Journal of Vision (2020) 20(13):11, 1–26

Explaining the effects of distractor statistics in visual search

Joshua Calder-Travis
Department of Experimental Psychology, University of Oxford, Oxford, UK
Department of Psychology, New York University, New York, NY, USA

Wei Ji Ma
Department of Psychology, New York University, New York, NY, USA
Center for Neural Science, New York University, New York, NY, USA

Visual search, the task of detecting or locating target items among distractor items in a visual scene, is an important function for animals and humans. Different theoretical accounts make differing predictions for the effects of distractor statistics. Here we use a task in which we parametrically vary distractor items, allowing for a simultaneously fine-grained and comprehensive study of distractor statistics. We found effects of target-distractor similarity, distractor variability, and an interaction between the two, although the effect of the interaction on performance differed from the one expected. To explain these findings, we constructed computational process models that make trial-by-trial predictions for behavior based on the stimulus presented. These models, including a Bayesian observer model, provided excellent accounts of both the qualitative and quantitative effects of distractor statistics, as well as of the effect of changing the statistics of the environment (in the form of distractors being drawn from a different distribution). We conclude with a broader discussion of the role of computational process models in the understanding of visual search.

Introduction

Animals and humans constantly engage in visual search, the process of detecting, locating, or identifying target objects in an image or scene (Eckstein, 2011). The golden eagle looking for a hare, the hare looking for predators, and the human looking for bread in a supermarket are all examples of visual search. A great deal of research over more than 50 years has aimed to build a mechanistic understanding of visual search (Neisser, 1964; Treisman & Gelade, 1980; Estes & Taylor, 1964). Such an understanding would not just be important in its own right but would contribute to our knowledge of the representations and algorithms used to perceive and act in the world. Additionally, it may have direct application in critical visual search situations, such as baggage scanning at the airport (Schwaninger, 2005).

Any satisfactory mechanistic account needs to explain key qualitative patterns in visual search data, such as those identified by Treisman and Gelade (1980). Treisman and Gelade (1980) suggested that there is a categorical difference between two kinds of search, which they called feature search and conjunction search. In feature search, the target can be distinguished from the distractor items using a single feature such as color. In conjunction search, there is no single feature that is present in the targets and absent in all distractors. Instead, two or more features are required to uniquely identify an item as the target. Feature search was highly efficient: Increases in the number of items in the display (the total number of targets and distractors) had little effect on the time taken to respond when a target was present. By contrast, efficiency in conjunction search was lower, and as set size increased, Treisman and Gelade (1980) found that response time increased markedly.

Duncan and Humphreys (1989) contested the idea of a dichotomy between feature and conjunction search and suggested that search efficiency varies along a continuum. They claimed that the similarity of the target to the distractors decreases search efficiency, variability of the distractors decreases search efficiency, and these quantities interact such that variability of distractors is most harmful when distractors are very similar to the target. One of the key motivations for their departure from the idea of a strict dichotomy was that, over a series of experiments using a wide range of stimuli, they could not find a consistent set of properties that could be identified as features. Duncan and Humphreys's (1989) account can still accommodate the findings of Treisman and Gelade (1980), if we claim

Citation: Calder-Travis, J., & Ma, W. J. (2020). Explaining the effects of distractor statistics in visual search. *Journal of Vision*, *20*(13):11, 1–26, https://doi.org/10.1167/jov.20.13.11.

https://doi.org/10.1167/jov.20.13.11 Received January 21, 2020; published December 17, 2020 ISSN 1534-7362

that Treisman and Gelade (1980) only explored part of the stimulus space spanned by target-distractor similarity and distractor variability; the appearance of a dichotomy would then stem from the use of stimuli drawn from two distinct clusters in this space.

While Duncan and Humphreys (1989) made a valuable contribution in suggesting that performance likely lies on a continuum, their exploration of this claim was necessarily limited by the stimuli that they used. Using letters and joined lines, they could not parametrically vary properties of these items along an easily quantified dimension. Instead, they could only make comparisons between specific sets of stimuli, which differed qualitatively in terms of target-distractor similarity and distractor variability. For example, in one experiment, Duncan and Humphreys (1989) studied search for an upright "L" in 90° clockwise or counterclockwise rotated "L"s. In the low-distractor variability condition, only one distractor type was used. In the high-distractor variability condition, both distractor types were used. Search was more efficient in the low-variability condition, supporting their suggested pattern of effects. However, it could be that Duncan and Humphreys (1989), while describing performance in a larger area of stimulus space than Treisman and Gelade (1980), missed areas of the space, along with distinctive qualitative effects.

Other visual search researchers have used stimulus items that can easily be parametrically varied (Rosenholtz, 2001; Palmer, Ames, & Lindsey, 1993; Palmer, 1994; Cameron, Tai, Eckstein, & Carrasco, 2004; Ma, Navalpakkam, Beck, van den Berg, & Pouget, 2011; Palmer, Verghese, & Pavel, 2000). For example, Cameron et al. (2004) used oriented Gabor patches, with the target being defined as a Gabor of particular orientation. Such stimuli make it possible to operationalize target-distractor difference precisely as the difference between the target orientation and the mean distractor orientation and to operationalize distractor variability as the variance of the distractors.

Parametric stimuli also allow us to apply formal models of behavior and cognition (Ma et al., 2011; Rosenholtz, 2001; Palmer et al., 2000). Computational modeling could provide a simple unified explanation of the full range of patterns observed in behavior, and could allow us to infer the precise mechanisms underlying visual search. Signal detection theory (SDT) is the leading framework for building such formal models. SDT has at its core the idea that observers only receive noisy representations of items in the stimulus (Palmer et al., 1993; Green & Swets, 1966). The observer combines these noisy representations into a single variable and then applies a threshold to this variable. If the variable exceeds a threshold, the observer reports the target is present, if not, absent.

SDT models predict graded changes in performance. Consider what happens as distractors become less similar to the target. The chance that noise will cause them to be confused with the target decreases, decreasing the false-alarm rate and therefore potentially increasing performance (Rosenholtz, 2001). In this respect, SDT models may make similar predictions to the claims of Duncan and Humphreys (1989). However, SDT models may also make contrasting predictions. Specifically, when the distractors and target are very similar, increasing distractor variability will spread the distractors out, away from the target. This will decrease the chance that they will be confused with the target, decreasing the false alarm rate and potentially increasing performance (Rosenholtz, 2001; see Figure 1). Note that this mechanism may be strongest on trials when the target is absent: When the target is present, the item that looks most like the target may be the target itself. Therefore, spreading distractors out may not make a difference to the item that appears closest to the target.

Using process models, such as SDT models, we can generate predictions for the relationship between any stimulus statistic and any behavioral statistic. This is because these models predict behavior on a trial-by-trial basis using the full stimulus, rather than a summary of the stimulus. Hence, we can always ask what the model predicts for stimuli low or high on any particular statistic. By contrast, if our theory is that distractor mean predicts accuracy in a certain manner, it remains completely unclear how other distractor statistics might predict accuracy or how mean might be related to another behavioral statistic such as hit rate. For example, low accuracy could be caused by completely random responding or by always picking the same response. It should be noted that Duncan and Humphreys (1989) developed a detailed account of how parts of a visual scene are grouped and compete for entry into visual short-term memory. They used this account to explain the effects of distractor statistics that they described. In this article, we do not attempt to convert the entirety of their underlying theoretical account into a process model but instead focus on the effects of distractor statistics.

Signal detection theory encompasses a range of approaches to visual search, of which the Bayesian approach is one (Green & Swets, 1966; Rosenholtz, 2001; Palmer et al., 2000). In the Bayesian approach, we assume that the observer computes a very specific single variable from the noisy stimulus representation, namely, the posterior ratio. This is the ratio of the probability that the target is present and the probability that the target is absent, given the observer's measurements (Palmer et al., 2000; Peterson, Birdsall, & Fox, 1954; Ma et al., 2011). We assume that the observer has learned the statistical structure of the task and computes the posterior ratio using this knowledge. This assumption results in a highly constrained model that has been shown to fit behavior well in a range of visual

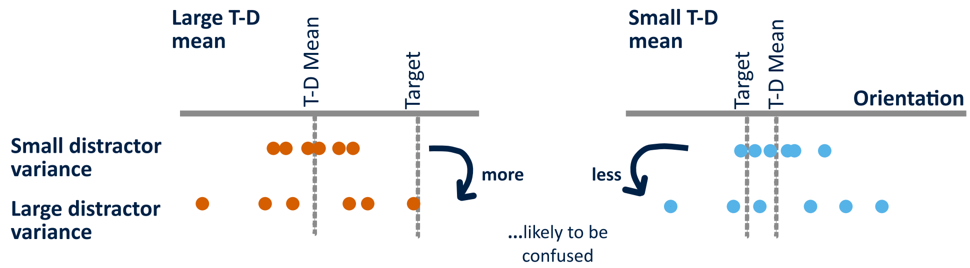

Figure 1. Interaction between target-to-distractor mean difference (T-D mean) and distractor variance on the probability of a confusing distractor (Rosenholtz, 2001). When T-D mean is large, then increasing distractor variance makes a distractor that closely resembles the target more likely. On the other hand, when T-D mean is small, increasing distractor variance actually makes it less likely that there is a distractor that is highly similar to the target.

search tasks (Ma et al., 2011; Mazyar, van den Berg, Seilheimer, & Ma, 2013, Mazyar, van den Berg, & Ma, 2012).

The present work has several goals. (a) We aim to describe the effect of target-distractor difference and distractor variability across the full stimulus space, comparing the results with the claims of Duncan and Humphreys (1989) and with SDT ideas (see Figure 1). (b) We examine whether a Bayesian optimal-observer model accounts for patterns in the data and whether the Bayesian model is consistent with the claims of Duncan and Humphreys (1989). (c) We examine whether simpler heuristic models can also account for the observed patterns.

Rosenholtz (2001) conducted closely related work aiming to test the idea that, under SDT, increasing distractor variability might actually improve performance when the target-distractor difference is initially low. The task used involved asking participants to search for a line with a particular orientation. Distractors took one of a small number of values (e.g., 30°, 50°, or 70°). The number of distractors at each value was manipulated to increase variance while only moving distractors further from the target orientation. In a display with a large number of items, Rosenholtz (2001) found evidence that increasing variability while moving distractors further from the target can harm performance. An SDT model could not explain this pattern. In a display of only eight items, no change in performance was detected. However, this pattern was still qualitatively inconsistent with the SDT model, which predicted that increasing variance by moving distractors away from the target should improve performance. Our study provides a new test of SDT models, one in which distractors can take any value, rather than a small number of values. Mazyar et al. (2013) also conducted an experiment closely related to ours but focused on the effects of number of items on precision, rather than the effects of distractor statistics on performance. Our study complements the work of Mihali and Ma (2020), who explored the effect of distractor statistics in different kinds of visual search tasks, while also exploring the effect of memory and stimulus spacing. We did not explore these factors here, but we did vary the distractor "environment": We compared two conditions in which the distractors were drawn from different distributions.

A subtle but sufficiently important point to warrant discussion at the outset is that we focus on one kind of distractor statistics. We explore the effects of statistics of sampled distractors. This contrasts with examining the effects of population statistics—the statistics of the distributions from which distractors are drawn. In the "Theory of Visual Selection" described by Duncan and Humphreys (1989, p. 444), both sample and population distractor statistics have a role. We focused on sample distractor statistics because experimental study of these effects is more feasible. To study the effects of population distractor statistics, participants would need to be trained on many different distractor distributions. As mentioned, we only trained participants on two distributions here. Both distractor distributions have a mean value equal to the target but differ in variance. Hence, we cannot study the effects of population distractor mean and can only study the effects of population distractor variance in the coarsest manner. To anticipate our results, it is unclear whether participants even learn the difference between these two distributions. Avoiding these important but technically challenging questions, we focus on exploring the full stimulus space, as characterized by sample distractor statistics.

Experimental methods

Participants

Fourteen participants were recruited consistent with a predetermined recruitment schedule (see Appendix A for age, gender, and handedness information). One participant was excluded from all analysis below as they were unable to complete all sessions. The study procedure was approved by the Institutional

Review Board of New York University and followed the Declaration of Helsinki (with the exception of registration on a public database prior to data collection). All participants gave informed consent.

Apparatus

Stimuli were presented on an LCD monitor at a 60-Hz refresh rate, with 1,920 × 1,080 resolution, and a viewable screen size of 51 × 29 cm. Stimuli were presented using Psychtoolbox in MATLAB on Windows (Brainard, 1997; Pelli, 1997; Kleiner, Brainard, & Pelli, 2007). A chin rest ensured participants viewed the stimuli at approximately 60 cm from the screen. Eye-tracking data was collected for future exploratory analysis. However, to date, these data have not been analyzed in any form.

Stimuli

Each item in the stimulus was a Gabor patch with a standard deviation of the Gaussian window of 0.25 degrees of visual angle (dva) and with 1.79 cycles per dva. Items were presented on a gray background (half the maximum of the RGB range). At the center of the Gaussian window, the peaks and troughs of the sinusoid were represented as grays at the maximum and minimum of the RGB range. The phase of the Gabors was set so there was a peak in the sinusoid at the center of the Gaussian window.

Six item locations were determined at the start of the experiment. These locations were equally spaced around the circumference of an imagined circle. Therefore, in the plane of the screen, Gabors were 60° apart from each other (the first patch was at 90° from vertical). Each location was at 4.99 dva from the imaginary line connecting the participant to the center of the screen. In trials where there were fewer than six items to display, a subset of the locations was randomly selected.

