Cognition 153 (2016) 196–210

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Original Articles

Building ensemble representations: How the shape of preceding distractor distributions affects visual search

Andrey Chetverikov [a,b,c,*], Gianluca Campana [d,e], Árni Kristjánsson [a]

[a] Laboratory for Visual Perception and Visuomotor Control, Faculty of Psychology, School of Health Sciences, University of Iceland, Reykjavik, Iceland
[b] Cognitive Research Lab, Russian Academy of National Economy and Public Administration, Moscow, Russia
[c] Department of Psychology, Saint Petersburg State University, Saint Petersburg, Russia
[d] Department of General Psychology, University of Padova, Padova, Italy
[e] Human Inspired Technology Research Centre, University of Padova, Italy

A R T I C L E I N F O

Article history:
Received 28 November 2015
Revised 27 April 2016
Accepted 29 April 2016
Available online 24 May 2016

Keywords:
Visual search
Ensemble representation
Summary statistics
Priming of pop-out
Line orientation

A B S T R A C T

Perception allows us to extract information about regularities in the environment. Observers can quickly determine summary statistics of a group of objects and detect outliers. The existing body of research has, however, not revealed how such ensemble representations develop over time. Moreover, the correspondence between the physical distribution of features in the external world and their potential internal representation as a probability density function (PDF) by the visual system is still unknown. Here, for the first time we demonstrate that such internal PDFs are built during visual search and show how they can be assessed with repetition and role-reversal effects. Using singleton search for an oddly oriented target line among differently oriented distractors (a priming of pop-out paradigm), we test how different properties of previously observed distractor distributions (mean, variability, and shape) influence search times. Our results indicate that observers learn properties of distractor distributions over and above mean and variance; in fact, response times also depend on the shape of the preceding distractor distribution. Response times decrease as a function of target distance from the mean of preceding Gaussian distractor distributions, and the decrease is steeper when preceding distributions have small standard deviations. When preceding distributions are uniform, however, this decrease in response times can be described by a two-piece function corresponding to the uniform distribution PDF. Moreover, following skewed distributions response times function is skewed in accordance with the skew in distributions. Indeed, internal PDFs seem to be specifically tuned to the observed feature distribution.

1. Introduction

Unlike the artificial displays used in many laboratory experiments, most real life scenes are internally consistent. Colors, lines and shapes all stem from objects in the world and we expect properties of one part of an object to be relatively similar to properties of another part of this object. Moreover, similar objects, such as leaves on trees or windows of a building, often appear together. The ability to extract regularities in the external world would therefore be useful for guiding vision. Studies of "ensemble representations" do, indeed, demonstrate that we are reasonably good at judging the summary statistics of stimuli (Albrecht & Scholl, 2010; Alvarez, 2011; Alvarez & Oliva, 2008; Ariely, 2001; Chong & Treisman, 2003, 2005; Dakin & Watt, 1997; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001). Such statistics could be more useful for predicting the environment than individual features, especially in peripheral vision (Balas, Nakano, & Rosenholtz, 2009; Rosenholtz, Huang, Raj, Balas, & Ilie, 2012). Yet, little is known about how the visual system represents such summary statistics and how such representations develop over time.

2. What is encoded?

Human observers' have the ability to estimate measures of central tendency such as a mean, for position, orientation, size, motion speed or direction, and other features (see reviews in Alvarez, 2011; Haberman & Whitney, 2012). It is less clear whether observers encode more complex information about feature distributions. Atchley and Andersen (1995) found that observers were able to

* Corresponding author at: Department of Psychology, University of Iceland, Saemundargata 2, 101 Reykjavík, Iceland.
E-mail address: andrey@hi.is (A. Chetverikov).

http://dx.doi.org/10.1016/j.cognition.2016.04.018
0010-0277/

pick the set of moving dots having a different velocity distribution from others in mean or variance but not in skewness or kurtosis. Similarly, Dakin and Watt (1997) found that when observers are required to make judgements related to the orientation of two intermixed sets of stimuli, they utilize information about variance but not skewness of the overall distribution (see review in Dakin, 2015). Additionally, Michael, de Gardelle, and Summerfield (2014) found that when observers judge the average shape or color of stimulus sets, they are more efficient if feature variance is constant in consecutive sets. While higher variance made the task more difficult overall, switching from high to low variance or vice versa affected performance even further. Importantly, they also demonstrated that this effect is not simply due to changes in task demands, and therefore does not reflect differences in cognitive control. The authors argued that variance priming of this sort is fast and automatic as it works for intervals as short as 100 ms and occurs both for task-relevant and task-irrelevant features. Norman, Heywood, and Kentridge (2015) used perceptual adaptation to show that observers automatically encode variance in orientation. Observers who adapted to a set of Gabor patches with high orientation variance subsequently perceived a test set as less variable while those who adapted to a low-variance set perceived a test set as more variable as compared to no adaptation. Following Morgan, Chubb, and Solomon (2008) and Norman et al. (2015) suggest that there might be a specific variance-detection mechanism broadly tuned to high or low variance. In sum, previous studies demonstrate that observers are able to detect variance of distributions in feature space. Unknown, however, is whether they encode variance simply as a range, use a distribution-based approximation, or a broad filter of some sort. Studies of observers' ability to extract more complex properties of distributions are scarce and the few existing ones (Atchley & Andersen, 1995; Dakin, 2015; Dakin & Watt, 1997) have yielded negative results.

A major methodological hurdle to the study of more complex statistics in ensemble representations involves a convenient way of assessing the distributions. For example, the point of subjective equivalence between two stimuli might be estimated by measuring thresholds. Similarly, repetition benefits in averaging (de Gardelle & Summerfield, 2011; Michael et al., 2014) can be used to show that information about variance is encoded. However, studying the point of subjective equivalence between two variances will yield little information about the representation of the variance because the variance can depend on range, standard deviation, absolute deviation, or on a number of other summary statistics; or might not even be statistical at all. A novel method for assessing ensemble perception is therefore needed.

3. Priming of pop-out as a way to assess ensemble representation

We propose a novel approach of studying ensemble representations and use it to show that observers encode the shape of distractor distributions in feature space over a series of visual search trials. We exploit the well-known "priming of pop-out" effect (Maljkovic & Nakayama, 1994) to assess learning of probability distributions. In priming of pop-out, trial-by-trial repetition of target or distractor features in a singleton search task leads to shortened response times (RTs), while switches between target and distractor features increase RTs even more than the appearance of novel features (Kristjánsson & Driver, 2008; Lamy, Antebi, Aviani, & Carmel, 2008; Wang, Kristjánsson, & Nakayama, 2005), and these effects are cumulative over multiple trials (Maljkovic & Nakayama, 1994). Such attentional priming can influence search at feature, feature-dimension, or object levels, depending on stimuli and task demands (Campana, Pavan, & Casco, 2008; Kristjánsson, 2006; Kristjánsson & Campana, 2010; Kristjánsson, Saevarsson, & Driver, 2013; Lamy et al., 2008; Meeter & Olivers, 2006; Muller, Reimann, & Krummenacher, 2003; Ásgeirsson & Kristjánsson, 2011). Moreover, not only specific features but the relationships between target and distractor features can be primed as well (Becker, 2008, 2010; Meeter & Olivers, 2014).

We suggest that priming of pop-out effects reveal internal models of target and distractors learned by observers. If the model is accurate and search parameters are consistent over time, RTs are faster. By varying the degree to which previously repeated distributions match the new one, the internal model can be assessed. Of particular interest is "distractor to target" role-reversal, that is, when the target has the features of a previously learned distribution of distractors. Such role-reversal usually results in particularly long search times (Kristjánsson & Driver, 2008). For example, if distractors are blue and targets red over several trials and the target then becomes blue, search is more difficult than when the target becomes yellow. The target represents a single point in a feature space and we suggest that it can therefore be used as a "probe" to determine the internal representation of probability of the distractor at this point in feature space. This probability can vary, for example, depending on the number of repetitions or the distribution of distractors features. If distractors have been blue and only blue, the probability of observing distractors in the "blue" part of feature space is high and for the others parts it is low. Hence, search will be longer when the target is blue. If distractors have been blue and sometimes green, then the probability of observing distractors is high in the blue part, lower in the green part, and lowest in other parts of the feature space. Hence, search times will be the slowest when the target is blue, somewhat faster if the target is green, and the fastest when target is neither green nor blue. By varying the position of a target in a feature space relative to a preceding distractor distribution it is possible to estimate observers' internal representations of the probability density functions (PDF) of that distribution. Unlike the methods previously used in studies of summary statistics this will allow a relatively precise estimation of corresponding internal models.

Comparison with statistical data analysis can be appropriate here: The same statistics (i.e., mean and standard deviation) can describe two qualitatively different distributions, while plotting the distribution density immediately reveals the differences (Fig. 1). Similarly, probing different points in feature space with role-reversal effects allows the "plotting" of the underlying probability density of previously learned distractor distributions in feature space using RTs as an indicator of probability.

In sum, distribution of stimuli features can be represented as a PDF that describes the relative probability that a feature will take a given value. We assume that the visual system operate in a similar way, assigning probabilities of observing specific stimuli or stimulus categories (such as target or distractors) for parts of feature space in accordance with previous experience. In a visual search task, response times would, among other factors, depend on the probabilities learned in previous trials. The exact mapping between physical features of stimuli and the probabilities assigned to them by observers and between these probabilities and response times is unknown. Although there are well-known psychophysical laws describing the mapping of single stimuli to their representations, there is no guarantee that these laws can be applied in the same way independently of stimulus probability, distance from the mean, spatial density, etc. (related issues are studied in case of texture representation, e.g. Rosenholtz, 2014, but probabilities do not play a central role there). But we can assume that the relationship between physical and perceptual domain is monotonic: the lower the probabilities in physical space, the lower are the probabilities in perceptual space and the

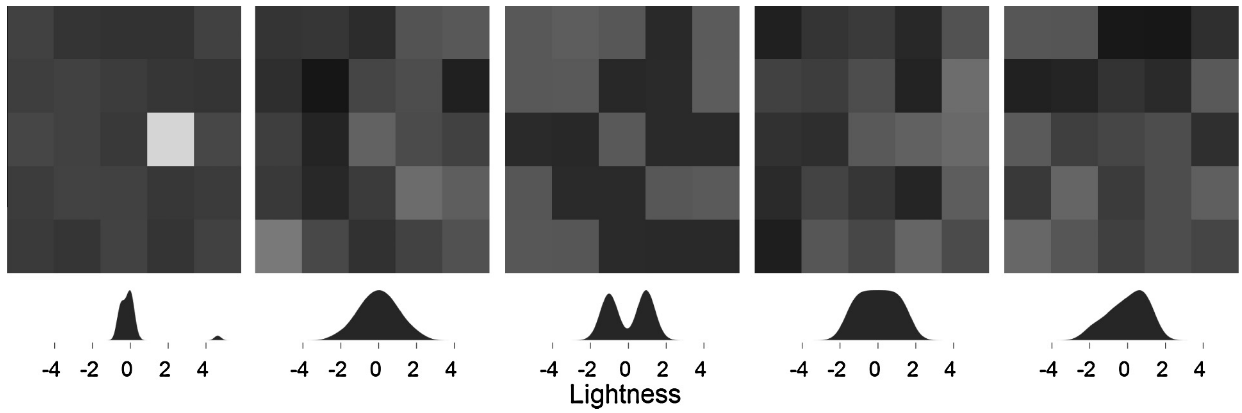

Fig. 1. Five examples of distributions of 25 tiles varying in lightness. While summary statistics (means and standard deviations) in each case are the same, plotting distribution densities (bottom row) reveals qualitative differences. These differences are also immediately perceivable when each distribution is translated to the lightness of tiles (top row).

response times are subsequently longer. Knowing the input (probabilities in physical space) and the output (response times) we can assess the corresponding internal model.

