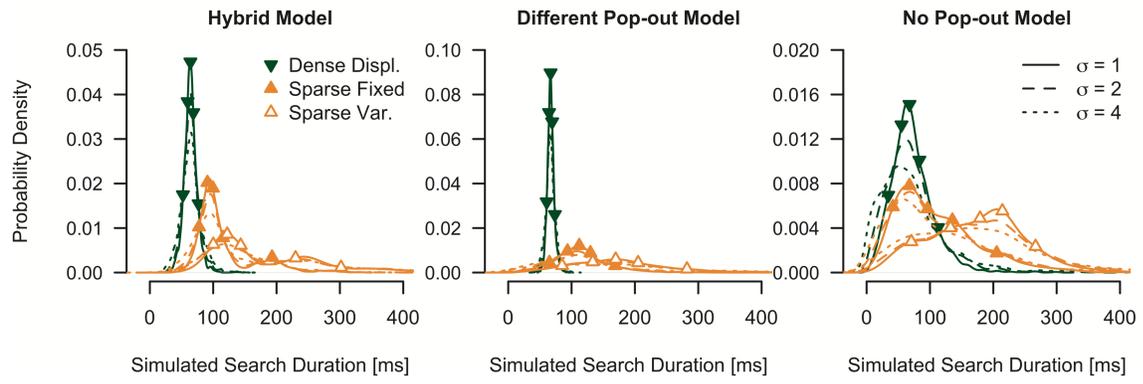


Figure 6

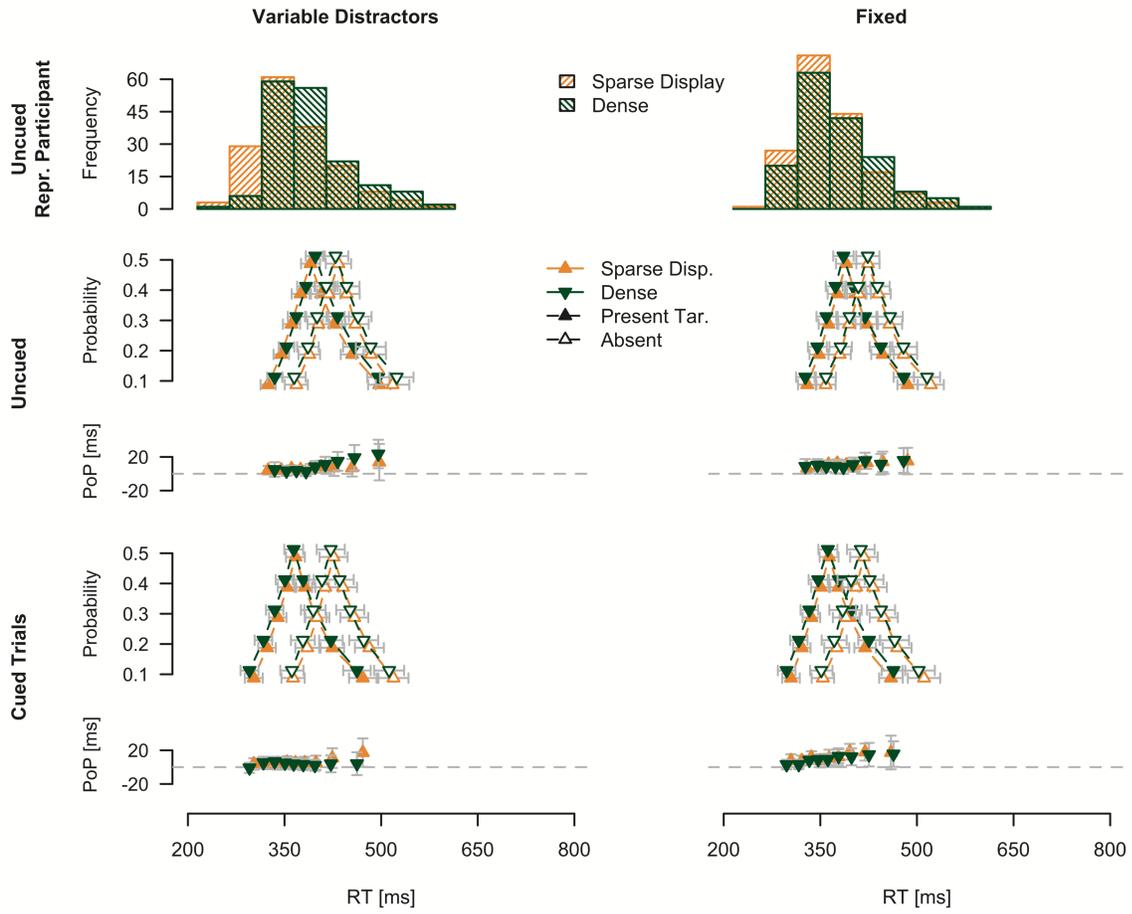