Trial procedure

Each trial began with a fixation cross presented at the center of the screen for 500 ms. Participants were presented with two, three, four, or six Gabor patches for 100 ms and asked to report the presence or absence of a target. Participants had unlimited time to respond "target present" with the "j" key, or "target absent" with the "f" key. The target was a Gabor patch oriented at 45° clockwise from vertical. Targets were present on 50% of trials. All other Gabors in the display, named distractors, were drawn from a probability distribution that depended on the distractor environment (described below).

Following a response, participants received feedback in the form of an orange or light blue fixation cross. This color-coded fixation cross was presented for 700 ms. Following a trial, there was a delay of at least 100 ms for setting up the eye tracker. The next trial would not begin until all keys had been released.

Structure of the experiment

There were two distractor environments, and these determined the probability distributions from which distractors were drawn (Figure 2). In the *uniform distractor environment*, distractors were drawn from a uniform distribution, and hence any orientation was equally likely. In the *concentrated distractor environment*, distractors were drawn from a von Mises distribution centered on the target orientation and with concentration parameter 1.5 (approximately equivalent to a wrapped normal distribution with an unwrapped standard deviation of 58°). The von Mises distribution is similar to the normal distribution but is the appropriate choice for circular variables (i.e., orientation). The von Mises provides a probability distribution over orientations from –180° to 180°. However, a Gabor patch with orientation θ° is identical to one with orientation $\theta + 180^\circ$. We deal with this in the usual way, by halving angles drawn from the von Mises distribution, so that this distribution only covers –90° to 90°. Each block either contained trials from the uniform distractor environment or trials from the concentrated distractor environment.

The experiment took place over four separate 1-hr sessions. In each session, there were eight test blocks of 64 trials. Uniform and concentrated distractor environment blocks were ordered as AABBBBAA. Whether "A" corresponded to the uniform or concentrated distractor environment was determined randomly at the beginning of each session.

Training

At the beginning of each session, the participant was presented with an image of the target. Beside this image was a series of example distractors from the uniform distractor environment, followed by a series of example distractors from the concentrated distractor environment. At the beginning of the first session, the participant also completed four training blocks. At the beginning of subsequent sessions, they completed two training blocks. Each training block contained 40 trials. Uniform and concentrated blocks alternated, with the first block being selected randomly at the beginning of each session (matching the first test block). During test blocks, every time the distractor environment switched, the participant was presented with a refresher, in the

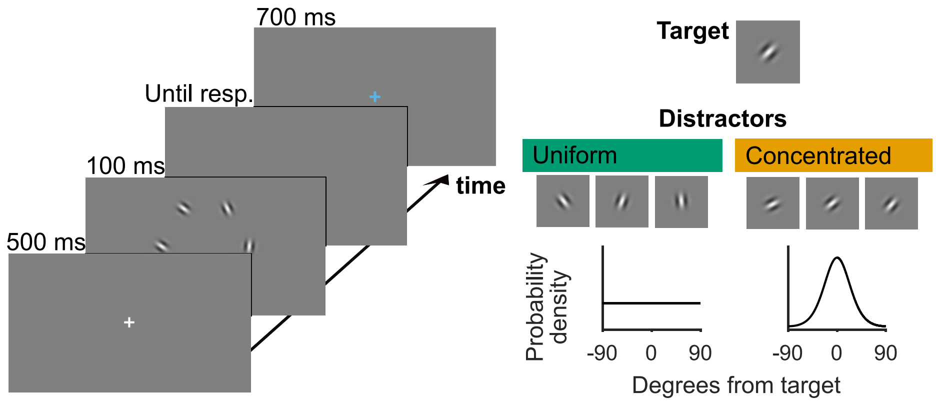

Figure 2. Participants performed a visual search task in which they had to report the presence or absence of a target (a Gabor oriented at 45° clockwise from vertical) in a briefly presented display. The display contained between two and six items. There were two distractor environments, one in which all distractor orientations were equally likely and one in which distractor orientations were more likely to be close to the target orientation. The plots of distractor distributions in the two conditions are accurate, while experiment "screenshots" are for illustration and not to scale.

form of another series of example distractors drawn from the upcoming distribution.

Analysis

Throughout the article, unless otherwise stated, we analyze the effect of distractor sample statistics. That is, on each trial, we calculate distractor statistics using the distractors that were actually presented. Throughout the article, target-to-distractor mean difference (T-D mean) will be used to refer to the absolute difference between the circular mean of the distractors and the target orientation. Distractor variance will refer to the circular variance, which ranges from 0 (all distractors the same) to 1 (e.g., two perpendicular distractors). Definitions of circular mean and variance are provided by Berens (2009). Minimum target-distractor difference (min T-D difference) refers to the absolute difference between the target orientation and the distractor orientation closest to the target orientation (Mazyar et al., 2012). For all circular statistics we used the CircStat toolbox (Berens, 2009). Prior to computation of circular statistics, we double all orientations to compensate for the halving of orientations drawn from the von Mises distribution as discussed above. In all plots, we map orientations (including T-D mean and min T-D difference) back to physical orientation.

In order to test the reliability of observed trends, we performed logistic regressions using distractor statistics as predictors and hits, false alarms (FA), or accuracy as outcome. (We included a constant as a predictor in each logistic regression.) We compared the fitted regression slopes to zero across participants. Prior to running the regression, we z-scored the predictors. Centering the variables allows interpretation of a main effect in the presence of an interaction, as the effect of the predictor at the mean value of all other predictors (Afshartous & Preston, 2011). We provide adjusted p-value significance thresholds, using the Bonferroni correction, to account for the number of regression slopes compared to zero in each individual regression analysis. As a measure of effect size, we computed the one-sample variant of Cohen's d (Cohen, 1988),

$$d = \frac{\mu}{\sigma},$$

where μ is the mean of the beta values being compared to zero, and σ is the estimated population standard deviation.

For this analysis (and not for computational modelling below), data from trials with only two Gabor items were excluded. The reason for this is that when there are only two items and one of them is a target, there is only one distractor, and the idea of distractor variability does not make sense. Throughout the article, unless labeled, plots reflect data from trials with three, four, and six items.

Plots

In order to visualize the effect of distractor statistics, we binned these variables. Specifically, we used quantile binning, separately for the data from each participant, and separately for each series in a plot. We took this approach as distractor statistic distributions can be highly nonuniform (Figure 3), and quantile binning ensures a reasonable number of data points in each bin. In order to determine where on the x-axis to plot a bin, we computed for each participant the average value in each bin and then averaged these across participants.

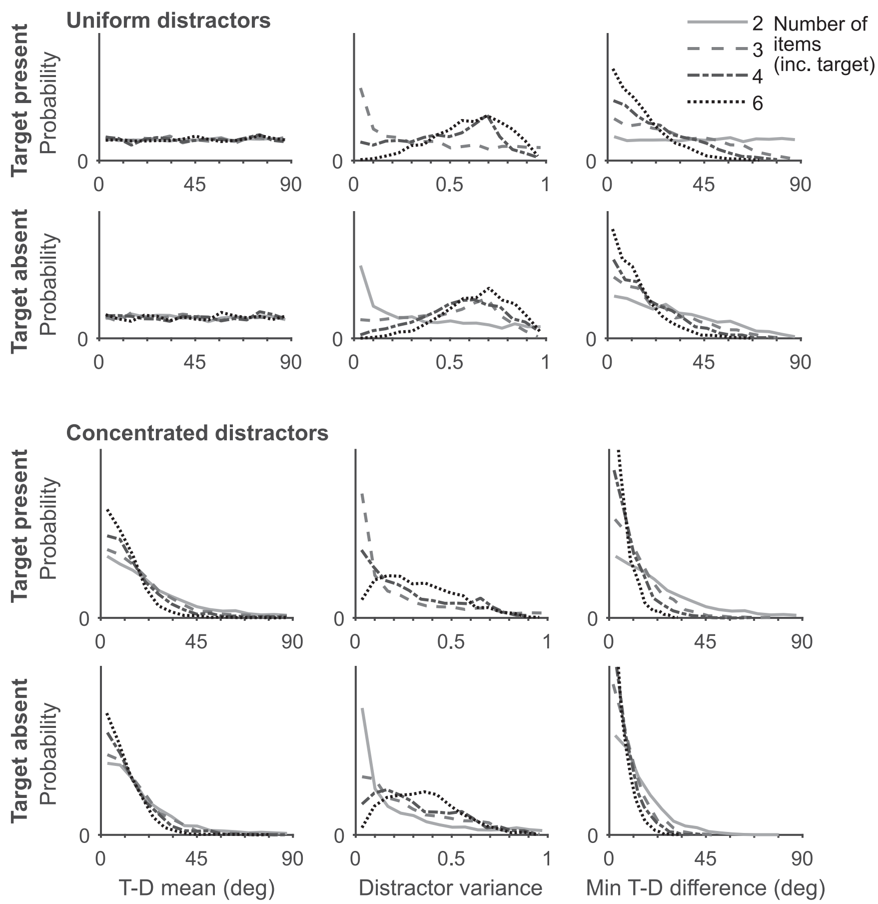

Figure 3. The distributions of distractor statistics, separately for the cases of two, three, four, and six items in the display (including the target). The area under the curves in all 12 plots is the same. Note that these distributions are determined by stimuli properties and are completely independent of participant behavior. Here and throughout the article, target-to-distractor mean difference (T-D mean) refers to the absolute difference between the circular mean of the distractors and the target orientation, and distractor variance refers to the circular variance of the distractors. Minimum target-distractor difference (min T-D difference) refers to the absolute difference between the target orientation and the distractor closest to this orientation. Data from all participants combined are shown. We can see that distractor statistic distributions are highly nonuniform. This is the motivation for, in all other plots than this one, quantile binning distractor statistics.

The location of a bin on the y-axis was determined by the mean value of the outcome variable across participants. Unless stated, error bars represent ±1 standard error of the mean.

Data and code availability

Anonymized data, together with all experiment and analysis code written for the study, will be made available upon publication at doi:10.17605/OSF.IO/NERZK.

Experimental results

We first examine the patterns in the data and turn to computational modeling of these patterns in later sections.

In an initial set of analyses, we focused on testing the pattern of effects suggested by Duncan and Humphreys (1989): T-D mean should improve performance, distractor variance should harm performance, and the harmful effect of distractor variance should be greatest at low T-D mean. For each participant, we conducted logistic regressions with target-to-distractor mean difference (T-D mean), distractor variance, and their interaction as predictors, and accuracy, hit rate, or false alarm (FA) rate as outcome. The resulting regression coefficients reflect the strength of the relationship between the predictor and the outcome. We compared these coefficients to zero (Figure 4, and Table 1). T-D mean, distractor variance, and their interaction significantly predicted accuracy. At the average value of distractor variance, increasing T-D mean increased accuracy, while at the average T-D mean, increasing distractor variance decreased accuracy.

Outcome	Predictor	*t*-value	Effect size (*d*)	*p*-value
Accuracy	T-D mean	5.5	1.5	1.3×10^{-4}
	Distractor variance	−3.1	−0.86	9.4×10^{-3}
	Mean-variance interaction	−3.2	−0.89	7.5×10^{-3}
Hit rate	T-D mean	−5.5	−1.5	1.3×10^{-4}
	Distractor variance	3.5	0.97	4.5×10^{-3}
	Mean-variance interaction	2.2	0.61	0.049
FA rate	T-D mean	−8.0	−2.2	4.0×10^{-6}
	Distractor variance	4.2	1.2	1.2×10^{-3}
	Mean-variance interaction	6.2	1.7	4.5×10^{-5}

Table 1. The effect of T-D mean, distractor variance, and their interaction, in the absence of additional predictors. When no other predictors are included in the model, all three variables have a significant effect on accuracy, hit rate, and FA rate, with the exception of the effect of the interaction on hit rate. Bonferroni-corrected *p*-value criterion 0.017.

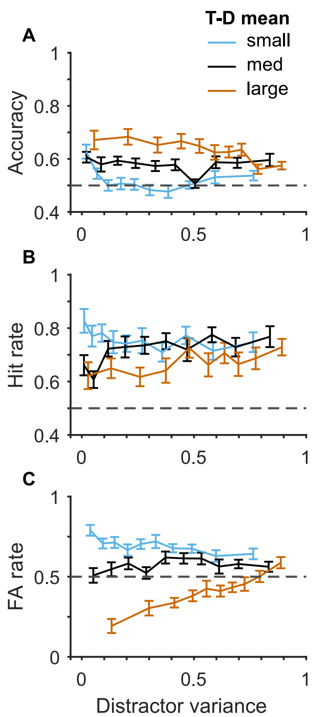

Figure 4. The effect of T-D mean and distractor variance on behavior. There was a particularly clear interaction effect on FA rate, consistent with a signal detection theory account. For the plot, T-D mean was divided into three bins, participant by participant. The average edges between bins were at 12° and 36°. Data from trials with three, four, and six items were used for plotting.

The two interacted such that at large T-D mean, the relationship between distractor variance and accuracy was more negative. This finding contradicts the idea of Duncan and Humphreys (1989) that increasing distractor variability would have relatively little effect on performance when target and distractors are very different from each other (high T-D mean). Whilst While these effects are significant, they are difficult to observe directly from the plot (particularly the effect of distractor variance; Figure 4A). The small T-D mean series appears to exhibit a "U" shape, with the lowest accuracy values at a distractor variance of about 0.35. If real, this effect represents a systematic deviation from a logistic relationship, an assumption of using logistic regression. Therefore, the result of this analysis should be interpreted with caution.

By contrast, T-D mean, distractor variance, and their interaction had a particularly clear effect on FA rate (Figure 4C). At the average value of distractor variance, increasing T-D mean decreased FA rate. At the average T-D mean, as distractor variance increased, FA also increased. There was also an interaction such that distractor variance had the most positive effect for large T-D mean. This pattern is completely consistent with the pattern we would expect from a SDT perspective (recall Figure 1). Specifically, for large T-D mean, increasing variance increases the probability of a distractor similar to the target. While with a small T-D mean, increasing variance actually makes a confusing distractor less likely, as distractor orientations are spread out away from the target orientation.