4. Summary statistics in visual search

Summary statistics in visual search have mostly been studied by researchers interested in effects of distractor heterogeneity (Avraham, Yeshurun, & Lindenbaum, 2008; Mazyar, van den Berg, Seilheimer, & Ma, 2013; Nagy, Neriani, & Young, 2005; Rosenholtz, 2001; Utochkin, 2013; Vincent, Baddeley, Troscianko, & Gilchrist, 2009), which has detrimental effects on visual search performance (Duncan & Humphreys, 1989). Importantly, Rosenholtz (2001) demonstrated that this effect can be successfully modelled using a best-normal model. This model describes a distractor distribution as the best Gaussian approximation available, involving the assumption that observers might fail to represent complex distributions and instead represent them only with means and standard deviations. The success of this model shows that observers may indeed represent distractor feature distributions as a set of summary statistics rather than as individual elements.

This conclusion is further supported by data obtained by Rosenholtz et al. (2012). They assumed that peripheral vision allows us to extract summary statistics but not information about individual items. Using scrambled stimuli ("mongrels") with the same summary statistics as regions of visual search displays, they found that several classical visual search asymmetries (e.g., search for Q among Os is more efficient than search for O among Qs) can be predicted by differences in discriminability between such mongrels. The results indicate that summary statistics can guide visual search by providing useful "snapshots" of the unattended regions.

However, showing that a model can predict performance does not necessarily imply that such models actually describe what occurs during human visual search. Corbett and Melcher (2014) took a step in that direction by demonstrating that stable summary statistics increase search efficiency even when individual items change from trial to trial. The effect of stable summary statistics accumulated over trials reach a plateau after four repetitions. They only studied effects of the distribution mean, however, keeping all other parameters constant.

How do observers represent distractor distributions in visual search and what predictions can be made regarding priming of pop-out? Based on the model proposed by Rosenholtz (2001) we might predict that observers will use Gaussian approximations for a relatively useful description of the distractors. From this perspective, further information about the shape of the distribution, such as its skewness, kurtosis, and higher order moments, is unnecessary and will not be encoded. Accordingly, RTs should depend on the interaction between present target position relative to the mean and standard deviation of previous distractor distributions. On the other hand, studies of orientation variance suggest that a specific mechanism for variance estimation may exist (Morgan et al., 2008; Norman et al., 2015). It is possible that observers simply encode variance as relatively low or relatively high along with the mean. We would then expect priming effects from the variance (Michael et al., 2014), and priming effects from the mean (Corbett & Melcher, 2014). The crucial point of that prediction is that the two effects should be independent, so that there will be no interaction between the position of the target (in feature space) relative to the previously learned mean of distractor distributions and its standard deviation. This account would also predict that the shape of the previously learned distribution should have no effect on RTs.

The final possibility is that observers encode information about the distribution shape. This could be possible using either a higher-order statistics or by density estimation at different points of feature space (these mechanisms are discussed in more details in General Discussion). Previous attempts to find effects of distribution parameters other than mean and standard deviation have involved single presentations where observers make explicit judgments about the distributions (Atchley & Andersen, 1995; Dakin & Watt, 1997). In our current experiment, in contrast, observers are exposed to multiple examples of the same distractor distribution over several search trials and no explicit judgments about distributions are needed.

Note that learning distribution shapes may have a functional significance. For example, if a uniform distribution is approximated as Gaussian this will lead to an underestimation of distractors from the tails of the distribution, and unexpected distractors should slow search down. This entails that after viewing several examples of stimuli drawn from the same distribution, observers need to learn not only its mean and standard deviation but also other parameters.

In this study we bridge two previously disparate research areas: serial effects in visual search and ensemble representations. We believe that both areas will benefit both theoretically and empirically from such integration. In fact, for serial effects this will create a new way of describing representations of targets and distractors; and for ensemble studies this will provide a way to assess how ensembles are formed and applied. Our main assumption is that representations underlying serial effects are not of specific distractors and target features, but rather representations of target and distractor feature distributions. Based on the three accounts outlined above we measured observers' ability to learn the parameters of distractor distributions. In Experiment 1, we test whether observers learn the mean and standard deviation separately or rather if the two parameters are learned together and may be treated as a probability density function in feature space. In Experiment 2 we further test whether observers can learn the shape of distractor distributions (i.e. uniform vs. Gaussian distributions) in addition to their mean and standard deviation. In Experiments 3A, 3B and 3C we aim to replicate the findings of Experiment 2 controlling for potential confounds; and in Experiment 4 we test whether observers can also learn another characteristic of the shape of distractor distributions, namely, their skewness.

5. Experiment 1

5.1. Participants

Ten observers (4 female, 20–30 years old, mean age 25.6 years; two were aware of the goal of the experiment, the rest naive) took part in two experimental sessions taking approximately 40 min each. The observers were students or staff from the cognitive psychology laboratory at the Faculty of Psychology, St. Petersburg State University, Russia.

5.2. Method

The experiment was run on a DELL Vostro 5470 laptop with a 14 in. display with 1366×768 pixel resolution using PsychoPy (Peirce, 2007). Viewing distance was ∼57 cm. Observers searched for an oddly oriented line in a search array of 36 lines arranged in a 6×6 grid subtending 16 deg $\times$ 16 deg in the center of a screen. The length of each line was 1 deg and their positions were slightly jittered by randomly adding a value between ±0.5 deg to both vertical and horizontal coordinates.

The distractor lines on each trial were picked randomly from a Gaussian distribution with a standard deviation (distractor standard deviation – DSD) of 5, 10, or 15 deg (see Fig. 2). The trials were organized in streaks of 5–7 trials, where the mean and standard deviation of the distractor distribution were constant (Fig. 3). Target orientation was set randomly for each streak so that target-to-distractor distance ranged from 60 to 120 deg. Within-streak target orientation was constant.

We used 9 gradations of shifts between distribution means from −80 to +80 deg in 20 deg steps (Fig. 4). Each gradation was repeated 5 times for each distribution pair (i.e., a transition from a DSD = 5 to a DSD = 15 and a +40 deg distance between means was repeated 5 times). The order of streaks was randomized for each observer by creating a string of minimal length from a randomly shuffled sequence of pairs of distributions with a different DSD. In total, there were 273 streaks (minimal possible length for all combinations of conditions) in each session for each observer, or approximately 1638 trials (as the number of trials within a streak was chosen randomly from 5 to 7, the total trial number varied).

Observers pressed the 'i' key if the target line was in the upper three rows and the 'j' key if the target line was in the lower three rows. If observers made an error, the word "ERROR!" appeared in red letters for 1 s. Observers were informed that their performance would be scored and were encouraged to respond as fast and as accurately as possible to increase their scores. The score for each trial was computed as $\text{Score} = 10 + (1 - \text{RT}) * 10$ for correct answers, where *RT* is response time in seconds, but for errors $= -|\text{Score}| - 10$. The score from the previous trial was shown in the top left corner (in green if positive, red if negative) of the screen along with the trial number and the total number of trials. The total score was shown during resting periods after the 100th and 200th streaks. The resting periods and scores were intended to increase observers' vigilance and concentration. Resting time was unlimited, but observers were encouraged not to take too much time for rest.

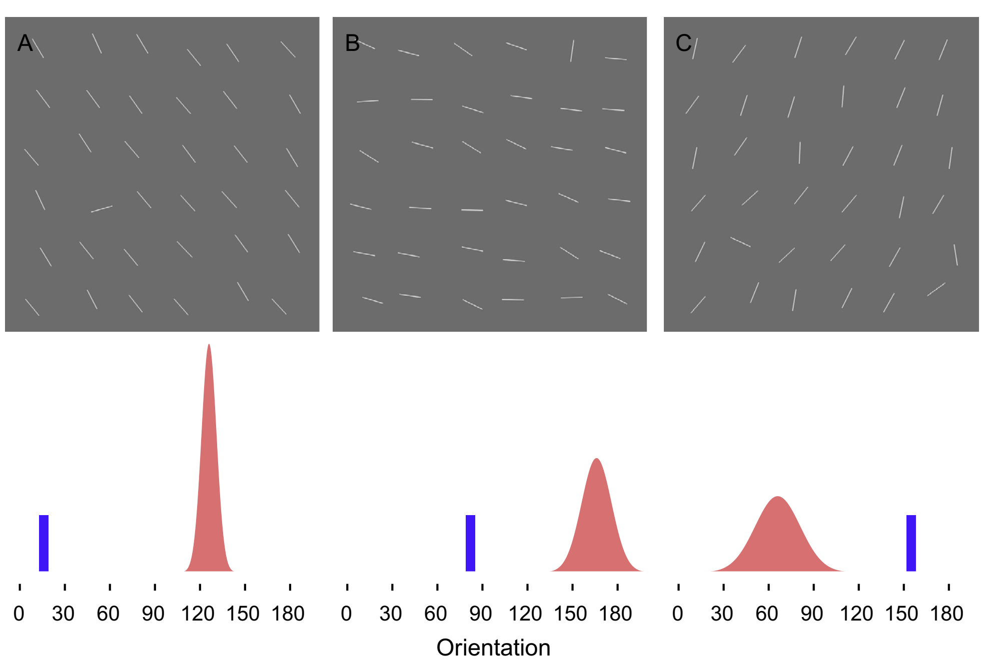

Fig. 2. Top row: examples of the stimuli in Experiment 1; (A) Distractor standard deviation (DSD) = 5, distractor mean = 126, target orientation = 16; (B) DSD = 10, distractor mean = 166, target orientation = 82 and (C) DSD = 15, distractor mean = 66, target orientation = 155. Bottom row: schematic depictions of targets (in blue) and distractor distributions (in red) from the corresponding examples in the top row. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