Figure 7

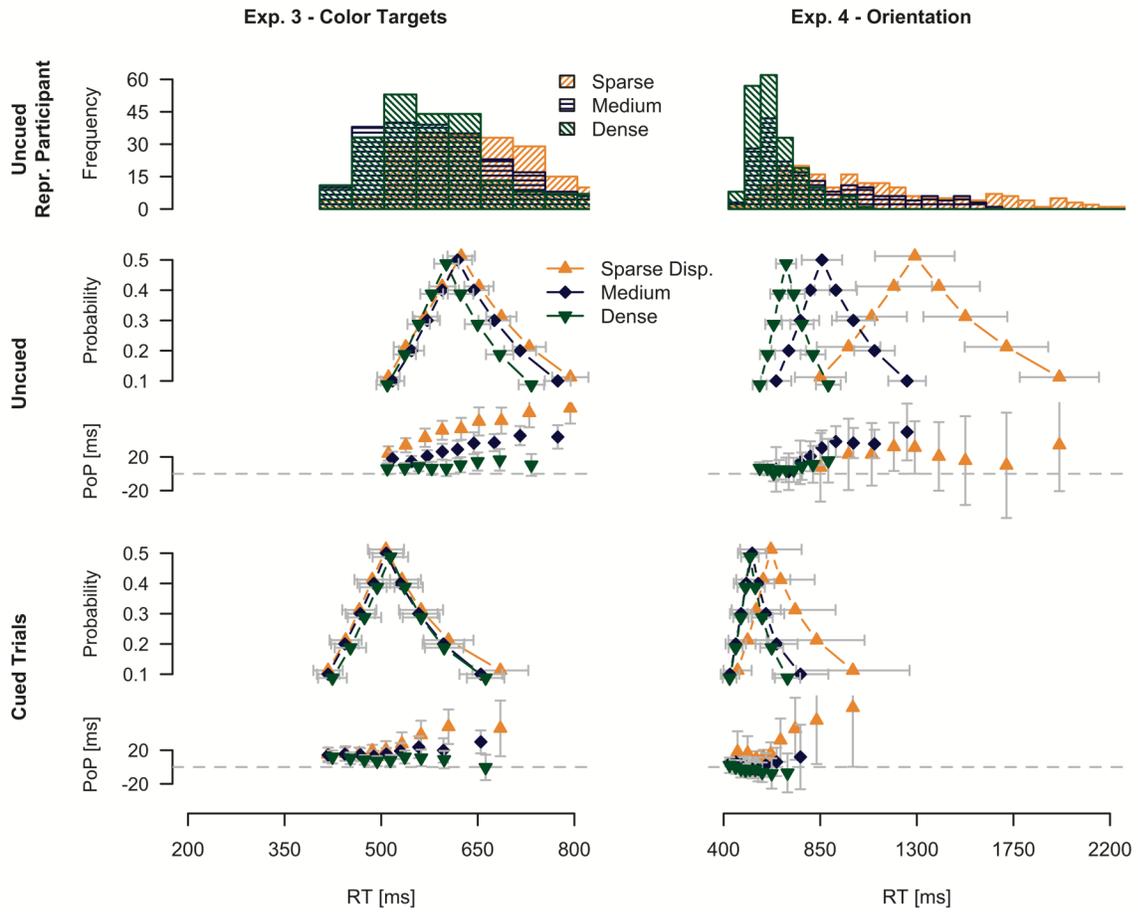


Figure 8