A similar pattern of effects was observed on hit rate, although the pattern is harder to observe in this case (Table 1, Figure 4B). Note that the considerations above, regarding T-D mean and distractor variance interacting to affect the probability of a confusing distractor, do not directly apply in the case of hit rate. The hit rate is calculated using trials on which the target was in fact present; hence, the target may itself be the most similar item to the target (from the perspective of the observer). This may explain why the effects of the distractors are diluted.

Having considered the effect of distractor mean and variance in isolation from other variables, we wanted to

Outcome	Predictor	*t*-value	Effect size (*d*)	*p*-value
Accuracy	T-D mean	1.6	0.46	0.13
	Distractor variance	−1.7	−0.46	0.12
	Mean-variance interaction	0.47	0.13	0.65
	Min T-D difference	3.6	1.0	3.4×10^{-3}
	Environment	−2.0	−0.55	0.073
	Number of items	−4.7	−1.3	4.8×10^{-4}
Hit rate	T-D mean	−0.73	−0.2	0.48
	Distractor variance	0.096	0.027	0.93
	Mean-variance interaction	−1.1	−0.3	0.31
	Min T-D difference	−5.4	−1.5	1.7×10^{-4}
	Environment	−3.0	−0.84	0.011
	Number of items	−1.4	−0.39	0.19
FA rate	T-D mean	−4.5	−1.2	7.2×10^{-4}
	Distractor variance	1.1	0.3	0.31
	Mean-variance interaction	3.7	1.0	3.1×10^{-3}
	Min T-D difference	−5.1	−1.4	2.4×10^{-4}
	Environment	−3.4	−0.94	5.5×10^{-3}
	Number of items	4.2	1.2	1.2×10^{-3}

Table 2. The effect of distractor and experiment variables on accuracy, hit rate, and FA rate. Surprisingly, T-D mean, distractor variance, and their interaction do not have a significant effect on accuracy. Instead, the effect of min T-D difference is significant. Bonferroni-corrected *p*-value criterion 8.3×10^{-3}.

explore whether the effects identified could be due to variability shared with additional variables. For each participant, we used T-D mean, distractor variance, their interaction, the minimum target-distractor difference (min T-D difference), distractor environment, and number of items as predictors in logistic regressions to predict accuracy, hit rate, or FA rate. As before, we compared the resulting coefficients to zero across participants. Only the min T-D difference and the number of items significantly predicted accuracy, and only the min T-D difference predicted hit rate (Table 2). A possible explanation for the difference between these findings, and the regressions without additional variables included, is that it is the min T-D difference that is the causally relevant variable. The effects of T-D mean and distractor variance may appear because T-D mean and distractor variance are correlated with the min T-D difference. This suggests that while Duncan and Humphreys (1989) may have identified distractor statistics that are related to performance, they make not be the cause of changes in performance.

A different pattern emerged when looking at the regression onto FA rate (Table 2). T-D mean and the interaction between T-D mean and distractor variance predicted FA rate in the same direction as they had without the inclusion of additional variables, although distractor variance was no longer a significant predictor. The min T-D difference had a large effect on FA rate, with FA rate decreasing as the min T-D difference increased. This finding provides evidence that T-D mean and distractor variance have some relevance to behavior, over and above their relationship with min T-D difference. Note that this finding does not rule out the possibility that the most similar item from the perspective of the observer is the only important variable in determining their response. This is because the min T-D difference, according to the participant, will not necessarily be the true min T-D difference, due to perceptual noise. Therefore, other items, not just the most similar, may affect behavior even if the observer only uses the item that appears most similar to them.

For the dedicated reader, we provide univariate analyses of the effects of distractor statistics in Appendix B.

To summarize, consistent with the account of Duncan and Humphreys (1989), there was evidence for an effect of T-D mean and distractor variance on accuracy. We detected an interaction effect between T-D mean and distractor variance. However, the effects of the interaction did not match the effects predicted by Duncan and Humphreys (1989). Plots suggested that the relationship between these variables and accuracy was not simple and may not be well described by the regression model used. The pattern of effects on FA rate was particularly clear and entirely consistent with SDT considerations. Analyses that also included the effects of additional variables suggested that, at least in the case of accuracy, min T-D difference is the causally relevant variable, not T-D mean or distractor variance.

Modeling methods

We next explored whether a computational model could provide a parsimonious explanation of the effects identified. We focus on a highly constrained Bayesian model initially and compare this model to other models below.

Generative model

We first specify how measurements are generated. The target is present on half of trials. Denote the presence of the target $C = 1$ and absence $C = 0$, then we have

$$p(C = 1) = p(C = 0) = \frac{1}{2}. \quad (1)$$

There are between two and six possible target locations, because there are between two and six items in a display. If the target is in location i, we write $T_i = 1$, and if it is absent in this location, $T_i = 0$. $\mathbf{T}$ indicates a vector containing T_i for every location. If the target is present at location i, then this item is at 45° clockwise from vertical. We express this using the Dirac delta function,

$$p(s_i|T_i = 1) = \delta(s_i).$$

Here, s_i represents the orientation of the Gabor item at the ith location in radians. Specifically, it represents twice the difference between the item orientation and the target orientation.

If the ith location contains a distractor, then the item orientation is drawn from a von Mises distribution with mean, $\mu = 0$, and concentration parameter κ_s,

$$p(s_i|T_i = 0) = \text{VM}(s_i; \mu, \kappa_s) = \frac{1}{2\pi I_0(\kappa_s)} e^{\kappa_s \cos(s_i - \mu)},$$

where VM indicates the von Mises distribution, and the second equality provides the definition of this function. I_0 are modified Bessel functions of the first kind and order zero. A von Mises distribution is similar to a normal distribution but is the appropriate distribution for a circular variable (i.e., orientation). A von Mises distribution with $\kappa_s = 0$ is the same as a uniform distribution. Therefore, we can model both distractor environments, uniform and concentrated, with this equation. Throughout, instead of directly reporting the von Mises concentration parameter, κ_s, we report,

$$\sigma_s = \frac{1}{2}\sqrt{-2 \log \frac{I_1(\kappa_s)}{I_0(\kappa_s)}}. \quad (2)$$

$2\sigma_s$ is the standard deviation of a wrapped normal distribution that closely approximates the relevant von Mises distribution (Stephens, 1963). For a uniform distribution ($\kappa_s = 0$), the corresponding wrapped normal distribution has $2\sigma_s = \infty$. The factor of $\frac{1}{2}$ is included as a very approximate way of compensating for the fact that s_i represents twice the difference between an item and the target.

Finally, we assume that the observer only receives noisy measurements of each item's orientation. We formalize this by assuming that measurements are drawn from a von Mises distribution centered on the true item orientation, but with concentration parameter κ,

$$p(x_i|s_i) = \text{VM}(x_i; s_i, \kappa). \quad (3)$$

Again, we do not report κ directly, but σ.

Optimal decision rule

Bayes's rule gives the optimal decision to make on the basis of noisy measurements. Here we only state our premises and conclusion, but the full derivation is provided in Appendix C.

We do not assume that the observer equally values hits and avoiding false alarms. Instead, we include in our models a parameter, p_{present}, which captures any bias toward reporting "target present." We use the fact that there is at most one target and assume that measurement noise at different locations is independent.

As shown in Appendix C, from these assumptions, we can derive the following rule for optimal behavior. The observer should report "target present" when

$$\log\left(\frac{1}{N}\sum_{i=1}^{N} e^{d_i}\right) + \log\frac{p_{\text{present}}}{1 - p_{\text{present}}} > 0, \quad (4)$$

where

$$d_i = \log\frac{p(x_i|T_i = 1)}{p(x_i|T_i = 0)} \quad (5)$$

$$= \kappa\cos(x_i) + \log\frac{I_0(\kappa_s)}{I_0\left(\sqrt{\kappa^2 + \kappa_s{}^2 + 2\kappa\kappa_s\cos(x_i)}\right)}, \quad (6)$$

and N is the number of items in the display.

Each d_i can be viewed as a "local" log-likelihood ratio (Ma et al., 2011; Palmer et al., 2000). For the case of uniform distractors, $\kappa_s = 0$, the expression simplifies to

$$d_i = \kappa\cos(x_i) - \log(I_0(\kappa)).$$

This is maximal when the measured orientation for item i matches the orientation of the target, $x_i = 0$, and decreases as the measured orientation moves further from the target orientation. Using a range of values, we found a similar pattern even when distractors are not uniformly distributed (Figure 5). Hence, the summation in (4) is a sum over apparently monotonically decreasing

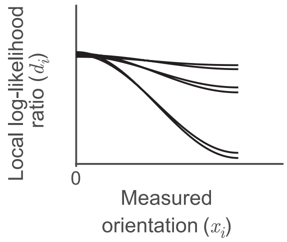

Figure 5. The relationship between the local log-likelihood ratio (d_i) and measured orientation (x_i). For a range of values ($\kappa = e^1, e^2, e^3$; $\kappa_s = 0, 1.5$), we observed that the local log-likelihood ratio was maximal when the measured orientation matched the target orientation (0) and decreased as measured orientation moved away from this value.

functions of the distance between each measured orientation and the target orientation.

Making predictions

We want to find the probability of a response, $\hat{C}$, given all stimuli orientations, $\mathbf{s}$. An item, s_i, generates measurements according to (3). Hence, for a particular set of items, we can simulate measurements and determine which responses these measurements lead to. By repeating this process many times, we can build an estimate of the probability, according to the model, that a particular set of items will lead to a "target present" or "target absent" response. For each trial, we simulated 1,000 sets of measurements and the associated decisions.

Lapses

We allow the possibility that some trials are the result of contaminant processes, such as getting distracted. On these "lapse" trials, the participant makes a random response. If we denote the probability of response, $\hat{C}$, according to the Bayesian observer model without lapses, $p_{\text{no lapse}}(\hat{C}|\mathbf{s})$, and the lapse rate λ, then the probability of a response is given by

$$p(\hat{C}|\mathbf{s}) = \frac{\lambda}{2} + (1-\lambda)p_{\text{no lapse}}(\hat{C}|\mathbf{s}). \tag{7}$$

Model fitting

Separately for each participant, we fitted lapse rate (λ), bias parameter (p_{present}), and the concentration parameter of measurement noise (κ) as free parameters. We allow the possibility that measurement noise varies with the number of items in the display and fit κ as four free parameters (one for each possible number of items in the display; Mazyar et al., 2013).

For any valid set of parameter values θ, we can calculate the likelihood. The likelihood is equal to the probability of the observed behavior, given the parameters and the stimulus shown. Assuming that responses in different trials are independent of each other, we can write the likelihood as a product of the probability of responses on each trial,

$$L(\theta) = p(\hat{C}^{(1)}, \hat{C}^{(2)}, \ldots|\theta, \mathbf{s}^{(1)}, \mathbf{s}^{(2)}, \ldots) \tag{8}$$

$$= \prod_i p(\hat{C}^{(i)}|\theta, \mathbf{s}^{(i)}), \tag{9}$$

where the product is taken over all the trials for a participant, $\hat{C}^{(i)}$ is the participant's response on the ith trial, and $\mathbf{s}^{(i)}$ denotes the stimulus on the ith trial. We used Bayesian adaptive direct search (BADS) to search for the parameters that maximized the log-likelihood (Acerbi and Ma (2017)). BADS is a well-tested optimization algorithm that alternates between a poll stage, in which nearby parameter values are evaluated, and a search stage, in which a Gaussian process model is fitted and used to determine promising parameter values to evaluate.

The model was fit separately for each participant. For each participant, we ran BADS 40 times. For each run, 150 parameter value sets were randomly selected and the likelihood evaluated at each. The set with the highest likelihood was used as the start point for the run. The bounds on the parameters during the search, the way in which initial parameter values were drawn, and the precise form in which the parameters were fit are described in Appendix F. Running the fitting procedure many times reduces the chance of getting stuck in local maxima and permits heuristic assessment of any problems local maxima may be causing (see supplementary methods of Acerbi, Dokka, Angelaki, & Ma, 2018). We found that fits to the same log-likelihood function often ended at different values of "maximum" log-likelihood, suggesting that we may have only found local maxima, rather than finding the global maximum. For a discussion of these issues, and our attempts to resolve them by reducing noise in the likelihood function, see Appendix G.

Alternative models

Given the strong effect of the minimum target-distractor difference on behavior, a heuristic that focuses on the measured orientation most similar to the target orientation might perform well. We compared the Bayesian observer to an observer who uses a very simple decision rule. If the absolute difference between the measured distractor orientation closest to the target orientation and the target orientation is below some threshold ρ, they report "target present"; otherwise, they report "target absent." Models of this kind have

Model number	1	2	3	4
Inference	Bayes	Bayes	Heuristic	Heuristic
Distractor environment use	True	False	True	False
Parameters				
Sensory noise (κ or σ)	4	4	4	4
Decision thresholds (ρ)	n/a	n/a	8	4
Bias (p_{present})	1	1	n/a	n/a
Distractor variability (κ_o or σ_o)	0	1	n/a	n/a
Lapse rate (λ)	1	1	1	1
Total	**6**	**7**	**13**	**9**

Table 3. Parameters of all models considered in the article. The main model is a Bayesian optimal observer (Model 1). We also considered an observer who applied a heuristic and made their decision entirely on the basis of the measured orientation closest to the target orientation (Models 3 and 4). "n/a" indicates that a model does not use a parameter.

been used extensively in visual search research (Ma, Shen, Dziugaite, & van den Berg, 2015). Note that, because the observer applies a criterion to a noisy variable to determine their response, this heuristic observer model is also a SDT model (Palmer et al., 2000, 1993). We make predictions for behavior in the same way that we did for the optimal observer model and fit the model in the same way.