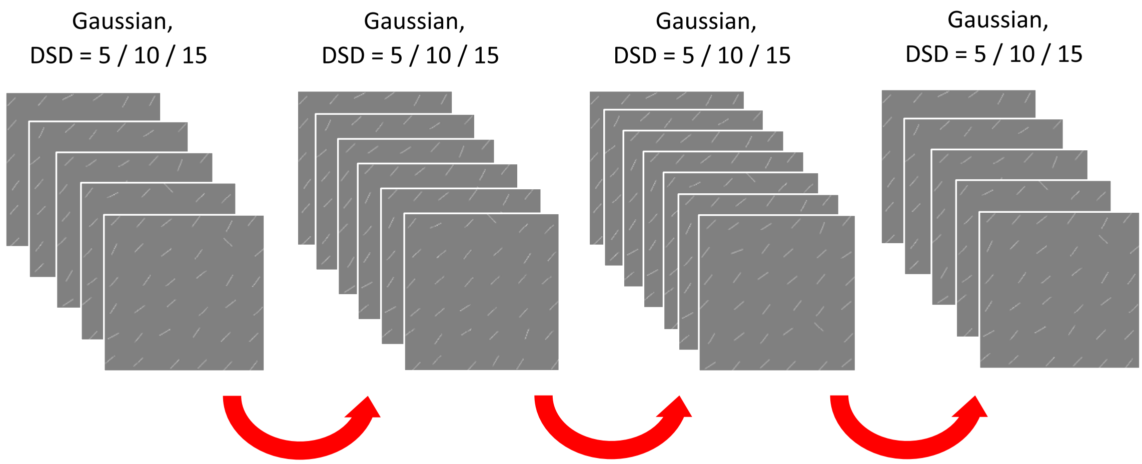

Fig. 3. Sequence of streaks in Experiment 1. Within each streak DSD and target orientation were constant.

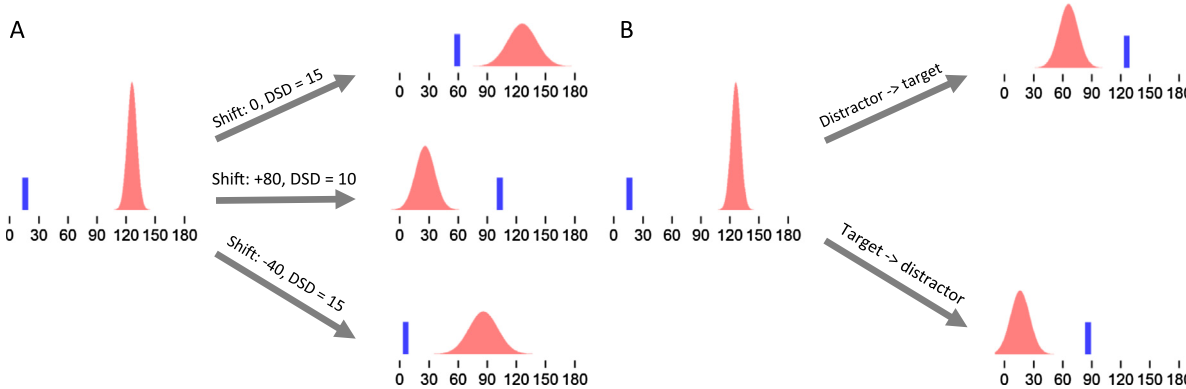

Fig. 4. Examples of distribution shifts used (panel A) and examples of role-reversals (panel B) in Experiment 1. In total, nine distractor distribution shifts were possible. Target orientation was chosen randomly for each streak, see details in text.

5.3. Results

We used linear mixed-effects regression (LMER) to analyze effects of target-distractor distance and distribution shifts on RTs. LMER takes non-independence in the data due to repeated measures on the same observers into account without data aggregation (see e.g., Jaeger, 2008). LMER does not provide p-values due to inherent uncertainty in the computation of the degrees of freedom for complex models. Instead, t-criteria values above 2 or below -2 can be used as guidance for significance, roughly corresponding to the usual $p < 0.05$ level (Baayen, Davidson, & Bates, 2008; Bates, 2006).

5.3.1. Average performance

As expected, RTs were longer and accuracy lower with larger distractor distributions (DSD, Table 1). A one-way repeated-measures ANOVA showed a significant effect of DSD on both RTs ($F(2,18) = 96.35$, $p < 0.001$, $\eta^2_G = 0.417$) (here and later we analyze log-transformed RTs but figures and tables show unmodified RTs for clarity), and accuracy ($F(2,18) = 44.25$, $p < 0.001$, $\eta^2_G = 0.497$).

Table 1
Response times and accuracy as a function of the standard deviation of the distractor distribution in Experiment 1.

DSD	Accuracy		RT (ms, correct responses)	
	M	SD	M	SD
5	0.95	0.02	620	57
10	0.93	0.03	713	80
15	0.88	0.05	820	106

As Fig. 5 shows, the more different the target was from the distractors, the easier the search. LMER demonstrated a negative linear target-distractor distance effect on RTs ($B = -1.94$ (0.37), $t = -5.22$), qualified by a quadratic effect ($B = 3.66$ (0.37), $t = 9.83$).

5.3.2. Repetition effects

Within streaks, RTs decreased following the first repetition, reaching a plateau approximately at the third trial. Accuracy tended to increase after the first trial in a streak (Fig. 6) for DSDs of 10 and 15. Linear mixed-effects regression with Helmert contrasts (comparing each level of the trial number with the average of the following levels) on RTs confirmed these observations indicating that the first and second trials in each streak differed from the later trials for $DSD = 10$ and $DSD = 5$ ($t < 2$ for the trials after the second one), but for $DSD = 15$ only the first trial differed from the rest. For accuracy, only the first trial for $DSD = 15$ differed from the rest. In sum, repetitions affected RTs more than accuracy.

5.3.3. Distribution shifts

We only analysed the first correct trials in each streak that were preceded by correct trials to avoid post-error slowing effects (Danielmeier & Ullsperger, 2011). Fig. 7 shows that RTs followed a U-shaped function: the larger the shift between means of distributions, the slower participants responded ($B = 5.17$ (0.37), $t = 14.17$ for the quadratic effect of distribution shift), and the more errors they made ($B = 10.12$ (3.37), $t = 3.00$).

Fig. 8 shows that not only the mean, but also the SD of previous distractor distributions (DSD) influences RTs. A mixed-effects regression controlling for the effects of target orientation,

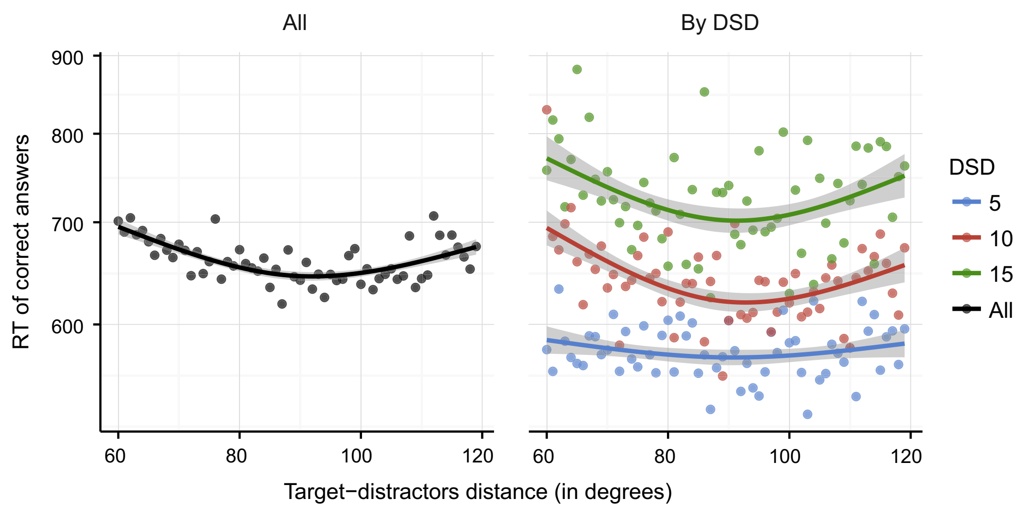

Fig. 5. Quadratic trends of target-distractor distance effects in Experiment 1. Dots show mean response times, shaded areas show 95% confidence intervals based on a quadratic regression fit.

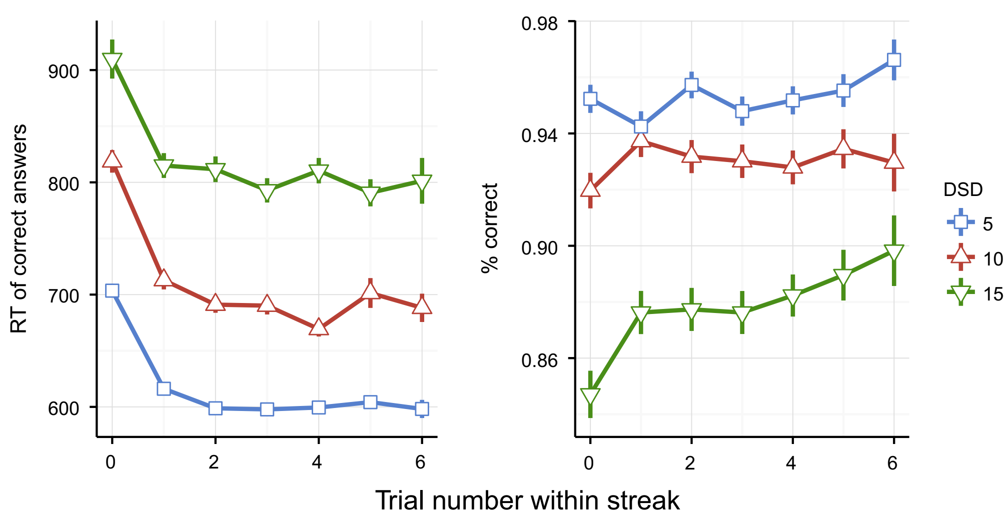

Fig. 6. Repetition effects within streak in Experiment 1. Bars show ±1 SEM.

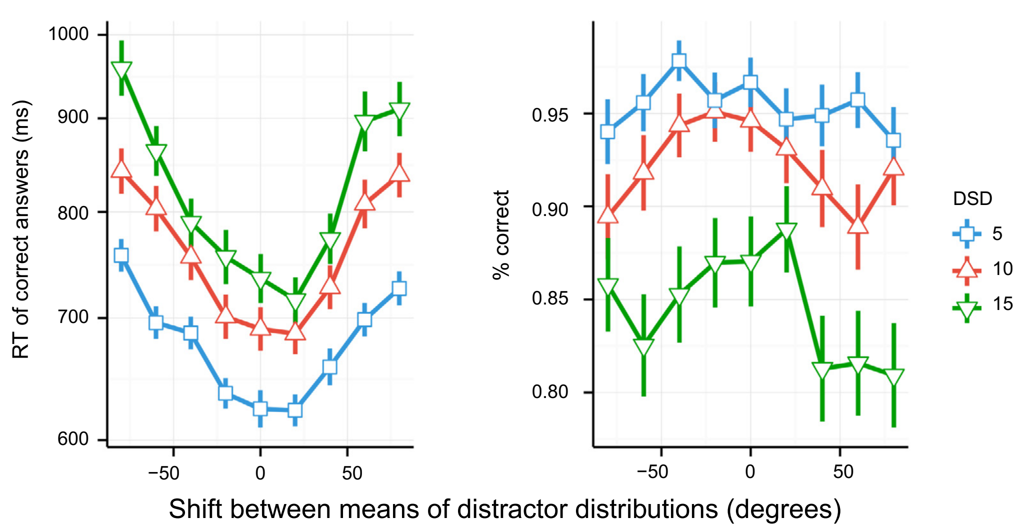

Fig. 7. Response times and accuracy as function of changes in distractor distributions between streaks in Experiment 1. Bars show ±1 SEM.

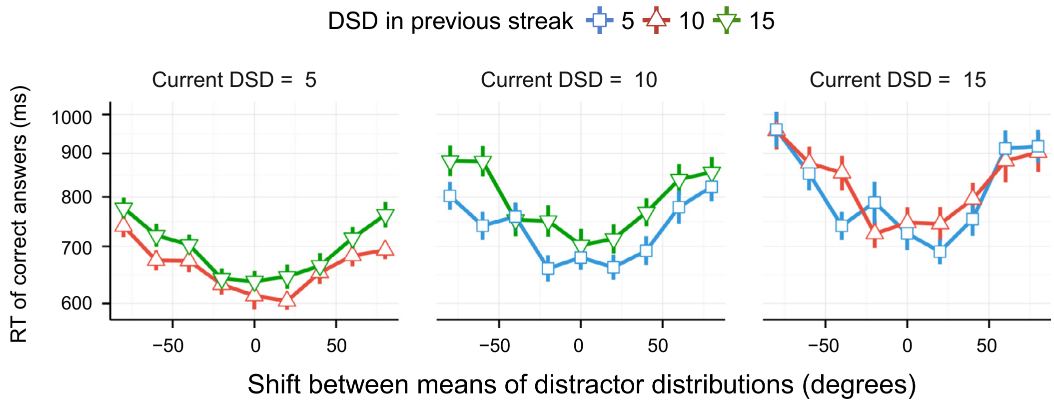

Fig. 8. Response times as function of previous and current distractor distributions and the shift between their means in Experiment 1. Bars show ±1 SEM.

target-to-distractor distance, and a shift in means showed that for *DSD* = 5, search takes longer when previous *DSD* = 15 than when previous *DSD* = 10, $t = 3.57$, $p < 0.001$. Similarly, for *DSD* = 10 search takes longer when previous *DSD* = 15 than when previous *DSD* = 5, $t = 4.07$, $p < 0.001$. However, for *DSD* = 15 previous DSD does not seem to influence RTs. No significant effects of previous DSD on accuracy were found in any condition.

5.3.4. Role-reversals

A target with an orientation close to the mean of the previous distractor distribution (distractor to target switch) or conversely, distractor distributions with a mean close to the previous target (target to distractor switch) may have detrimental effects on performance over and above distribution shifts because of target/distractor role reversal effects (Kristjánsson & Driver, 2008; Lamy et al., 2008). We analysed distractor to target switches treating RTs on the first trial in a streak as a function of the distance between the *current target* and the mean of the *previous distractor* distribution (target distance to previous distractor, T-PD). Given that target-distractor distance was randomized for each streak this effect is partially independent of the shift between means of distractor distributions, and could therefore be analysed separately.

If a distractor distribution is inhibited during repetitions, RTs should increase when a target falls within that distribution on a new streak. Fig. 9 shows that role-reversals have a strong effect for all DSDs as RTs gradually increase with the decrease of T-PD distance. Gaussian distribution implies a non-linear decrease of probability and hence the effect of previous DSD in interaction with T-PD should also be non-linear. Fig. 9 shows evidence of such non-linear dependency. To quantify it we used two complimentary approaches. First, we split the T-PD distance into 15 deg bins (15 deg cover 99.97% of the distribution with SD = 5) and compared the RTs as a function of previous DSDs in each of the bins and for each current DSD, using mixed-effects regression with pre-set contrasts. RTs were different depending on the previous DSD ($t > 2$) for bins [15, 30] and [30, 45] for current DSD = 10, and for bin [0, 15] for current DSD = 5. We then used segmented regression (with the *segmented* package in R; Muggeo, 2003, 2008) to test, whether RTs decreased monotonically as a function of T-PD or whether there was a significant change in the rate of decrease at some

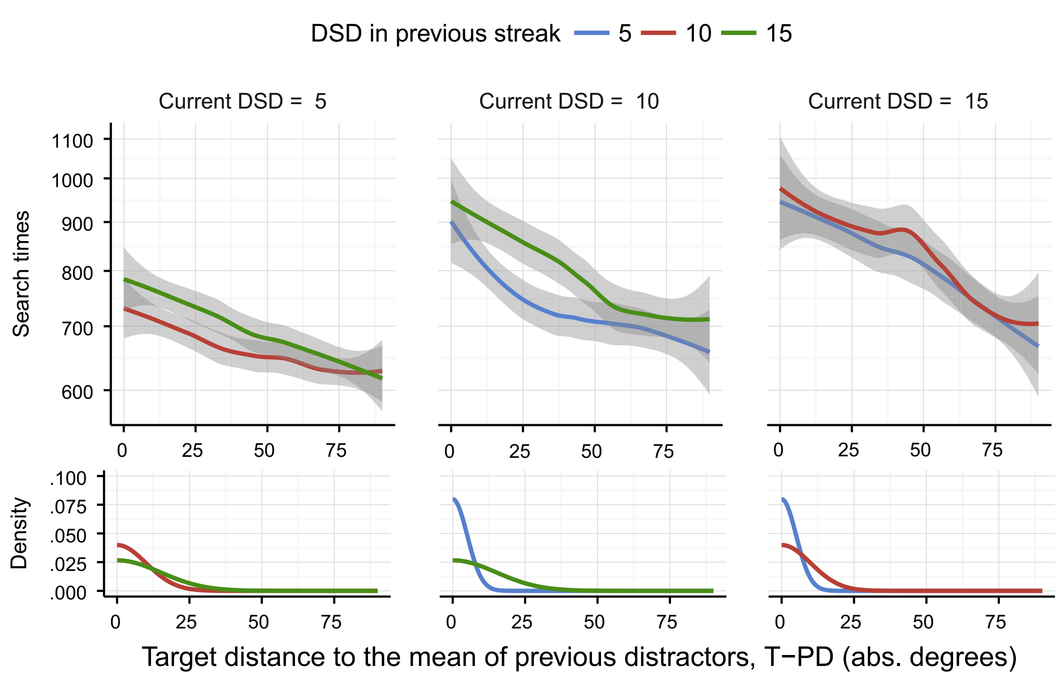

Fig. 9. Response times in Experiment 1 as a function of the distance between targets and previous distractor mean (top row) and corresponding probability densities (bottom row). Shaded areas show 95% confidence intervals based on local regression fit.

T-PD value. For $DSD = 15$, there were no significant breakpoints indicating a monotonic decrease of RT as function of T-PD regardless of previous DSD. For $DSD = 10$, however, there was a change in T-PD effect on RTs following $DSD = 5$ with a breaking point located at 9.75 (Davies test $p = 0.020$), but not following $DSD = 15$ (the best estimate for break point was at 67.00, $p = 0.161$). The effect of T-PD on RT following $DSD = 5$ changed from $B = -30.60$, 95% $CI = [-56.57, -4.62]$ to $B = -1.48$, 95% $CI = [-2.73, -0.23]$, that is, from a steeper to a shallower decrease. For DSD = 5 following DSD = 10 there was a tendency-level breaking point at 77.31 deg distance, $p = 0.082$, with slopes changing from $B = 2.17$, 95% $CI = [-8.68, 13.01]$ to $B = -1.40$, 95% $CI = [-2.24, -0.57]$, indicating a flat end of the RT pattern as function of the T-PD curve. For DSD = 5 following DSD = 15 no significant break point was found (the best estimate was at 4.97, $p = 0.210$).

The target to distractor switch effect, analysed as the distance between the mean of the *current* distractor distribution and the previous target (distractor distance to previous target, D-PT), could not explain the results. When both D-PT and T-PD are included in the regression model, the effect of T-PD is significant ($B = -0.31$ (0.04), $t = -7.75$), while the effect of D-PT is not ($B = -0.01$ (0.02), $t = -0.54$).

5.4. Discussion

Experiment 1 shows that observers can learn information about the variability of distractor sets during visual search. We replicated classic effects of distractor variability and target-distractor distance (Duncan & Humphreys, 1989). We further demonstrated that priming of pop-out (Kristjánsson & Campana, 2010; Maljkovic & Nakayama, 1994) can depend on a *statistic* of a feature distribution (mean orientation), instead of a specific feature value (orientation) replicating the results of Corbett and Melcher (2014). Analysis of POP effects further demonstrated that *SD*s of the preceding distractor distribution (DSD) affect RTs: following a more heterogeneous distribution, observers were slower in finding a target, though accuracy did not change.

The exact shape of the distribution (determined in this experiment by *SD*) affects RTs for role-reversals, when a target falls within the range of the previous distractor distribution. Learning of distribution shape modulates RTs on switch trials (Fig. 9, top rows) roughly following the PDF of the learned distribution (Fig. 9, bottom rows). The comparison between the response time following a distribution with $DSD = 5$ and that following a distribution with $DSD = 15$ (middle column) reveals a steeper decrease of RTs as a function of increasing distance between current target and a previous distractor (T-PD) for the former than for the latter. The non-linear relationship between T-PD and RTs is confirmed both by analysis of T-PD split into bins and by segmented regression and corresponds to the non-linear probability density function of distractor distributions.

These results suggest that mean and variance are not encoded independently. RTs change as a function of the interaction between mean and standard deviation of the preceding distribution. If the current target is far from the mean of the previous distractor distribution, RTs vary less as a function of its variance than when the target is within this range. This can be expected if observers encode mean and variance together to describe the distractor distribution. However, the variance of a preceding distribution has an effect even when the target is outside its range. Moreover, if present distractors have low heterogeneity ($DSD = 5$), the mean and variance of previous distributions have independent effects. It is possible then that variance may be encoded independently in addition to the co-encoding of mean and variance.