We fitted two variants of this model. In Model 3 (see Table 3), the threshold used by the observer varies with different numbers of items in the display and varies in different distractor environments. There are four possible numbers of items in the display (2, 3, 4 and 6) and two distractor environments, giving a total of eight thresholds that were fitted as free parameters. In Model 4, we allowed the threshold to vary with number of items in the display but assumed that it was fixed across distractor environments, as if participants ignored the difference between the environments when making their decisions.

We also included a variant on the Bayesian observer model in our model comparison. The Bayesian observer discussed in the previous section is Model 1, but we also consider a model in which the observer is Bayesian, except they ignore the difference between the two distractor environments. Instead, this observer assumes all items, regardless of distractor environment, are distributed following a von Mises distribution with a concentration parameter that we fit, κ_o. This is Model 2. (As with all concentration parameters, we do not report κ_o directly, but σ_o. See Equation 2.) A list of all parameters and models is shown in Table 3.

We used the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) to compare the performance of these models. The AIC and BIC take into account the likelihood of the fitted models and the flexibility of each model in terms of the number of fitted parameters. A lower AIC and BIC indicates better fit. For each information criterion, we found the best-fitting model across participants using the mean value of the information criterion. To determine whether the difference in fit between the best-fitting model and the other models was meaningful, we calculated, for each participant, the difference in information criterion between the overall best-fitting model and each of the other models. By bootstrapping these differences 10,000 times, we computed 95% confidence intervals around the mean difference between the best-fitting model and each of the other models. If the confidence interval on the mean difference does not include zero for all competitor models, then we can conclude that the best-fitting model fits better than all other models.

Modeling results

We explore whether a Bayesian observer model can explain the effects of distractor statistics by fitting such a model, before simulating data using the parameter values fitted for each participant. By plotting both the real data and model-simulated data on the same plot, we can visually inspect whether the model successfully accounts for the trends in human behavior. For plots, we simulated 24,000 trials per participant. In plots, we use error bars for data and shading for model fits. Shading, like error bars, covers ±1 standard error of the mean. Beyond exploring the fit of the model to the data, of particular interest is whether the model re-creates patterns suggested by Duncan and Humphreys (1989), namely, a beneficial effect of T-D mean and a harmful effect of distractor variance, which is maximal when T-D mean is low.

We first looked at whether the model can successfully account for the individual effects of the distractor statistics. Looking at Figure 6, we can see that the model fits closely match the observed data and that all qualitative patterns are recovered. Consistent with the data and with Duncan and Humphreys (1989), the model predicts increased performance with increasing target-to-distractor mean difference (T-D mean). In contrast to the account of Duncan and Humphreys (1989), but consistent with the data, the model does not predict a strong relationship between distractor variance and performance.

We noted in the experimental results that minimum target-distractor difference (min T-D difference) may be the causally relevant variable. The strength of the effect of min T-D difference is accurately captured by the Bayesian model (Figure 6C). It is interesting to ask why the Bayesian model would predict such a strong effect of just one distractor, when the Bayesian

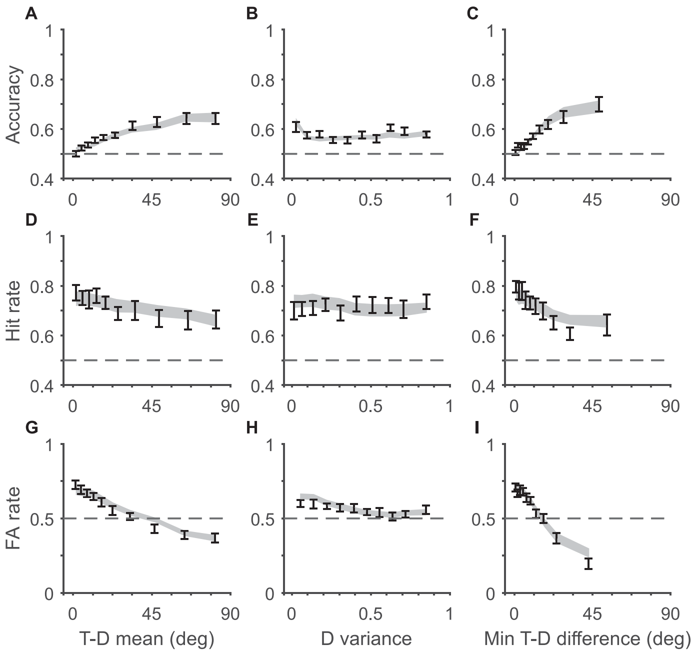

Figure 6. The effect of all summary statistics when considered individually (error bars) and model fits for these effects (shading). The Bayesian observer model captures the observed effects well.

observer combines all distractor measurements in their decision rule (see Equations 4 and 6). It turns out that, under specific conditions, the Bayesian observer closely approximates an observer who makes their decisions only on the basis of the item that appears closest to the target. Figure 7 shows the decision threshold the Bayesian observer applies to their measurements. The x- and y-axes represent measurements of two items. If the measurements fall within the marked area, the observer reports "target present." The decision thresholds are shown for a range of distractor environments, including the uniform and concentrated environments used in our study ($\sigma_s = \infty$ and $\sigma_s = 29^\circ$). The shape of the decision thresholds are also shown for an observer who uses a heuristic based on the item that appears closest to the target (see Modeling methods). We can see that for distractor environments with more variable distractors (the ones used in our experiment), the decision thresholds are very similar in shape to those used by an observer who makes their decision on the basis of the item which appears closest to the target, regardless of the other measurement.

We next looked at whether the Bayesian observer model could capture the observed interaction between T-D mean and distractor variance. The model captures the interaction between T-D mean and distractor variance on FA rate, including the decrease in FA rate with distractor variance at small T-D mean (Figure 8C). As discussed, this interaction on FA rate is predicted by signal detection theory accounts, because T-D mean

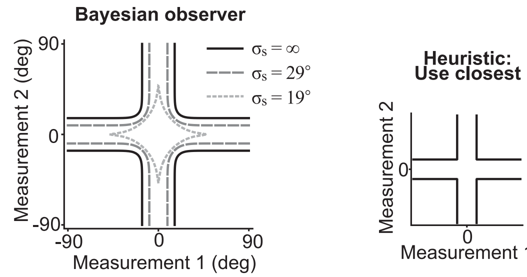

Figure 7. Decision thresholds used by the Bayesian observer for the case of two items. Also shown are the shape of the decision thresholds corresponding to a heuristic strategy in which the decision is based on the item closest to the target (see Modeling methods). The axes represent measurements of the items made by the observer, relative to the target orientation. If the measurements fall within the marked area, the observer reports "target present." The Bayesian observer thresholds were calculated using $\sigma = 10^\circ$ and under a range of values for σ_s, including those used in the experiment (uniform environment, $\sigma_s = \infty$; concentrated environment, $\sigma_s = 29^\circ$). For high σ_s, the Bayesian observer effectively only uses the measurement closest to the target to make their decision.

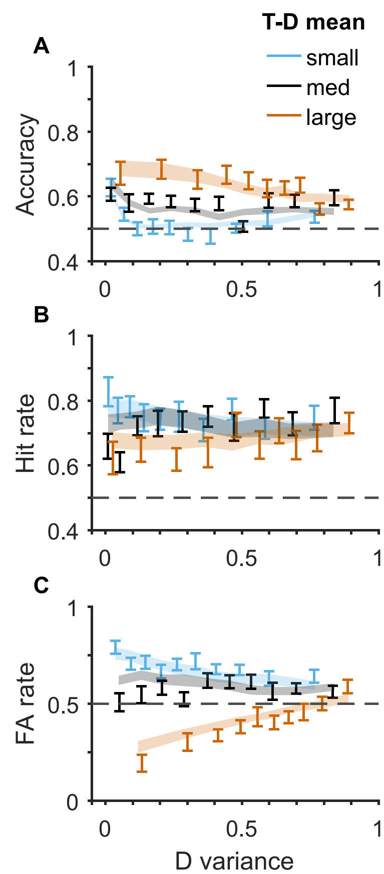

Figure 8. Interaction between T-D mean and distractor variance and model fits for these effects. The model captures the interaction of T-D mean and distractor variance on FA rate, along with the trends in accuracy.

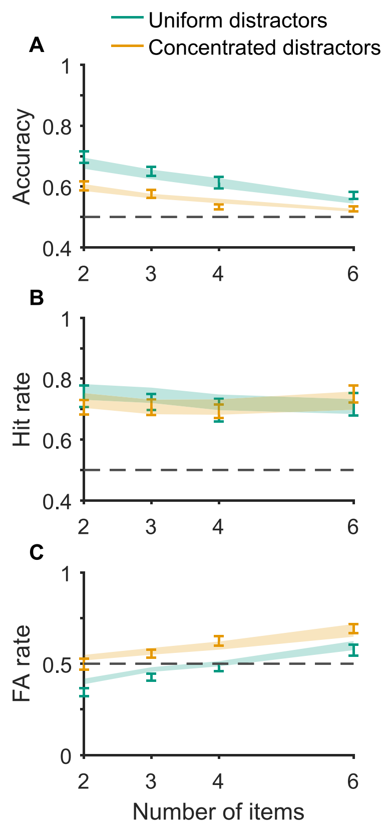

Figure 9. The effect of number of items. The Bayesian observer model captured the reduction in accuracy with more items and the increase in false alarms.

and distractor variance have an interactive effect on the probability of a confusing distractor (Rosenholtz, 2001). The Bayesian observer, as a specific kind of SDT observer, inherits this effect.

The model also captures effects on accuracy and hit rate. In particular, it accounts for the weak relationship between distractor variance, T-D mean, and hit rate (Figure 8B). Looking at the model fits for accuracy, we can see that the model largely captures the quantitative patterns (Figure 8A). Interestingly, the model captures the "U"-shaped relationship between accuracy and variance for small T-D mean trials. This relationship would be overlooked using regressions or logistic regressions alone but emerges out of an optimal observer model.

Much previous research has focused on the effect of number of items in the display (e.g., Treisman & Gelade, 1980; Duncan & Humphreys, 1989). We examined whether the Bayesian model could account for the effects of number of items on accuracy, hit rate, and FA rate (Figure 9). As in previous work (Mazyar et al., 2012, 2013), Bayesian observer model fits were highly accurate. The model captured the reduction in accuracy with number of items, the largely flat effect on hit rate, and the increase in false alarms. It captured the effect of number of items in both distractor environments and, additionally, captured the difference between the environments.

Finally, we looked at whether the model could account for fine-grained details by looking at the effect of distractor statistics separately for different numbers of items (uniform environment: Figure 10; concentrated environment: Figure 11). Mazyar et al. (2013) established that Bayesian observer models can account for the effects of min T-D difference in displays with different numbers of items. We looked at these effects here, but also looked at the effect of T-D mean and distractor variance. The model largely captures the effect of distractor statistics for all item numbers. We note that there appear to be some systematic deviations in the model fits. For example, for two items and concentrated distractors, the model does not capture an apparent dip in hit rate at median values of min T-D difference (Figure 11C). If reliable, this is an intriguing phenomenon: When the most similar distractor is very different to the target, "target present" responses are more probable than when the most similar distractor is just somewhat different. A similar pattern has been observed before (Mazyar et al., 2012). This suggests that some part of the mechanism of visual search may not be captured by the Bayesian model.

Having seen that a Bayesian observer model captures trends in the data well, we wanted to explore whether other models could also explain the data as well or better. Of particular interest is the question of whether a model in which the observer uses a heuristic might explain the observed data better. We compared two heuristic observer models (3 and 4) to two Bayesian observer models (1 and 2). The heuristic observer applies a threshold on the distractor that appears most similar to the target (from their perspective) to determine their response.

The results of the model comparison are presented in Figure 12. According to the AIC, a heuristic observer (Model 3) fit best. While Model 4 is similar to Model 3, in Model 3, the observer applies different decision thresholds depending on the distractor environment. Confidence intervals on the difference in fit between this and the other models did not include zero, suggesting Model 3 fit reliably better according to the AIC. Figure 12 also shows that, according to the AIC, a majority of participants were best fit by Model 3. In contrast, according to the BIC, the Bayesian optimal observer (Model 1) fit best. However, confidence intervals on the BIC differences suggested that the difference in fit between this model and the other models was not reliable. According to the BIC, a majority of participants were best fit by Model 1. The fits of model 3 to data are provided in Appendix D.