The interaction between the mean and variance of preceding distributions needs further explanation. Two models discussed in the introduction may explain this result: a normal approximation suggested by Rosenholtz (2001) or encoding of distribution shape. It is impossible to distinguish between them by the data from Experiment 1 as we only used Gaussian distributions. In Experiment 2 we therefore compared Gaussian and uniform distributions to further test the hypothesis that observers encode distribution shape. Additionally, the *SD* of the distribution in Experiment 1 was confounded with its range. The range of the distribution affects commonly used measures of distractor heterogeneity, such as *cover* – the ratio of distractor range in perceptual space to the distance between target and the nearest distractor (Avraham et al., 2008). To disentangle the effect of the range and the shape of the distribution, the range was kept constant.

6. Experiment 2

6.1. Participants

Ten observers (all but one previously participated in Exp. 1; the new naïve observer was trained on a singleton search task for 1000 trials) took part in two experimental sessions taking approximately 20–30 min each. The observers were students or staff from the cognitive psychology laboratory at the Faculty of Psychology, St. Petersburg State University, Russia.

6.2. Method

The same general procedure as in Experiment 1 was used, while this time the distractor lines on each trial were picked randomly from either a Gaussian or a uniform distribution. The Gaussian distribution had DSD = 10 (Gauss10) or DSD = 15 (Gauss15). The uniform distribution (Uniform) had a range of 60 deg ($\pm 2 * 15$ deg, the area accounting for 95% of Gauss15; Fig. 10). To keep the range exactly the same, we added two lines with an orientation of Mean + 30 deg and Mean − 30 deg to the Gauss15 and Uniform distributions and replaced all lines outside the ±30 deg range with novel ones. The range of both the Gauss15 and Uniform distributions was therefore 60 deg on each trial, but the PDFs of the distributions within that range differed.[1]

We used a Gaussian distribution with $DSD = 10$ as the testing one and the other two distributions as priming ones to compare the influence of distributions with the same range. The distribution with DSD = 10 was chosen to avoid floor and ceiling effects. We therefore counterbalanced the experiment with regards to the distribution shifts between streaks so that Gauss10 was preceded five times by each of two other distributions (Fig. 11) for each gradation of the distance between previous and current distribution means (−80 to 80 deg in 20 deg steps). In total, there were 180 streaks with a pause after 100 streaks.

6.3. Results

6.3.1. Average performance

Performance was comparable for the two distributions with the larger range, though for the uniform distribution RTs were longer and accuracy slightly lower (Table 2). A one-way repeated measures ANOVA revealed a significant effect of distractor distribution on both RTs, $F(2,18) = 53.31$, $p < 0.001$, $\eta^2_G = 0.234$, and accuracy,

[1] Note that this manipulation also changed the *SD* of the distributions: for the uniform distribution the resulting $SD = 18.26$ (on average on a given trial) while for the Gaussian distribution the $SD = 14.65$. The effect of *SD* was assessed in Experiment 3B. The labels in text for the Gaussian distribution show $SD = 15$ to account for the fact that the distractors lines were drawn from a distribution with this standard deviation although range constraints change the resulting value.

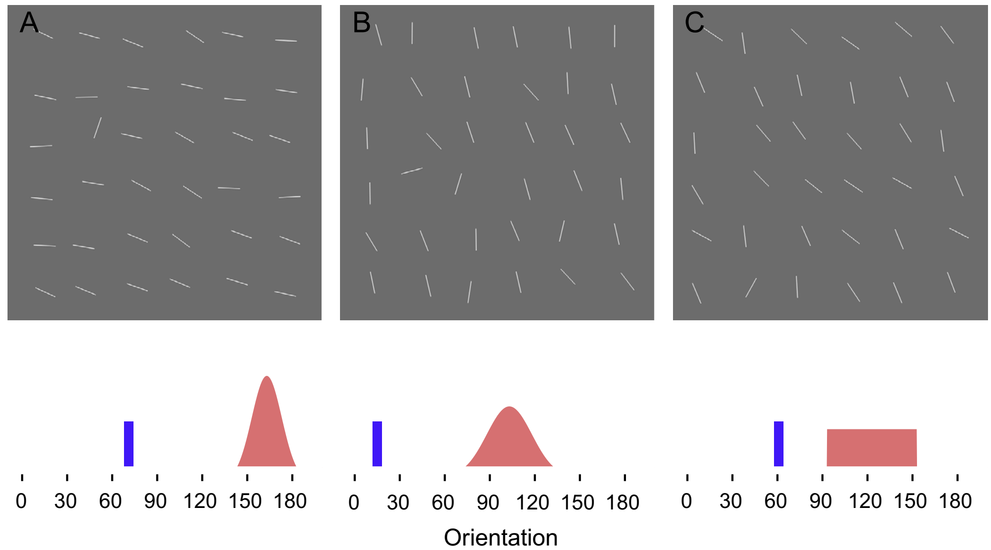

Fig. 10. Top row: example stimuli in Experiment 2; (A) Gauss10, distractor mean = 163 deg, target orientation = 71 deg; B: Gauss15 = 15 deg, distractor mean = 103 deg, target orientation = 15 deg; C: Uniform, distractor mean = 123 deg, target orientation = 61 deg. Bottom row: target (in blue) and distribution of distractors (in red) from the examples in the top row. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

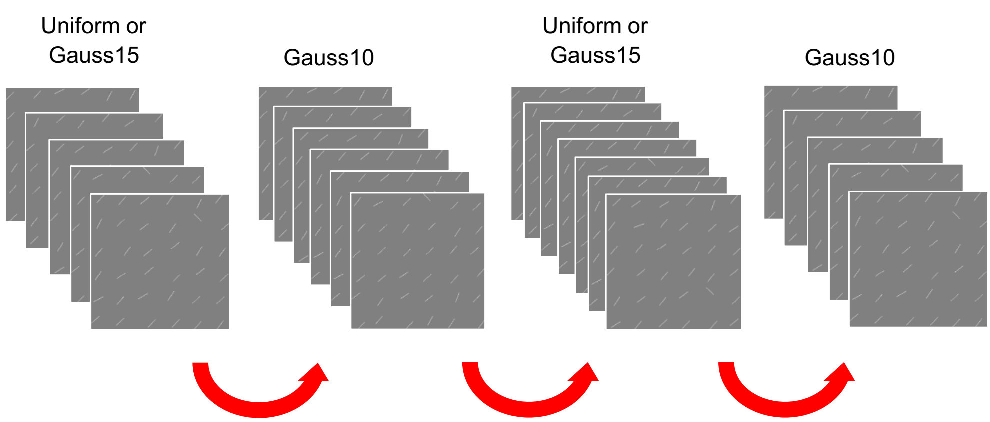

Fig. 11. Streak sequence in Experiment 2.

Table 2
Response times and accuracy as a function of the distractor distributions in Experiment 2.

Distractors	Accuracy (%)		RT (ms, correct responses)	
	M	SD	M	SD
Gaussian, DSD = 10	0.95	0.02	621	73
Gaussian, DSD = 15	0.92	0.03	725	107
Uniform	0.89	0.05	764	118

$F(2, 18) = 23.67$, $p < 0.001$, $\eta^2_G = 0.362$. A comparison between the two larger-range distributions also showed a significant effect on RTs, $t(9) = 3.16$, $p = 0.012$, and accuracy, $t(9) = 4.08$, $p = 0.003$.

As in Experiment 1, the more different the target was from distractors, the easier the search (Fig. 12). A quadratic effect of target distance was found for Gauss10, $B = 1.32$ (0.27), $t = 4.96$, Gauss15, $B = 3.49$ (0.35), $t = 10.10$, and Uniform distributions, $B = 1.64$ (0.38), $t = 4.34$. The effect of target distance was more pronounced for the Gauss15 distribution, $B = 4.74$ (0.79), $t = 5.98$, than for Gauss10, but the difference between Gauss10 and Uniform distributions was not significant, $B = 1.39$ (0.80), $t = 1.73$.

6.3.2. Repetition effects

As in Experiment 1, RTs decreased for the first repetition reaching a plateau approximately at the second trial. Accuracy increased following the first trial in a streak (Fig. 13). Linear mixed-effects regression with Helmert contrasts confirmed these observations indicating that the first and second trial in each streak differed from the later trials for Gauss10 ($t < 2$ for the trials after the second one), but for Gauss15 and Uniform only the first trial differed from the rest. For accuracy, the first trial for Gauss10 and Gauss15 differed from the rest, and for Uniform both the first and the second trials were less accurate than the others.

6.3.3. Distribution shifts

Fig. 14 shows effects of distribution shifts on RTs and accuracy of the Gauss10 distribution. Both RTs and accuracy were the same whether the previous distribution was Uniform or Gaussian (all $ts < 2$).

6.3.4. Role-reversals

The PDF of the uniform distribution is different from the PDF of the Gaussian in that it consists of two different parts – a uniform probability within the distribution range and zero probability elsewhere. Consequently, one would expect the response time function

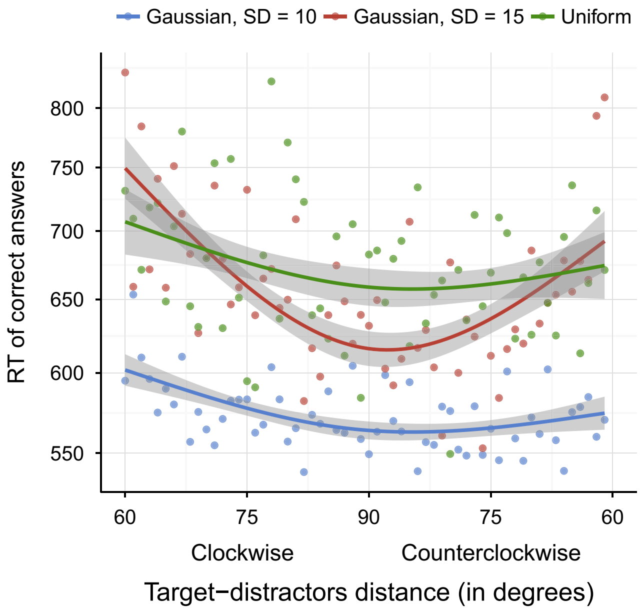

Fig. 12. Quadratic trends of target-distractors distance effect in Experiment 2. Dots show mean RTs, shaded area shows 95% confidence interval based on quadratic regression fit.

to have two parts as well. To quantify the differences, we used a segmented regression. The results showed that following a Uniform distribution, RTs could be approximated by two separate regressions, split at 32.00 deg of previous distractor-to-target distance (Fig. 15). The slope of the first part did not significantly differ from zero, $B = 0.79$, 95% $CI = [-2.90, 4.48]$ (values represent slope and CI for untransformed RTs, log-transformed data yielded the same results). For the second part, the slope was significantly negative, $B = -3.90$, 95% $CI = [-5.41, -2.40]$. Davies' test confirmed that the difference in slopes for the two parts was significant, $p = 0.009$. Following the Gauss15 distribution, there was no significant breakpoint; the RTs decreased monotonically with increasing target-to-previous-distractor-mean distance (Fig. 15).

6.4. Discussion

Experiment 2 demonstrates that not only are observers sensitive to the variance of the distribution, but also to its shape. If distractors were uniformly distributed during the preceding streak, RTs were similar when the target was within the range of this distribution and decreased when the target did not overlap with the distribution. If, on the other hand, distractors were drawn from a Gaussian distribution, RTs decreased monotonically the further the target was oriented away from the mean of the previous distractor distribution.

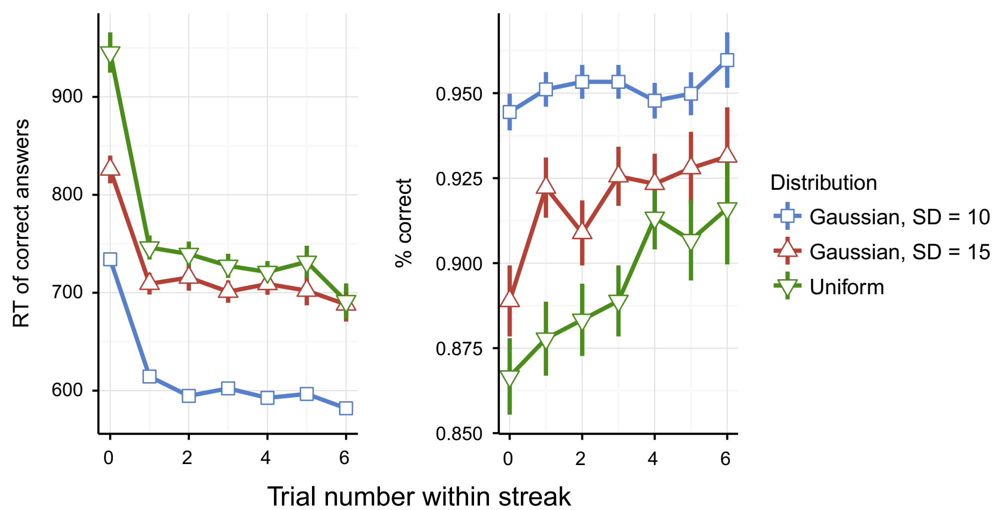

Fig. 13. Repetition effects within streak in Experiment 2. Bars show ±1 SEM.

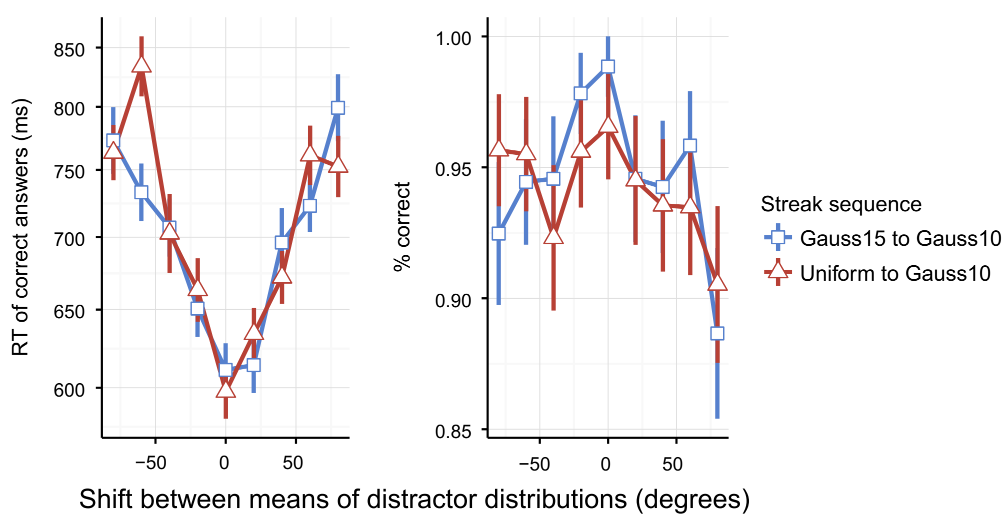

Fig. 14. Response times and accuracy as function of changes in distractor distributions between streaks in Experiment 2. Bars show ±1 SEM.

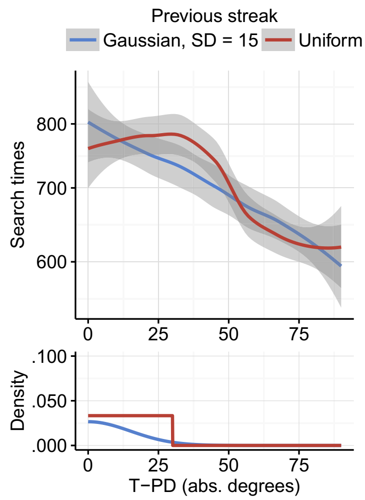

Fig. 15. Top panel: Response times as function of the distance between targets and previous distractors' mean in Experiment 2. Shaded areas show 95% confidence intervals based on local regression fit. Bottom panel: PDFs of the previous distractors distribution.

The data from Experiment 2 confirm the findings from Experiment 1: mean and variance of the distribution are co-encoded. More importantly, however, the results show that observers do not use the normal approximation (Rosenholtz, 2001) when the distractor distribution is not normal. Instead, the shape of the distribution is taken into account. To our best knowledge, this is the first demonstration of the encoding of distribution shape over the course of several trials in visual search.

7. Experiments 3A–3C

The findings of Experiment 2 clearly indicate that observers encode the shape of the distractors distribution. Given the novelty of this findings we replicated this experiment with additional controls for possible confounds.

7.1. Method

Experiments 3A–3C followed the same procedure as Experiment 2. Experiment 3A was a direct replication. In Experiment 3B, in order to directly assess the role of the shape vs. the role of variance of previous distractor distributions, instead of matching the range of uniform and Gaussian ($SD = 15$) distributions, we matched their SD. Due to the fact that the two parameters are connected (for the uniform distribution $\mathrm{SD} = \sqrt{\frac{1}{12}\mathrm{range}^2}$), in order to have the same SD = 15 we had to decrease the range of the uniform distribution to 52 deg. In Experiment 3C, in order to hinder learning of target's parameters so to isolate the effect of distractor distributions, the target was chosen randomly on each trial instead of having constant orientation in each streak. Furthermore, in Experiments 3B and 3C, streaks of trials with a Gaussian distribution and $SD = 10$ had only 1 or 2 trials because these streaks were used only for testing the effects from previous distribution. All observers were trained before the experiments on a singleton search task for 500 trials.

7.2. Participants

Eleven observers (four female, age $M = 35.67$) participated in Experiments 3A–3B. All of them were staff or students at the School of Health Sciences, University of Iceland. Two of them were students who participated as part of a course requirement while the rest participated without additional reward. The response times for one observer in Experiment 3A were too slow ($M = 1612$ ms, as opposed to $M = 1054$ ms for the rest of observers) and the data from this observer in Experiment 3A were replaced with the data from a new observer.

7.3. Results

7.3.1. Average performance

Table 3 shows that the uniform distribution was more difficult for observers than the Gaussian distribution with $SD = 15$ in Experiments 3A ($t(10.0) = -5.39$, $p < 0.001$ and $t(10.0) = 2.25$, $p = 0.048$ for RT and accuracy, respectively) and 3C ($t(10.0) = -7.52$, $p < 0.001$ and $t(10.0) = 1.95$, $p = 0.080$), but in 3B the effect was the opposite ($t(10.0) = 3.94$, $p = 0.003$ and $t(10.0) = -7.87$, $p < 0.001$). In 3B, the standard deviation (SD) but not range was controlled, and as a consequence the range of the uniform distribution was smaller. These results indicate that both range and SD are important for search efficiency. The effects of target-distractor distance, repetition effects, and the effects of distribution shifts were the same as in Experiment 2 and are not reported here for the sake of brevity.

7.3.2. Role-reversals

Role-reversals were analysed in the same way as in Experiment 2 (Fig. 16). For Experiment 3A, after a switch from a uniform distribution, a segmented regression showed a breakpoint at 24.97 deg (Davies' test $p = 0.003$), with the flat segment ($B = 5.47$, 95% $CI = [-5.17, 16.12]$) followed by a segment with a negative slope ($B = -6.38$, 95% $CI = [-8.94, -3.82]$). Following the Gaussian distribution, however, the breakpoint was not significant ($p = 0.732$). Similarly, in Experiment 3B the breakpoint was significant following the uniform distribution (20.04 deg, $p = 0.019$) with a flat first segment and a negative-slope in the second segment ($B = 2.50$, 95% $CI = [-5.90, 10.90]$ and $B = -3.40$, 95% $CI = [-4.78, -2.03]$). Again, following the Gaussian distribution, there was no significant breakpoint ($p = 0.358$). Finally, in Experiment 3C, again following the uniform distribution the breakpoint was significant (45.93 deg, $p = 0.010$), and the first segment was flat while the second had a negative slope ($B = 0.14$, 95% $CI = [-1.96, 2.24]$ and $B = -4.84$, 95% $CI = [-7.19, -2.50]$). Unlike previous experiments, however, following the Gaussian distribution in Experiment 3C there was a significant breakpoint, but it was located at

Table 3
Response times and accuracy as a function of the distractor distributions in Experiments 3A–3C.

Distractors	Exp. 3A: Replication				Exp. 3B: Equal SD				Exp. 3C: Random Target			
	Acc.		RT_{corr}		Acc.		RT_{corr}		Acc.		RT_{corr}	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Gaussian, *SD* = 10	0.96	0.03	803	179	0.96	0.03	822	189	0.97	0.02	775	188
Gaussian, *SD* = 15	0.93	0.05	1022	270	0.91	0.04	960	260	0.94	0.03	877	228
Uniform	0.92	0.05	1172	366	0.94	0.04	897	219	0.93	0.04	975	263

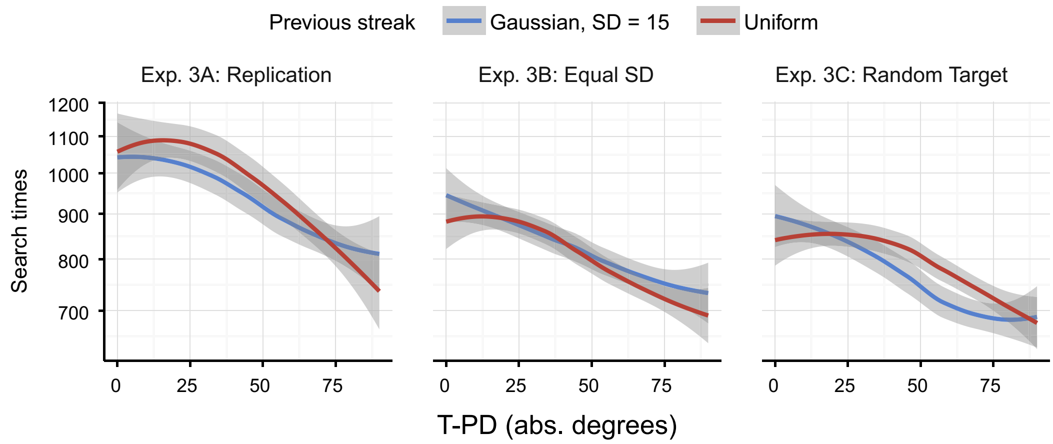

Fig. 16. Response times as function of the distance between targets and previous distractors' mean in Experiments 3A–3C. Shaded areas show 95% confidence intervals based on local regression fit.