We do not want the conclusions of our research to depend on the fit metric used. Therefore, in this case, we cannot draw conclusions about which model fit

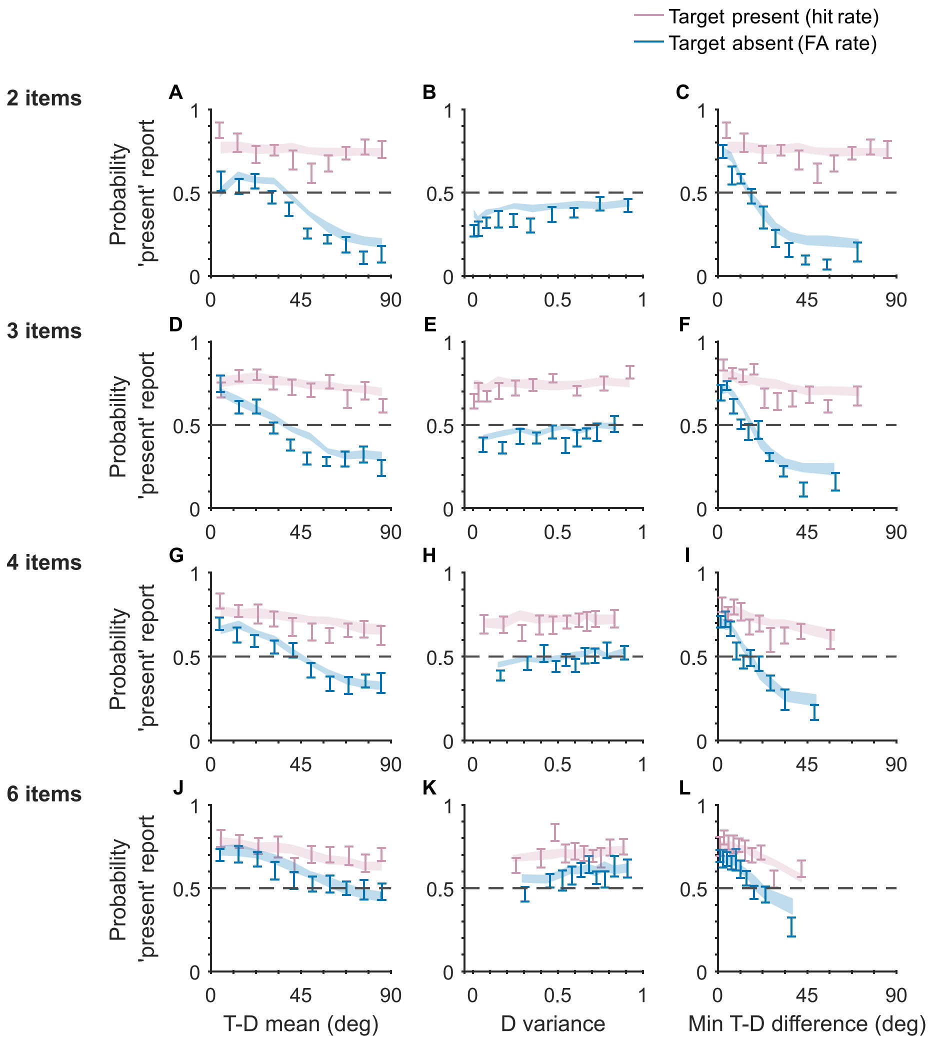

Figure 10. Effect of distractor statistics in the uniform environment, at different numbers of items. The Bayesian observer model successfully accounts for most effects at all numbers of items considered, although there appear to be some systematic deviations.

best. The differences between the AIC and BIC results stem from the fact that the BIC penalizes extra model parameters more harshly. Model 3, best according to the AIC, has the most parameters out of all the models and so would be penalized heavily by the BIC (see Table 3). Using the AIC and BIC alone, we cannot say whether this penalization is fair or not.

To explore these results further, we performed model recovery analysis. For each model, we simulated new data using this model and its fitted parameters. The data set simulated using a model was the same size as the real data set, with one simulated participant for each real participant. This gave us four simulated data sets. We then fit all four models to all four simulated data sets. The only case in which the model used to simulate the data was not the best-fitting model, according to the information criteria, was when data were simulated using Model 3. On the AIC, Model 3 fit best, as expected. However, according to the BIC, Model 1 fit best. This suggests that in the present case, BIC may be unreasonably harsh on complex models. We note also that the median fitted lapse rate for Model 1 was 0.31, which seems unreasonably high (Appendix E). In contrast, the median fitted lapse rate for Model 3 was 0.14. Hence, there is very tentative evidence pointing to Model 3 as the best model.

Bayesian observer models, and heuristic observer models of the kind considered here, have proved difficult to distinguish in previous work where the target takes a single value, and distractors are of equal reliability (Ma et al., 2015, Section 2.3.2). We had hoped that the task used here, with two different distractor environments, could tease apart the models, but that proved not to be the case. The discussion above may help us understand why the Bayesian and heuristic observer models are difficult to distinguish: Under certain parameter values, the Bayesian observer effectively only uses the measured orientation closest to the target orientation to make their decision, just as the heuristic observer does (Figure 7). In Figure 7, the shape of the decision

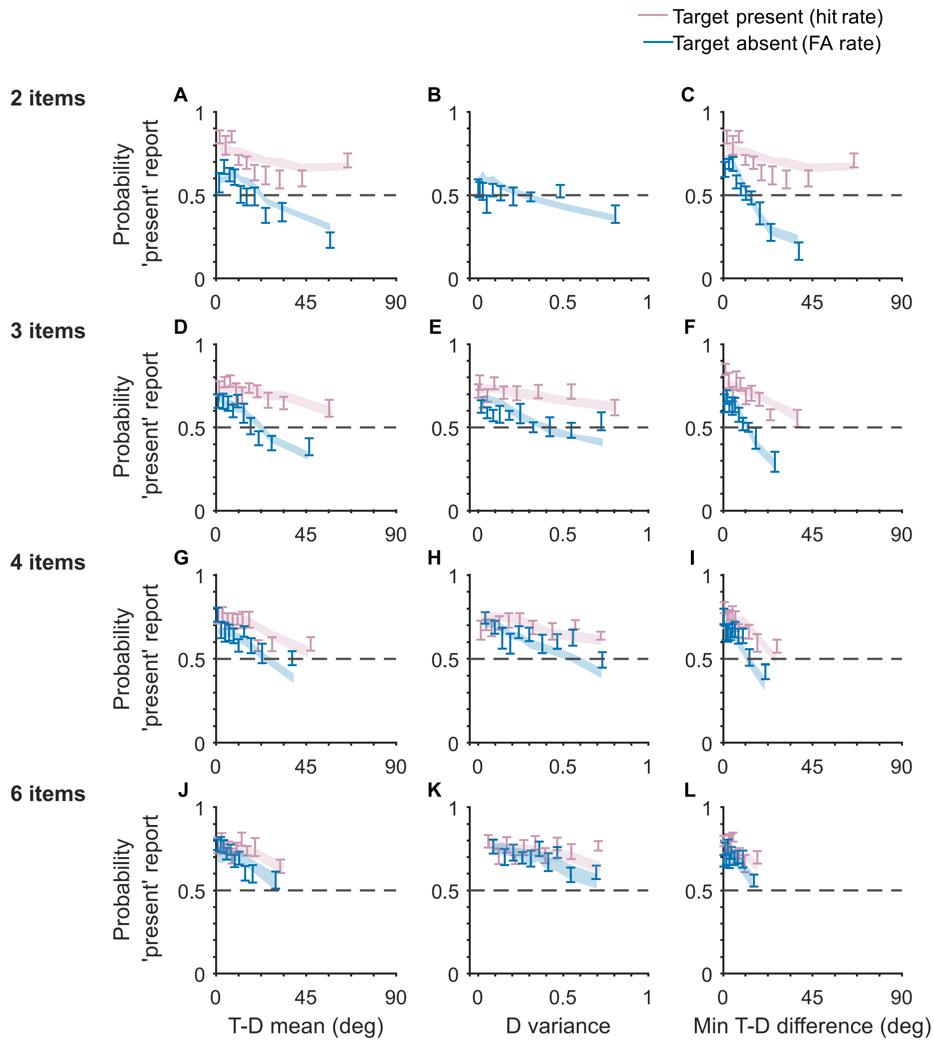

Figure 11. Effect of distractor statistics in the concentrated environment, at different numbers of items. The Bayesian observer model successfully accounts for most effects observed.

thresholds used by the Bayesian and heuristic observers is qualitatively different for a distractor environment that we did not use, where $\sigma_s = 19^\circ$. Provided this does not make the task too difficult for observers, using such an environment may make it possible to distinguish between the two models.

We could also not decisively say whether observers used or ignored the difference between the uniform and concentrated distractor environments. Model 2 and Model 4, the models in which observers ignored the difference (Table 3), fit reliably worse according to the AIC (Figure 12). However, the difference in fit was not reliably worse according to the BIC. Both models that used and those that ignored the difference between distractor environments could predict different effects of distractor statistics in the two environments. For example, Figure 13 shows data and model fits for Model 2, a Bayesian model in which the observer ignores the difference between the two distractor environments. Despite this fact, the model predicts differences in the effect of distractor variance between the two environments. Such effects must be due to correlations between distractor environment and other distractor statistics that do have an effect on behavior.

Parameter estimates from the fits are provided in Appendix E.

General discussion

In this study, we asked participants to perform a visual search task with heterogeneous distractors. We looked for the distractor statistic effects identified by Duncan and Humphreys (1989)—a beneficial effect of T-D mean and a harmful effect of distractor variance that is maximal when T-D mean is low—but the results were mixed. We found some evidence for an effect of target-to-distractor mean difference (T-D mean), and distractor variance on accuracy. There was also

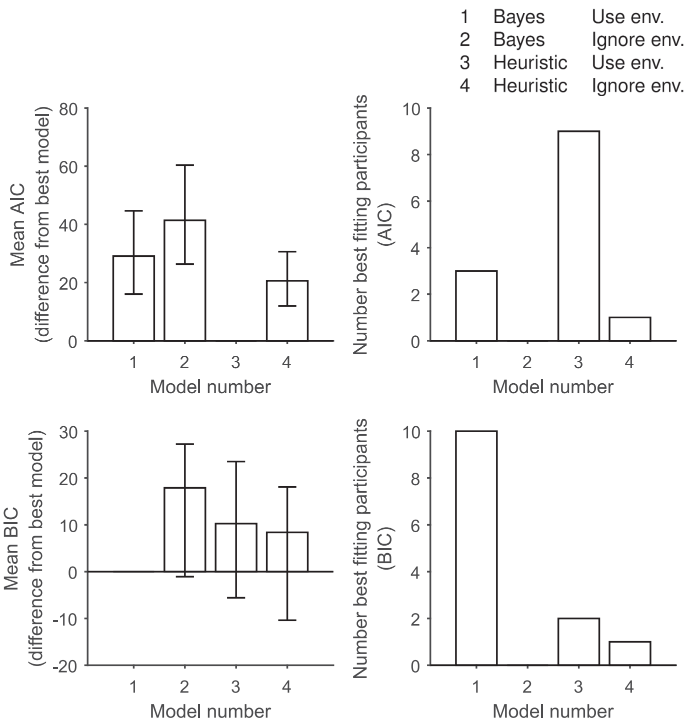

Figure 12. Mean AIC and BIC relative to the best-fitting model and the number of participants best fit by each model. See Table 3 for details of the models. Model comparison results were inconclusive because a consistent pattern of results was not found across AIC and BIC. Unlike in other plots, error bars here reflect 95% bootstrapped confidence intervals.

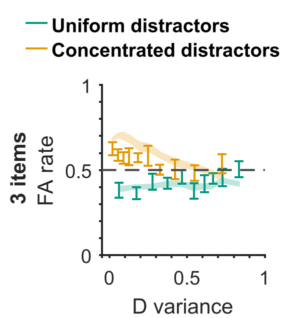

Figure 13. Data and Model 2 fits for the effect of distractor variance on FA rate. Model 2 assumes observers ignore the difference between the two distractor environments. Nevertheless, the model can predict differences between the two environments. This is likely because distractor environment correlates with other distractor statistics.

evidence for an interaction between T-D mean and distractor variance on accuracy, but this interaction led to effects that were different from those predicted by Duncan and Humphreys (1989). We found that a statistic not explicitly considered by Duncan and Humphreys (1989), namely, minimum target-distractor difference (min T-D difference), had a strong effect on behavior, and that the effects of T-D mean and distractor variance may in fact be consequences of the effect of min T-D difference.

One potential reason for the discrepancy between our results and the account of Duncan and Humphreys (1989) is that we explored the difficulty of visual search through accuracy, while their primary variable of interest was response time (to be precise, the increase in response time as the number of items increased). Specifically in the context of visual search, there is evidence that stimuli that generate low accuracy also generate slow responses (Eckstein, Thomas, Palmer, & Shimozaki, 2000; Palmer, 1998; Geisler & Chou, 1995). Hence, difference in primary variable seems an unlikely explanation of the discrepancy between our results and the account of Duncan and Humphreys (1989). Nevertheless, it would be worthwhile to explore our data with a process model that makes predictions for response time as well as accuracy.

A second potentially important difference between our work and the work of Duncan and Humphreys (1989) involves the calculation of distractor statistics. In the present work, we explored the effects of statistics of sampled distractors. Duncan and Humphreys (1989, p. 444) held that both sample and population distractor statistics (statistics of the population from which distractors are drawn) have a role. Studying the effect of population distractor statistics would involve training participants on a wide range of probability distributions. We only used two distractor distributions (uniform and concentrated distractors). As discussed, we could not decisively say whether participants learned and used the difference between environments, suggesting training participants on a wide range of distributions would be challenging. Future studies could explore the effects of population statistics using large numbers of participants and a between-participants design. Alternatively, in the right setting, rapid learning of distractor distributions may be feasible (Chetverikov, Campana, & Kristjánsson, 2017). It remains possible, then, that the patterns identified by Duncan and Humphreys (1989) accurately describe the effects of population distractor statistics.

The conclusion that T-D mean or distractor variance matter for performance, but only because of their effects on min T-D difference, provides a reinterpretation that is broadly compatible with the findings of Duncan and Humphreys (1989). However, this conclusion appears to be in conflict with more recent findings. Rosenholtz (2001) found that performance in visual search suffers when distractors are made more variable, even when this is done in a way that does not move any distractors closer to the target. This effect was found in a stimulus featuring 36 items, which were within a diameter of 3.75°. Crowding may play a key role in this stimulus (Levi, 2011). Indeed Rosenholtz (2001) did not find a decrease in performance with increased variance when only eight items were used. Crowding is not likely to play a role in our experiment, where stimuli are separated by a distance greater than half their eccentricity (Rosen, Chakravarthi, & Pelli, 2014; Pelli & Tillman, 2008).

In the second half of this article, modeling revealed that a Bayesian model with only six free parameters could account for a rich pattern of effects of distractor statistics. It captured the way T-D mean, distractor variance, and min T-D difference affected accuracy, false alarm (FA), and hit rate. It also accounted for the interaction between T-D mean and distractor variance, various effects of set size, and effects of distractor statistics at different set sizes. A model comparison of the Bayesian model with a variant and with a heuristic observer model was inconclusive. This may be due in part to the similarity of the decision rule for the Bayesian and the heuristic observer: The Bayesian observer may effectively only use the item that looks most similar to the target (Figure 7), the policy of the heuristic observer. Potential solutions include making distractors even more concentrated (Figure 7) or using items of differing reliability within the same stimulus (Stengård & van den Berg, 2019).

While we were unable to determine which model fit the data best, our findings suggest that SDT models (of which the Bayesian observer model and heuristic observer model are variants) can provide parsimonious explanations for a large set of distractor statistic phenomena. This work highlights some of the advantages of computational modeling. In particular, by building a process model of how stimuli are mapped to response, we were able to make predictions for a very wide range of effects. In fact, we could make predictions for how any distractor statistic affects any statistic summarizing behavior.

These findings complement other work showing that SDT models can provide parsimonious explanations of apparently complex phenomena in visual search. For example, SDT models can account for the apparent distinction between feature and conjunction search discussed in the introduction. This apparent distinction emerges from an SDT model that makes sensible assumptions about how multidimensional items are encoded, but the model does not need to treat feature and conjunction searches as qualitatively different processes (Eckstein et al., 2000). The results also complement work showing that SDT models provide a good explanation of behavior across a wide range of visual search tasks and conditions. Mihali and Ma (2020) found that a Bayesian SDT model could account for data from different forms of visual search. Specifically, Mihali and Ma (2020) fit models to data from a visual search task in which observers had to detect the presence versus absence of a target and a visual search task in which observers had to indicate the location of the target. In both these cases, whether observers had to hold the search array in memory was also manipulated. A Bayesian SDT model provided a reasonable or good fit in all conditions studied. The success of SDT models highlights the value of using stimuli that can be quantified and continuously varied, and of building quantitative process models that can make predictions for behavior on a trial-by-trial basis.

On the other hand, the performance of SDT models observed in this study and in the examples discussed conflicts with the finding of Rosenholtz (2001), that an SDT model provided a poor fit to performance as properties of distractors were varied. Poor fits were obtained with large and small numbers of items in stimuli. A key difference between our studies may be that Rosenholtz (2001) used distractors that took one of a small number of values. We used stimuli in which distractors could take any value. One explanation for these findings is that, as hypothesized by Duncan and Humphreys (1989), distractors in close proximity that share the same value may be grouped together and treated as a single item.

Our work has a number of limitations in scope, stemming from the fact that the experimental setup was chosen to ensure the experiment was well controlled and that modeling of behavior was tractable. As already noted, we have not studied the effects of crowding on visual search or the effects of perceptual grouping. Another important aspect of the experimental design is that stimuli were only presented for 100 ms, precluding the possibility of saccades. Preventing saccades is a common experimental choice (e.g. Eckstein et al., 2000; Palmer et al., 2000; Mazyar et al., 2013); the rationale is that it limits the complexity of the system under investigation. Specifically, we can ignore processes such as saccade selection and integration of previously gathered information and instead focus on encoding and decision (see Table 4). Nevertheless, much visual search research has focused on tasks in which viewing time is unlimited. Importantly for the present discussion, many of the experiments in Duncan and Humphreys (1989) featured unlimited viewing time. Discrepancies between our results and the work of Duncan and Humphreys (1989) might, therefore, stem from the effects of processes studied by Duncan and Humphreys (1989), but not here. For instance, distractor variability might have a negative effect on the quality of saccades selected.

Another decision that may limit the scope of the results is the use of relatively simple search displays. The items in the stimuli were easily distinguished from the surround and from each other, varied only along a single dimension, and were not correlated with each other (Palmer et al., 2000; Bhardwaj, van den Berg, Ma, & Josić, 2016). It is possible that the effects identified here would not generalize to tasks with more complex displays. However, in an important sense, the

Component processes in visual search		Example study
a	Isolating objects from background	Wolfe, Alvarez, Rosenholtz, Kuzmova, and Sherman (2011)
b	Encoding of sensory information	Shen and Ma (2019)
c	Decision mechanism	This article
d	Saccade selection	Najemnik and Geisler (2005)
e	Integration of information from previous saccades	Horowitz and Wolfe (1998)

Table 4. Naturalistic visual search involves a large number of processes. In the present study, through the design of the experiment, we focused on processes (b) and (c).

complexity of the stimuli used here is greater than that of some typically used displays: In many studies, distractors have only taken one of a small set of values (e.g., Treisman & Gelade, 1980; Rosenholtz, 2001). In our study, the items could take an infinite number of values, and all items in the display took different values to each other. It would be premature, then, to dismiss the conclusions of the present study on the grounds that the task used is simpler than tasks in previous research.

Even more important than whether the results can be compared to previous research is the question of whether the results generalize to naturalistic visual search. Palmer et al. (2000) highlighted many ways in which naturalistic visual search differs from conditions in lab studies. In real-world visual search, targets are unlikely to take specific values (e.g., you want to detect any car or motorcycle approaching, not just one specific car), vary along a single dimension (e.g., cars vary in lots of ways), appear at a fixed location (e.g., a car could be anywhere along a road), involve small numbers of items, or be presented against a plain background (e.g., cars will be in a scene with signs, pedestrians, houses, and trees). By using simple, briefly presented stimuli, we have clearly not studied all the processes involved in naturalistic visual search (Table 4).

As scientists, our shared aim is to build a complete understanding of visual search, not just as it operates in the lab, but in naturalistic settings. Nevertheless, there are definite advantages to studying component processes separately. The choice of stimuli in the present study allowed us to explore the encoding and decision mechanisms of visual search in isolation. Thus, this choice vastly simplified the research problem and increased the chances of producing intelligible results. We have seen in this article just how successful our models of single stages (here the decision stage) can be. Moreover, there is much research on the other processes that make up visual search (see Table 4). The present article contributed to this shared effort to understand naturalistic visual search by demonstrating that models of the decision mechanism provide an excellent account of the effects of distractor statistics.

Keywords: Bayesian inference, visual search, computational modeling

Acknowledgments

The authors thank the reviewers for helpful comments and points, including the point that crowding may be a key factor in understanding the difference between our and previous results, and the point that in certain distractor environments, it may be possible to distinguish the Bayesian and heuristic models. They also thank Andra Mihali and Heiko Schütt for helpful discussions and advice over the course of the project. The authors are very grateful to Luigi Acerbi for extensive advice on dealing with model-fitting difficulties. Wei Ji Ma was supported by Grant R01EY020958 from the National Institutes of Health. The authors gratefully acknowledge the use of the University of Oxford Advanced Research Computing (ARC) facility in carrying out this work. http://dx.doi.org/10.5281/zenodo.22558.

Commercial relationships: none.
Corresponding author: Joshua Calder-Travis.
Email: j.calder.travis@gmail.com.
Address: Institute of Neurophysiology and Pathophysiology, UKE, Hamburg, Germany.

References

Acerbi, L., Dokka, K., Angelaki, D. E., & Ma, W. J. (2018). Bayesian comparison of explicit and implicit causal inference strategies in multisensory heading perception. *PLoS Computational Biology, 14*, e1006110, doi:10.1371/journal.pcbi.1006110.

Acerbi, L., & Ma, W. J. (2017). Practical Bayesian optimization for model fitting with Bayesian adaptive direct search. In *Advances in Neural Information Processing Systems 30* (pp. 1836–1846). Retrieved from http://papers.nips.cc/paper/6780-practical-bayesian-optimization-for-model-fitting-with-bayesian-adaptive-direct-search.pdf.

Afshartous, D., & Preston, R. A. (2011). Key results of interaction models with centering. *Journal of Statistics Education, 19*(3), doi:10.1080/10691898.2011.11889620.

Berens, P. (2009). CircStat: A MATLAB toolbox for circular statistics. *Journal of Statistical Software, 31*(10), 1–21, doi:10.18637/jss.v031.i10.

Bhardwaj, M., van den Berg, R., Ma, W. J., & Josić, K. (2016). Do people take stimulus correlations into account in visual search? *PLoS ONE, 11*(3), 1–16, doi:10.1371/journal.pone.0149402.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433–436, doi:10.1/156856897X00357.

Cameron, E. L., Tai, J., Eckstein, M., & Carrasco, M. (2004). Signal detection theory applied to three visual search tasks identification, yes/no detection and localization. *Spatial Vision, 17*, 295–325, doi:10.1163/1568568041920212.

Chetverikov, A., Campana, G., & Kristjánsson, Á. (2017). Rapid learning of visual ensembles. *Journal of Vision, 17*(2), 21, doi:10.1167/17.2.21.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum Associates.

Drugowitsch, J., Wyart, V., Devauchelle, A.-D., & Koechlin, E. (2016). Computational precision of mental inference as critical source of human choice suboptimality. *Neuron, 92*, 1–14, doi:10.1016/j.neuron.2016.11.005.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96*, 433–458.

Eckstein, M. P. (2011). Visual search: A retrospective. *Journal of Vision, 11*, 14, doi:10.1167/11.5.14.

Eckstein, M. P., Thomas, J. P., Palmer, J., & Shimozaki, S. S. (2000). A signal detection model predicts the effects of set size on visual search accuracy for feature, conjunction, triple conjunction, and disjunction displays. *Perception & Psychophysics, 62*, 425–451, doi:10.3758/BF03212096.

Estes, W. K., & Taylor, H. A. (1964). A detection method and probabilistic models for assessing information processing from brief visual displays. *Proceedings of the National Academy of Sciences of the United States of America, 52*, 446–454.

Geisler, W. S., & Chou, K.-L. (1995). Separation of low-level and high-level factors in complex tasks: Visual search. *Psychological Review, 102*, 356–378.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: Wiley.

Horowitz, T. S., & Wolfe, J. M. (1998). Visual search has no memory. *Nature, 394*, 575–577, doi:10.1038/29068.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception, 36*. ECVP Abstract Supplement.

Levi, D. M. (2011). Visual crowding. *Current Biology, 21*(18), R678–R679, doi:10.1016/j.cub.2011.07.025.

Ma, W. J., Navalpakkam, V., Beck, J. M., van den Berg, R., & Pouget, A. (2011). Behavior and neural basis of near-optimal visual search. *Nature Neuroscience, 1*, 783–790, https://doi.org/10.1038/nn.2814.

Ma, W. J., Shen, S., Dziugaite, G., & van den Berg, R. (2015). Requiem for the max rule? *Vision Research, 116*, 179–193, https://doi.org/10.1016/j.visres.2014.12.019.

Mazyar, H., van den Berg, R., & Ma, W. J. (2012). Does precision decrease with set size? *Journal of Vision, 12*(6), 10, doi:10.1167/12.6.10.

Mazyar, H., van den Berg, R., Seilheimer, R. L., & Ma, W. J. (2013). Independence is elusive: Set size effects on encoding precision in visual search. *Journal of Vision, 13*(5), 8, doi:10.1167/13.5.8.

Mihali, A., & Ma, W. J. (2020). The psychophysics of visual search with heterogeneous distractors. bioRxiv, doi:10.1101/2020.08.10.244707.

Murray, R. F., & Morgenstern, Y. (2010). Cue combination on the circle and the sphere. *Journal of Vision, 10*, 15, doi:10.1167/10.11.15.

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature, 434*, 387–391, doi:10.1038/nature03390.

Neisser, U. (1964). Visual search. *Scientific American, 210*, 94–103.

Palmer, J. (1994). Set-size effects in visual search: The effect of attention is independent of the stimulus for simple tasks. *Vision Research, 34*, 1703–1721, doi:10.1016/0042-6989(94)90128-7.

Palmer, J. (1998). Attentional effects in visual search: Relating search accuracy and search time. In R. D. Wright (Ed.), *Visual attention* (pp. 348–388). New York, NY: Oxford University Press.

Palmer, J., Ames, C. T., & Lindsey, D. T. (1993). Measuring the effect of attention on simple visual search. *Journal of Experimental Psychology, 19*, 108–130.

Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. *Vision Research, 40*, 1227–1268, doi:10.1016/S0042-6989(99)00244-8.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437–442, doi:10.1163/156856897X00366.

Pelli, D. G., & Tillman, K. A. (2008). The uncrowded window of object recognition. *Nature Neuroscience, 11*, 1129–1135, doi:10.1038/nn.2187.

Peterson, W., Birdsall, T., & Fox, W. (1954). The theory of signal detectability. *Transactions of the IRE Professional Group on Information Theory, 4*, 171–212, doi:10.1109/TIT.1954.1057460.

Rosen, S., Chakravarthi, R., & Pelli, D. G. (2014). The Bouma law of crowding, revised: Critical spacing is equal across parts, not objects. *Journal of Vision, 14*(6), 10, doi:10.1167/14.6.10.

Rosenholtz, R. (2001). Visual search for orientation among heterogeneous distractors: Experimental results and implications for signal-detection theory models of search. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 985–999, doi:10.1037/0096-1523.27.4.985.

Schwaninger, A. (2005). Increasing efficiency in airport security screening. *WIT Transactions on The Built Environment, 82*, 12, doi:10.2495/SAFE050401.

Shen, S., & Ma, W. J. (2019). Variable precision in visual perception. *Psychological Review, 126*, 89–132, doi:10.1037/rev0000128.

Stengård, E., & van den Berg, R. (2019). Imperfect Bayesian inference in visual perception. *PLoS Computational Biology, 15*, e1006465, doi:10.1371/journal.pcbi.1006465.

Stephens, M. A. (1963). Random walk on a circle. *Biometrika, 50*, 385–390, doi:10.2307/2333907.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97–136, doi:10.1016/0010-0285(80)90005-5.