68.00 deg (p = 0.019). The segment before this breakpoint had negative slope ($B = -3.59$, 95% $CI = [-4.66, -2.52]$) while the second segment had a flat slope (B = 0.41, 95% $CI = [-4.99, 5.81]$). In contrast to the uniform distribution where breakpoints indicate that observers respond similarly slowly within the range of previous distractors distribution, this breakpoint indicates that observers responded similarly fast when the target was far from the range of the previous distractors distribution.

7.4. Discussion

Experiments 3A–3C replicate the results obtained in Experiment 2 and further demonstrate that they cannot be explained by differences in standard deviations of the Gaussian and uniform distributions or by the learning of target parameters. Interestingly, Experiment 3C where target was selected randomly on each trial showed a presence of a breakpoint following a Gaussian distribution. But, in contrast to uniform distribution, this breakpoint indicated that the segment outside the range of Gaussian distribution had a flat slope. Given that the probability of observing a distractor in this part is zero, a flat slope is to be expected. Yet, a flat slope was only found in this experiment. It is possible that this is a result of the randomly changing target as it might force observers to analyze perceptual space with more scrutiny. However, it also could be a result of an extended training because observers participated in this experiment after completing Experiments 3A and 3B in addition to a training session. Nevertheless, we believe that this result is interesting enough to be investigated in future research.

8. Experiment 4

In previous experiments we used different variants of symmetric distributions. But, according to our proposal that observers can learn distractor set distributions, observers should be sensitive to asymmetries in distribution density functions as well. To test this, in Experiment 4 we analysed the effects of learning skewed distributions.

8.1. Method

Experiment 4 followed the same procedure as Experiment 3B. But, instead of a Gaussian distribution with SD = 15 and uniform distribution we used two variants of triangular distributions. Both had a range of -30 to 30 deg relative to the distribution center which varied in the same way as distribution mean in previous studies. The "Triangular, left" distribution was skewed to the left with a peak at 25 deg, resulting in a longer right tail. The "Triangular, right" distribution was skewed to the right with a peak at -25 deg, resulting in a longer left tail.

8.2. Participants

The same observers participated as in experiments 3A–3C. One observer had to leave before finishing the experiment, thus the resulting analyses were run on the data from 10 observers (4 female, age M = 33.40).

8.3. Results

8.3.1. Average performance

Table 4 shows that both variants of the triangular distribution were slightly more difficult than the Gaussian distribution with an SD = 10. A repeated measures ANOVA confirmed this showing a significant effect of distribution type both in RT, $F(2,18) = 3.81$, $p = 0.045$, $\eta^2_G = 0.00$, and in accuracy, $F(2,18) = 7.42$, $p = 0.008$, $\eta^2_G = 0.03$.

8.3.2. Role-reversals

To test for the effects of distribution skewness we first estimated the breakpoint of RT as function of T-PD distance. In contrast to previous experiments, we analysed the signed distance and not the distance in absolute degrees because the distribution of distractors was not symmetric (Fig. 17). The breakpoint therefore simply corresponds to the maximum RT as function of T-PD. We then tested for the difference in slopes to the left and to the right of the breaking point using simple linear regression with interaction between side (left or right) and distance to breaking point. The analysis indicated that for triangular distribution with longer right tail, the breaking point was at -16.15 deg with a steeper slope on the left B = 3.76, 95% CI = [2.16, 5.35] than on the right

Table 4
Response times and accuracy as a function of the distractor distributions in Experiment 4.

Distractors	Accuracy (%)		RT_{corr} (ms)	
	M	*SD*	*M*	*SD*
Gaussian, SD = 10	0.96	0.02	724	195
Triangular, left	0.95	0.02	752	202
Triangular, right	0.95	0.03	758	213

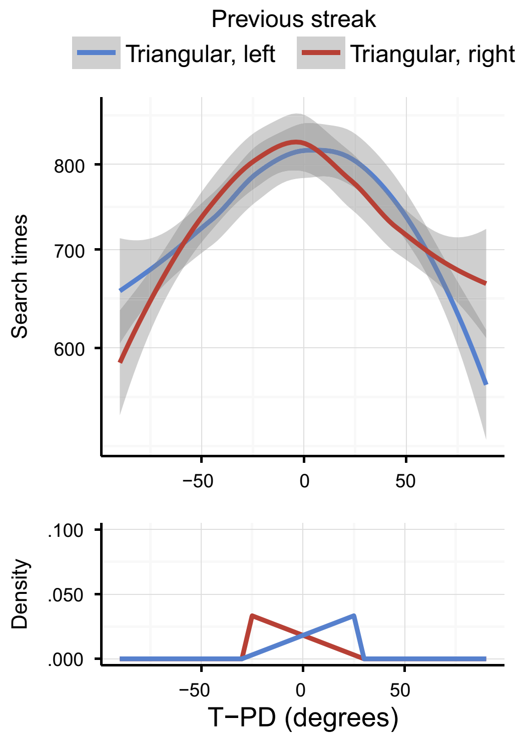

Fig. 17. Top panel: Response times as function of the distance between targets and previous distractors' mean in Experiment 4. Shaded areas show 95% confidence intervals based on local regression fit. Bottom panel: PDFs of the previous distractors distribution.

$B = 2.01$, 95% $CI = [2.96, 1.07]$, $t = -2.68$, $p = 0.008$. For the distribution with longer left tail, in contrast, the breaking point was at 19.72 deg with shallower slope on the left $B = 1.80$, 95% $CI = [0.97, 2.63]$ than on the right $B = 4.07$, 95% $CI = [5.75, 2.39]$, $t = 2.79$, $p = 0.005$.

8.4. Discussion

The results of Experiment 4 are straightforward: with asymmetric distributions of distractors the response times on the first trial in the next streak are also asymmetric. Following a distribution with a left skew, response times as function of the distance between target and previous distractors distribution have a steeper slope on the left than on the right. The opposite happens after a right-skewed distribution. Interestingly, the peak of the RT as function of T-PD (as estimated with breaking point from segmented regression) is closer to the distribution center than the actual peak of the triangular distribution. It also does not correspond to the means or medians of the distributions we used (the means are ±8.3, the medians are ±10.55). Speculatively, observers might use a weighting scheme that assigns higher weights to a particularly important points, such as borders of the distribution.

9. General discussion

Our results demonstrate two novel and important points. Firstly, that priming of pop-out effects can occur based on visual statistics rather than on specific features or their conjunctions. Corroborating the results obtained by Corbett and Melcher (2014) and Michael et al. (2014) we found that observers are sensitive to the mean and standard deviation of the previous distractor distribution. In contrast to previous studies, we demonstrate that learning of mean and SD is an integrated process as shown by different effects of SD depending on the distance from target to the mean of preceding distractors in feature space. Speculatively, observers use the mean and SD to describe a part of feature space as related to distractors. Secondly, we demonstrate that observers can learn information about the shape of distractor distributions. In particular, we show that following a change in distractor distribution:

(1) RTs decrease monotonically with increased distance between target and the mean of the preceding distractor distribution (current target distance to previous distractor), indicating that the mean is learned;
(2) larger standard deviations of preceding distractor distributions result in slower responses, indicating that information about variance is encoded;
(3) steeper Gaussian probability density functions (PDF's) result in a steeper decrease of RTs as a function of previous distractor to target distance;
(4) a uniform distribution results in a two-piece function corresponding to its' PDF;
(5) a triangular distribution results in an asymmetrical distribution of response times again corresponding to its PDF.

In general, RTs follow the shape of the preceding distractor distribution, indicating that distribution shape is encoded along with the mean and standard deviation. This occurs for Gaussian, uniform and skewed triangular distributions. This last result is the most important and novel one. Learning of distribution shapes means that observers build a probabilistic model representing an ensemble of distractor features. That is, over several trials they are able to gather and integrate information from separate distractors into a summary representation. It is unlikely that such information can be gathered from a single trial, since lines were picked randomly on each trial so they do not accurately represent the PDF of the distribution they were drawn from. For example, picking 35 lines from a uniform distribution with a range of 60 deg will provide on average about 27 uniquely oriented lines that cover less than half of the distribution. However, we did not test the effect of learning duration explicitly and it is possible that learning satiates after two or three trials. Future studies may help to clarify the satiation levels for statistical learning in visual search.

Our results suggest that the best-normal approximation suggested by Rosenholtz (2001) is not used when observers encounter non-normal distributions. Instead, the shape of the distribution is taken into account. Two distinct mechanisms can explain the observed effects. First, it is possible that observers approximate the distractor distribution using several overlapping Gaussians resembling a density estimation with Gaussian kernel function. A number of different approximation methods can be used as well, so the choice of particular method for theoretical models depends on their biological plausibility and requires further study. Second, it is possible that higher-order summary statistics, such as skewness or kurtosis, are encoded. Previous studies have not shown that observers use such statistics (Atchley & Andersen, 1995; Dakin & Watt, 1997). However, we used repetitive displays, which provided observers with different examples of distractor sets drawn from the same distribution. Observers therefore had much more information to estimate higher-order statistics. Describing the distribution with several statistics can be more efficient than using a density approximation as it will allow fast comparisons between distributions and outlier detection (Alvarez, 2011; Utochkin, 2015). This mechanism entails that observers ignore local deviations, which cannot be described by the statistics that define the distribution. With sufficient learning the former mechanism (approximation with multiple Gaussians or other functions) should adequately describe such local deviations. The latter mechanism (approximation with a set of statistics) will ignore them regardless of the amount of learning. Future studies are necessary to understand the specific mechanism responsible for the learning of distribution shape.