Wolfe, J. M., Alvarez, G. A., Rosenholtz, R., Kuzmova, Y. I., & Sherman, A. M. (2011). Visual search for arbitrary objects in real scenes. *Attention, Perception & Psychophysics, 73*, 1650–1671, doi:10.3758/s13414-011-0153-3.

Appendix A: Participant demographics

Gender	Female	10
	Male	4
	Non binary	0
	Prefer not to say	0
Age	18–25	10
	26–35	3
	36–45	1
	46–55; 56–65; 66+	0
Handedness	Right	10
	Left	3
	Neither	1

Table 5. Aggregated gender, age, and handedness information for participants in the study.

Appendix B: Univariate analysis of distractor statistic effects

In addition to the analysis discussed in the main text, we also looked at the effect of distractor statistics when considered individually (i.e., ignoring variance shared with other distractor statistics and experimental variables). For each participant, we used target-to-distractor mean difference (T-D mean), distractor variance, or the minimum target-distractor difference (min T-D difference) in a logistic regression to predict accuracy, hit rate, or FA rate. We compared the regression coefficients to zero across participants.

As expected, if the mean of the distractors was further from the target orientation, participants were less likely to report “target present” (Figure 14D and G; Table 6). Increasing T-D mean also increased accuracy of responses (Figure 14A; Table 6). Surprisingly, distractor variance was only related to FA rate. Increasing distractor variance predicted fewer false alarms (Figure 14H; Table 6). Like T-D mean, the min T-D difference strongly predicted accuracy and hit and FA rate. As the min T-D difference increased, the probability of a “target present” report decreased

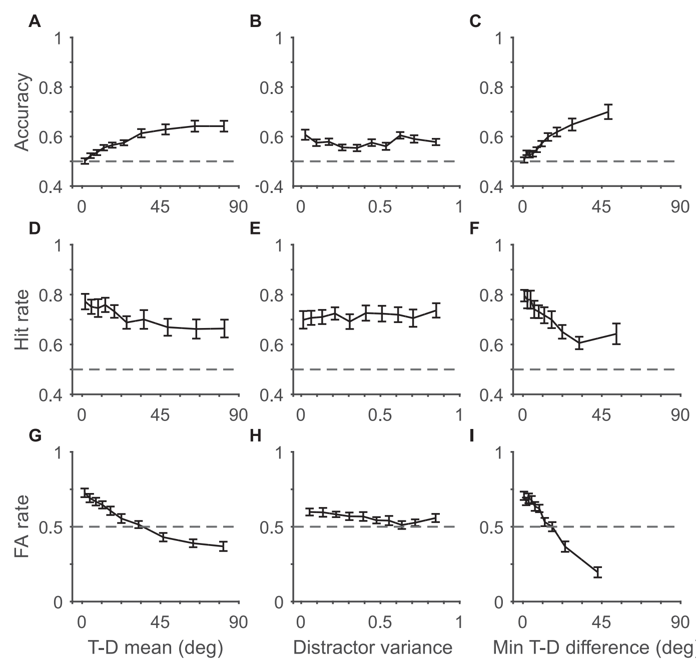

Figure 14. The individual effect of three distractor statistics on accuracy, hit rate, and FA rate. As the T-D mean and min T-D difference increased, performance also increased and “target present” responses decreased. Surprisingly, distractor variance only had an effect on FA rate.

Outcome	Predictor	*t*-value	Effect size (*d*)	*p*-value
Accuracy	T-D mean	6.1	1.7	5.5×10^{-5}
Accuracy	Distractor variance	0.22	0.061	0.83
Accuracy	Min T-D difference	5.0	1.4	3.3×10^{-4}
Hit rate	T-D mean	−5.1	−1.4	2.7×10^{-4}
Hit rate	Distractor variance	1.7	0.48	0.11
Hit rate	Min T-D difference	−8.8	−2.4	1.5×10^{-6}
FA rate	T-D mean	−8.9	−2.5	1.3×10^{-6}
FA rate	Distractor variance	−3.0	−0.84	0.011
FA rate	Min T-D difference	−7.0	−1.9	1.4×10^{-5}

Table 6. The effect of distractor statistics on accuracy, hit rate, and FA rate, when these effects are considered independently of the effects of other distractor statistics (including the other distractor statistics in this table), and experiment variables.

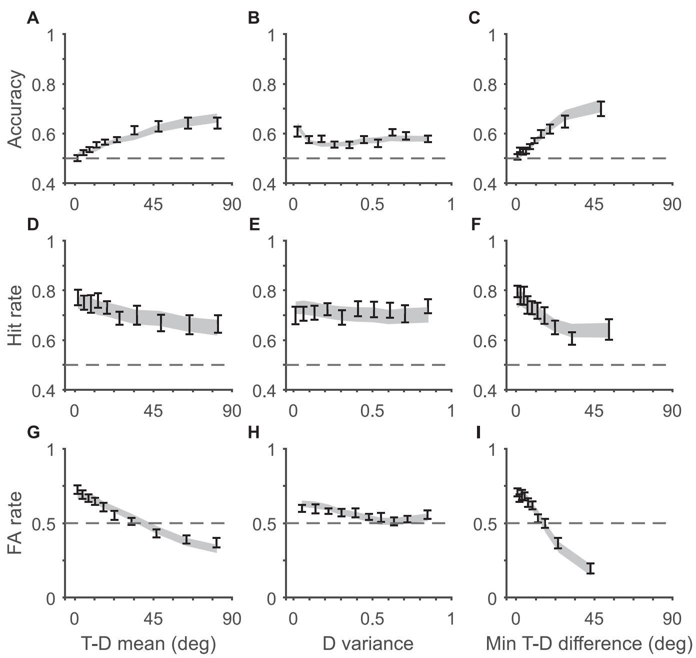

Figure 15. The effect of all summary statistics when considered individually (error bars) and Model 3 fits for these effects (shading).

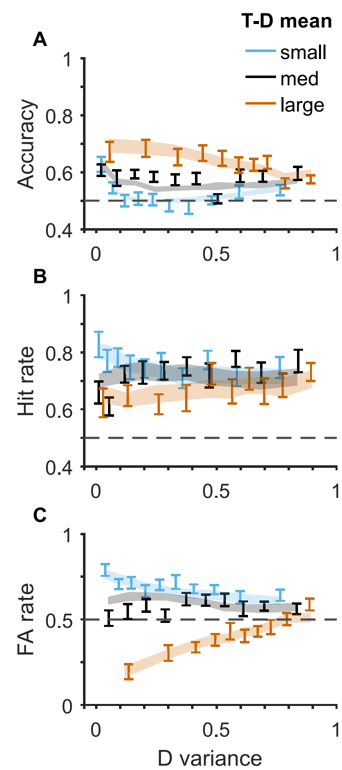

Figure 16. Interaction between T-D mean and distractor variance.

(Figure 14F, I; Table 6). At the same time, accuracy increased (Figure 14C; Table 6).

The effects of T-D mean and min T-D difference on "target present" responses, both when the target was present and when it was absent, suggest that participants were using a sensible strategy to perform the task: If the distractors were less like the target, participants were less likely to report "target present." In addition, we observed that with increasing T-D mean and min T-D difference, performance improved. This finding suggests that similarity of target and distractors is an important determinant of performance. The lack of an effect of variance on performance may be because, as discussed in the introduction, increasing variance can make easily confused distractors either more or less likely, depending on the value of the T-D mean.

Appendix C: Derivation of the optimal decision rule

A Bayes-optimal observer uses the true generative model and evidence received in the form of measurements, to infer the probability of "target present" and "target absent." An observer who equally values hits and avoiding false alarms responds that the target is present when this is more likely than the target being absent. This is the same as using a criterion of 1 on the ratio of these probabilities or 0 on the log of this

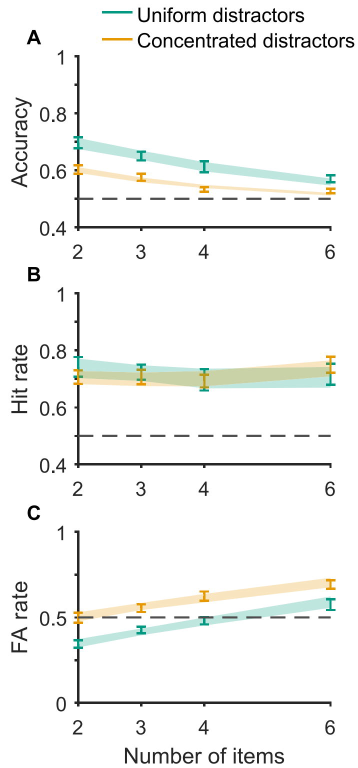

Figure 17. The effect of number of items.

ratio. We can write the condition for responding "target present" as

$$\log \frac{p(C=1|\mathbf{x})}{p(C=0|\mathbf{x})} > 0, \tag{10}$$

where $p(C=1|\mathbf{x})$ is the probability of the target being present after incorporating the information provided by the measurements (the posterior probability). $\mathbf{x}$ represents a vector of x_i for all i.

We will not assume that the observer equally values hits and avoiding false alarms. Instead, much like in SDT, we allow for the possibility that the observer values hits more than avoiding false alarms or vice versa. Hence, in our models, we will use the following decision rule,

$$d = \log \frac{p(C=1|\mathbf{x})}{p(C=0|\mathbf{x})} + \log \frac{p_{\text{present}}}{1-p_{\text{present}}} > 0, \tag{11}$$

where p_{present} is the parameter that captures any bias towards reporting "target present," and d denotes the sum of the posterior ratio and the bias term. Using Bayes's rule and taking logarithms, we have

$$\log \frac{p(C=1|\mathbf{x})}{p(C=0|\mathbf{x})} = \log \frac{p(\mathbf{x}|C=1)}{p(\mathbf{x}|C=0)} + \log \frac{p(C=1)}{p(C=0)}$$

$$= \log \frac{p(\mathbf{x}|C=1)}{p(\mathbf{x}|C=0)},$$

where the second line follows from Equation 1. Hence, the optimal observer will report "target present" when

$$d = \log \frac{p(\mathbf{x}|C=1)}{p(\mathbf{x}|C=0)} + \log \frac{p_{\text{present}}}{1-p_{\text{present}}} > 0. \tag{12}$$

Assuming that there is at most one target and that measurement noise at different locations is independent, it has been shown that the log-likelihood ratio is given by (Ma et al., 2011; Palmer et al., 2000)

$$\log \frac{p(\mathbf{x}|C=1)}{p(\mathbf{x}|C=0)} = \log\left(\frac{1}{N}\sum_{i=1}^{N} e^{d_i}\right), \tag{13}$$

where N indicates the total number of Gabor patches in the display, and d_i indicates the local log-likelihood ratio for location i. The local log-likelihood ratio is defined as

$$d_i = \log \frac{p(x_i|T_i=1)}{p(x_i|T_i=0)}. \tag{14}$$

Marginalizing over s_i and substituting in expressions from the generative model, we find

$$d_i = \log \frac{\int p(x_i|s_i)p(s_i|T_i=1)ds}{\int p(x_i|s_i)p(s_i|T_i=0)ds}$$

$$= \log \frac{\mathrm{VM}(x_i; 0, \kappa)}{\int \mathrm{VM}(x_i; s_i, \kappa)\mathrm{VM}(s_i; \mu, \kappa_s)ds}.$$

The denominator in this expression is the product of two von Mises distributions. Murray and Morgenstern (2010) state that the product of two von Mises is a new, scaled, von Mises. Any von Mises distribution, integrated over all angles, gives 1, because it is a probability distribution. Hence, when we integrate over all s_i, we will only be left with the scaling. Using the formula from Murray and Morgenstern (2010), we have

$$d_i = \log \frac{\mathrm{VM}(x_i; 0, \kappa)}{\frac{I_0\left(\sqrt{\kappa^2+\kappa_s{}^2+2\kappa\kappa_s\cos(x_i-\mu)}\right)}{2\pi I_0(\kappa)I_0(\kappa_s)}}.$$

Substituting in the definition of a von Mises distribution, rearranging, and using the fact that for both distractor distributions in our experiment, $\mu = 0$, we find

$$d_i = \kappa\cos(x_i) + \log \frac{I_0(\kappa_s)}{I_0\left(\sqrt{\kappa^2+\kappa_s{}^2+2\kappa\kappa_s\cos(x_i)}\right)}. \tag{15}$$

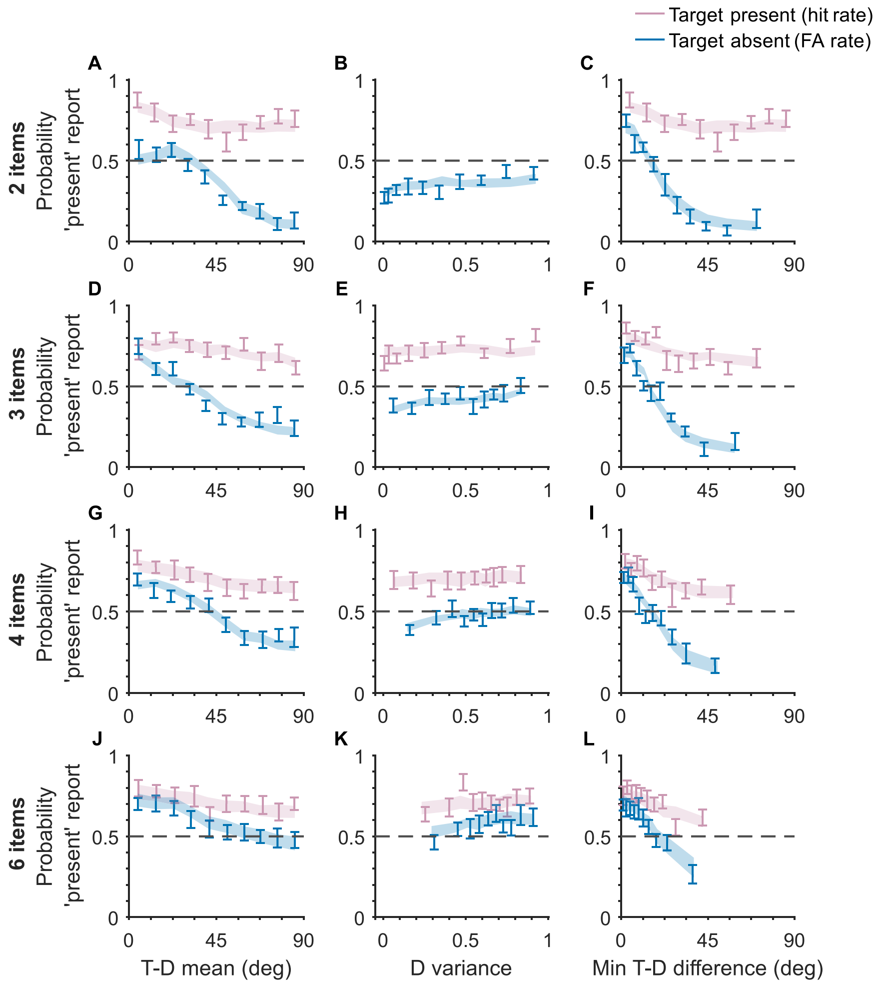

Figure 18. Effect of distractor statistics in the uniform environment, at different numbers of items.

For the case of uniform distractors, $\kappa_s = 0$, and we have,

$$d_i = \kappa \cos(x_i) - \log(I_0(\kappa)).$$

Substituting these expressions into Equation 13 will give us the log-likelihood ratio. In turn, using the log-likelihood ratio in (12) gives us the optimal observer's decision rule. That is, it tells us, for any combination of measurements $\mathbf{x}$, what the optimal observer would do.

Appendix D: Heuristic observer model fits

In the main text, we compared behavior simulated using the fitted Model 1 to behavior observed in the real data. Here we provide the corresponding plots for Model 3 in Figures 15–19.

Appendix E: Parameter estimates

Across participants, we computed the median parameter estimate for each parameter, along with the 25th and 75th percentiles. We present the parameters in two ways. In Table 7, we present the parameters in forms that are easier to interpret. In Table 8, we present the parameters in the forms in which they were fitted.

As noted in the main text, the median lapse rate was very high in Model 1 (0.31). It was lower in Model 3 (0.14), although this still seems fairly high. When interpreting this value, it is important to bear in mind that, aside from one participant who could not complete all sessions, no participants and no trials were excluded, and no performance-based exclusion criteria were applied. Additionally, the models used do not explicitly account for a number of possible sources of variability in responses, such as variable encoding precision (Mazyar et al., 2012) and computational imprecision (Stengård & van den Berg, 2019; Drugowitsch, Wyart, Devauchelle, & Koechlin, 2016). Explicitly including

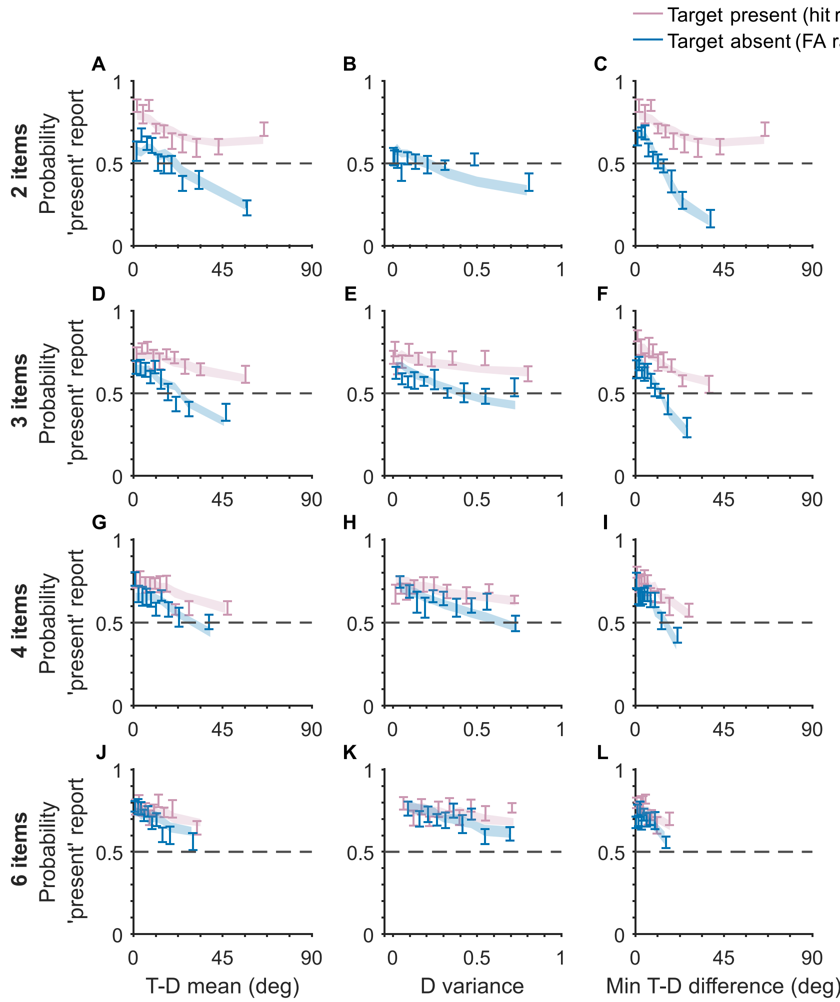

Figure 19. Effect of distractor statistics in the concentrated environment, at different numbers of items.

these sources of variability would likely lower the estimated lapse rate parameter.

Stengård and van den Berg (2019) performed an orientation discrimination task using single ellipses presented for 67 ms. They reported estimated sensory noise standard deviations ranging from approximately 1° to 8°. Sensory noise will increase as the number of items in the display increases and, as noted, the computations required to perform visual search likely introduce additional noise (Mazyar et al., 2012; Stengård & van den Berg, 2019; Drugowitsch et al., 2016). Mazyar et al. (2012) reported mean values for $p_{present}$ of approximately 0.42 to 0.54.

Appendix F: Parameter bounds

During fitting with Bayesian adaptive direct search (BADS), we applied bounds to the values that the parameters could take and specified plausible bounds within which we expected to find the parameter values. The bounds are specified below in two ways. In Table 9, we present the bounds on the parameters in forms that are easier to interpret. In Table 10, we present the bounds used on the parameters, when the parameters are considered in the actual form in which they were fitted. (Note: Several variables were log transformed for fitting.) Initial parameter values were drawn from uniform distributions on the interval between the plausible lower and upper bounds shown in Table 10. For κ and ρ, sets of initial values were drawn until a set was drawn in which the parameters decreased monotonically with increasing number of items on the screen.

Appendix G: Problems with local maxima

For each model and participant, we performed maximum likelihood fitting 40 times. Fitting 40 times allowed us to estimate the probability that our best fits

Model	Parameter	Median	25th percentile	75th percentile
1	σ	6.9	6.0	12
		10	7.6	18
		10.0	7.9	16
		15	9.5	33
	λ	0.31	0.29	0.47
	$p_{present}$	0.57	0.53	0.62
2	σ	6.5	6.0	11
		7.3	6.9	13
		9.9	6.8	29
		12	9.4	29
	λ	0.37	0.33	0.54
	$p_{present}$	0.60	0.52	0.71
	σ_o	43	30	89
3	σ	11	9.5	14
		12	10.0	18
		17	12	26
		19	16	41
	ρ (uniform.)	15	11	20
		13	12	16
		13	12	16
		13	10	16
	ρ (conc.)	11	8.5	16
		9.9	7.7	12
		9.5	8.4	12
		9.5	7.8	14
	λ	0.14	0.087	0.28
4	σ	9.2	8.6	13
		10	8.9	17
		14	11	26
		20	13	28
	ρ	12	8.8	16
		12	9.2	14
		11	9.7	14
		12	9.1	17
	λ	0.22	0.18	0.31

Table 7. Parameter estimates in a more intuitive form. σ, σ_o, and ρ are specified in degrees. The computation of σ and σ_o is described in the main text. ρ is measured in physical degrees, as opposed to degrees of the circular space defined by the Gabors (which as discussed in the main text only runs from -90° to 90° physical degrees).

Model	Parameter	Median	25th percentile	75th percentile
1	$\log\kappa$	2.9	1.9	3.2
		2.1	1.2	2.7
		2.2	1.3	2.6
		1.4	0.17	2.3
	λ	0.31	0.29	0.47
	$p_{present}$	0.57	0.53	0.62
2	$\log\kappa$	3.0	2.0	3.1
		2.8	1.7	2.9
		2.2	0.44	2.9
		1.8	0.45	2.23
	λ	0.37	0.33	0.54
	$p_{present}$	0.60	0.52	0.71
	$\log\kappa_o$	−0.37	−4.1	0.37
3	$\log\kappa$	2.1	1.6	2.3
		1.8	1.2	2.2
		1.3	0.56	1.8
		1.1	−0.29	1.4
	$\log\rho$ (uniform)	−0.67	−0.92	−0.37
		−0.76	−0.85	−0.57
		−0.77	−0.86	−0.57
		−0.78	−1.0	−0.58
	$\log\rho$ (conc.)	−0.99	−1.2	−0.60
		−1.1	−1.3	−0.85
		−1.1	−1.2	−0.92
		−1.1	−1.3	−0.73
	λ	0.14	0.087	0.28
4	$\log\kappa$	2.3	1.7	2.4
		2.1	1.3	2.4
		1.6	0.61	2.0
		1.0	0.47	1.7
	$\log\rho$	−0.85	−1.2	−0.56
		−0.91	−1.1	−0.71
		−0.95	−1.1	−0.74
		−0.84	−1.1	−0.52
	λ	0.22	0.18	0.31

Table 8. Parameter estimates in the form in which they were fitted. Parameters are defined in the main text. ρ is in radians in the circular space defined by the Gabors. As discussed in the main text, this circular space only corresponds to the angles between -90° to 90° in physical space.

Parameter	Lower bound	Plausible lower bound	Plausible upper bound	Upper bound
σ	1.1	2.9	88	100
ρ	0.071	1.8	68	90
λ	0	0.0050	0.40	1.0
$p_{present}$	0	0.20	0.80	1.0
σ_o	1.1	2.9	100	100

Table 9. Parameter bounds in a more intuitive form. σ, σ_o, and ρ are specified in degrees. The computation of σ and σ_o is described in the main text. ρ is measured in physical degrees, as opposed to degrees of the circular space defined by the Gabors (which as discussed in the main text only runs from -90° to 90° physical degrees).

Parameter	Lower bound	Plausible lower bound	Plausible upper bound	Upper bound
$\log \kappa$	-6.0	-4.0	4.6	6.6
$\log \rho$	-6.0	-2.8	0.86	1.1
λ	0	0.0050	0.40	1.0
$p_{present}$	0	0.20	0.80	1.0
$\log \kappa_o$	-6.0	-6.0	4.6	6.6

Table 10. Bounds on the parameters, when the parameters are considered in the form in which they were fitted. Parameters are defined in the main text. ρ is in radians in the circular space defined by the Gabors. As discussed in the main text, this circular space only corresponds to the angles between -90° to 90° in physical space.

were reaching the true maximum likelihood, as opposed to getting stuck in local maxima (see supplementary methods of Acerbi et al., 2018). To do this, we looked at how many fits ended up close to the best value of the likelihood found. If few fits get close to the maximum found, this suggests that the optimization algorithm is struggling, and the global maximum may not have been found. For two models (2 and 3), many fits did not end up close to the best value of the likelihood (Figure 20A), suggesting problems with fitting and that an even greater value of the likelihood might exist but has been missed.

To explore the possibility of issues with local maxima, we ran a further 40 fits for each model and participant, starting from the 40 points found in the first round of fitting. To reduce noise in the likelihood function, we simulated 5,000 sets of measurements and associated decisions per trial, instead of 1,000 as before. To make this approach computationally tractable, each likelihood evaluation, we drew 5,000 samples from each von Mises distribution (one for each value of κ) and resampled with replacement from these when von Mises samples were required. This improved the number of fits ending close to the best likelihood found for Model 3, although there was still some evidence of potential issues with Model 2 (Figure 20B).

Since the results of the model comparison remained largely unchanged, we report the results for the first 40 fits in the main text.

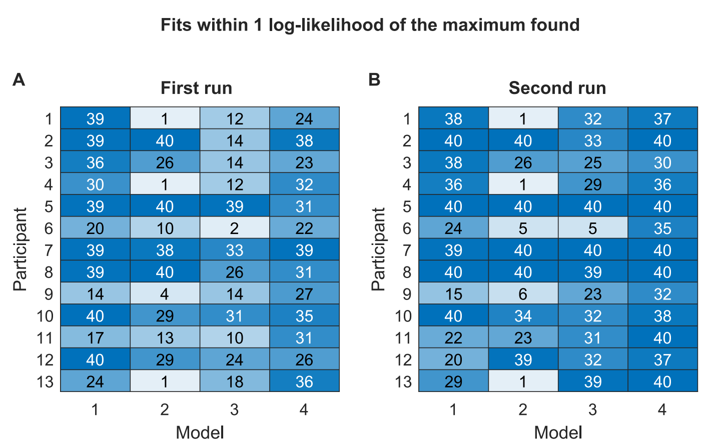

Figure 20. Number of fits out of 40 resulting in a log-likelihood within 1 point of the maximum log-likelihood found. Model numbers refer to Table 3. More saturated colors represent higher success rates. (A) First run. (B) Second run. The differences between the runs are described in the text.