The results pose an important question for the literature on attentional priming: are homogenous distractors typically used in priming studies approximated as a special kind of distribution? For example, RTs may change as a function of previous distractor to target distance in feature space as observed here. In other words, a target close but not equal in feature space to previous distractors may have detrimental effects on search efficacy. Given that we found such effects using distractors from a Gaussian distribution with DSD = 5, which is nearly homogenous, this is quite probable. This would mean that even when it would be optimal to ignore or focus on only a very small part of perceptual space, observers nevertheless apply a broad filter. A switch from a "categorical" view where distractor groups are treated as isolated entities towards a "continuous" view where they are considered parts of distributions will bring new answers and new challenges to the field. A potential limitation is that we used only one particular type of search – for the odd orientation. We believe that future studies will clarify the generalizability of the findings. Particularly interesting is which stages of processing are affected by repetition of summary statistics.

Finally we note that from a methodological perspective, our study shows that priming of pop-out can be used to assess the development of perceptual representations. This is especially important for studies of prediction-related frameworks. According to hierarchical predictive accounts of cognition, mental representations are probabilistic models tested for external validity and iteratively corrected through Bayesian updating (Arnal & Giraud, 2012; Clark, 2013; den Ouden, Friston, Daw, McIntosh, & Stephan, 2009; den Ouden, Kok, & de Lange, 2012; Friston, 2009, 2010; Hohwy, 2012, 2014). We propose that the language of summary statistics naturally suits predictive coding. Information about distribution parameters allows the use of distribution probability density functions. This enables prediction of both the probabilities of appearance of existing stimuli and of novel stimuli within the same feature space. For example, if we conceive of the brain as a Bayesian observer, it is vital to assess the models used by observers and how they are updated with new information. The paradigm we have introduced provides a convenient way of achieving this.

Summing up, our study demonstrates that visual statistics accumulate over and above simple parameters such as means and standard deviations: the shapes of distributions are learned as well. Note that means and standard deviations are most useful only when the information comes from a Gaussian distribution. However, distributions of line orientations in the real world, are far from normal (Coppola, Purves, McCoy, & Purves, 1998). Moreover, what is important to us changes, and the distributions of signals that should be attended and of those to ignore change consequently, highlighting the need to learn probability functions other than Gaussian approximations. Observers need to learn more than means and SDs to adapt to the environment, and as we show they can build ensemble representations with surprising precision.

Supplementary data

The data from the experiments reported in this paper is available at https://osf.io/3wcth/.

Acknowledgements

This paper was supported by Russian Foundation for Humanities (#15-36-01358A2), Saint Petersburg State University (research grant #8.38.287.2014), and a grant from the Icelandic Research Fund (IRF #152427).

References

Albrecht, A. R., & Scholl, B. J. (2010). Perceptually averaging in a continuous visual world: Extracting statistical summary representations over time. *Psychological Science, 21*(4), 560–567. http://dx.doi.org/10.1177/0956797610363543.

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences, 15*(3), 122–131. http://dx.doi.org/10.1016/j.tics.2011.01.003.

Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science, 19*(4), 392–398. http://dx.doi.org/10.1111/j.1467-9280.2008.02098.x.

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science, 12*(2), 157–162.

Arnal, L. H., & Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences, 16*(7), 390–398. http://dx.doi.org/10.1016/j.tics.2012.05.003.

Ásgeirsson, Á. G., & Kristjánsson, Á. (2011). Episodic retrieval and feature facilitation in intertrial priming of visual search. *Attention, Perception & Psychophysics, 73*(5), 1350–1360. http://dx.doi.org/10.3758/s13414-011-0119-5.

Atchley, P., & Andersen, G. J. (1995). Discrimination of speed distributions: Sensitivity to statistical properties. *Vision Research, 35*(22), 3131–3144. http://dx.doi.org/10.1016/0042-6989(95)00057-7.

Avraham, T., Yeshurun, Y., & Lindenbaum, M. (2008). Predicting visual search performance by quantifying stimuli similarities. *Journal of Vision, 8*(9), 1–22. http://dx.doi.org/10.1167/8.4.9.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. http://dx.doi.org/10.1016/j.jml.2007.12.005.

Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision, 9*(12). http://dx.doi.org/10.1167/9.12.13. 13.1–18.

Bates, D. M. (2006). [R] lmer, *p*-values and all thathttps://stat.ethz.ch/pipermail/r-help/2006-May/094765.html. Retrieved 18.01.15.

Becker, S. I. (2008). The mechanism of priming: Episodic retrieval or priming of pop-out? *Acta Psychologica, 127*, 324–339. http://dx.doi.org/10.1016/j.actpsy.2007.07.005.

Becker, S. I. (2010). The role of target-distractor relationships in guiding attention and the eyes in visual search. *Journal of Experimental Psychology: General, 139*(2), 247–265. http://dx.doi.org/10.1037/a0018808.

Campana, G., Pavan, A., & Casco, C. (2008). Priming of first- and second-order motion: Mechanisms and neural substrates. *Neuropsychologia, 46*(2), 393–398. http://dx.doi.org/10.1016/j.neuropsychologia.2007.07.019.

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research, 43*(4), 393–404. http://dx.doi.org/10.1016/S0042-6989(02)00596-5.

Chong, S. C., & Treisman, A. (2005). Statistical processing: Computing the average size in perceptual groups. *Vision Research, 45*(7), 891–900. http://dx.doi.org/10.1016/j.visres.2004.10.004.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *The Behavioral and Brain Sciences, 36*(3), 181–204. http://dx.doi.org/10.1017/S0140525X12000477.

Coppola, D. M., Purves, H. R., McCoy, a. N., & Purves, D. (1998). The distribution of oriented contours in the real world. *Proceedings of the National Academy of Sciences of the United States of America, 95*(7), 4002–4006. http://dx.doi.org/10.1073/pnas.95.7.4002.

Corbett, J. E., & Melcher, D. (2014). Stable statistical representations facilitate visual search. *Journal of Experimental Psychology: Human Perception and Performance, 40*(5), 1915–1925. http://dx.doi.org/10.1037/a0037375.

Dakin, S. C. (2015). Seeing statistical regularities: Texture and pattern perception. In J. Wagemans (Ed.), *The oxford handbook of perceptual organization* (pp. 150–167).

Dakin, S. C., & Watt, R. J. (1997). The computation of orientation statistics from visual texture. *Vision Research, 37*(22), 3181–3192. http://dx.doi.org/10.1016/S0042-6989(97)00133-8.

Danielmeier, C., & Ullsperger, M. (2011). Post-error adjustments. *Frontiers in Psychology, 2*(September), 233. http://dx.doi.org/10.3389/fpsyg.2011.00233.

de Gardelle, V., & Summerfield, C. (2011). Robust averaging during perceptual judgment. *Proceedings of the National Academy of Sciences, 108*(32), 13341–13346. http://dx.doi.org/10.1073/pnas.1104517108.

den Ouden, H. E. M., Friston, K. J., Daw, N. D., McIntosh, A. R., & Stephan, K. E. (2009). A dual role for prediction error in associative learning. *Cerebral Cortex (New York, N.Y. : 1991), 19*(5), 1175–1185. http://dx.doi.org/10.1093/cercor/bhn161.

den Ouden, H. E. M., Kok, P., & de Lange, F. P. (2012). How prediction errors shape perception, attention, and motivation. *Frontiers in Psychology, 3*(December), 548. http://dx.doi.org/10.3389/fpsyg.2012.00548.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96*(3), 433–458.

Friston, K. J. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences, 13*(7), 293–301. http://dx.doi.org/10.1016/j.tics.2009.04.005.

Friston, K. J. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127–138. http://dx.doi.org/10.1038/nrn2787.

Haberman, J., & Whitney, D. (2012). Ensemble perception: Summarizing the scene and broadening the limits of visual processing. In J. Wolfe & L. Robertson (Eds.), *From perception to consciousness: Searching with Anne Treisman* (pp. 339–349). Oxford: Oxford University Press.

Hohwy, J. (2012). Attention and conscious perception in the hypothesis testing brain. *Frontiers in Psychology, 3*(April). http://dx.doi.org/10.3389/fpsyg. 2012.00096.

Hohwy, J. (2014). The self-evidencing brain. *Noûs*. http://dx.doi.org/10.1111/nous.12062.
Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language, 59*(4), 434–446. http://dx.doi.org/10.1016/j.jml.2007.11.007.
Kristjánsson, Á. (2006). Simultaneous priming along multiple feature dimensions in a visual search task. *Vision Research, 46*(16), 2554–2570. http://dx.doi.org/10.1016/j.visres.2006.01.015.
Kristjánsson, Á., & Campana, G. (2010). Where perception meets memory: A review of repetition priming in visual search tasks. *Attention, Perception & Psychophysics, 72*(1), 5–18. http://dx.doi.org/10.3758/APP.72.1.5.
Kristjánsson, Á., & Driver, J. (2008). Priming in visual search: Separating the effects of target repetition, distractor repetition and role-reversal. *Vision Research, 48*(10), 1217–1232. http://dx.doi.org/10.1016/j.visres.2008.02.007.
Kristjánsson, Á., Saevarsson, S., & Driver, J. (2013). The boundary conditions of priming of visual search: From passive viewing through task-relevant working memory load. *Psychonomic Bulletin & Review, 20*(3), 514–521. http://dx.doi.org/10.3758/s13423-013-0375-6.
Lamy, D. F., Antebi, C., Aviani, N., & Carmel, T. (2008). Priming of pop-out provides reliable measures of target activation and distractor inhibition in selective attention. *Vision Research, 48*(1), 30–41. http://dx.doi.org/10.1016/j.visres.2007.10.009.
Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: I. Role of features. *Memory & Cognition, 22*(6), 657–672.
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), 1–8. http://dx.doi.org/10.1167/13.5.8.
Meeter, M., & Olivers, C. N. L. (2006). Intertrial priming stemming from ambiguity: A new account of priming in visual search. *Visual Cognition, 13*(2), 202–222. http://dx.doi.org/10.1080/13506280500277488.
Meeter, M., & Olivers, C. N. L. (2014). Target features and target-distractor relation are both primed in visual search. *Attention, Perception & Psychophysics, 76*(3), 682–694. http://dx.doi.org/10.3758/s13414-013-0611-1.
Michael, E., de Gardelle, V., & Summerfield, C. (2014). Priming by the variability of visual information. *PNAS, 111*(21), 7873–7878. http://dx.doi.org/10.1073/pnas.1308674111.
Morgan, M. J., Chubb, C., & Solomon, J. A. (2008). A "dipper" function for texture discrimination based on orientation variance. *Journal of Vision, 8*(2008), 1–8. http://dx.doi.org/10.1167/8.11.9.Introduction.
Muggeo, V. M. R. (2003). Estimating regression models with unknown break-points. *Statistics in Medicine, 22*, 3055–3071.
Muggeo, V. M. R. (2008). Segmented: An R package to fit regression models with broken-line relationships. *R News, 8*(1), 20–25.
Muller, H. J., Reimann, B., & Krummenacher, J. (2003). Visual search for singleton feature targets across dimensions: Stimulus- and expectancy-driven effects in dimensional weighting. *Journal of Experimental Psychology: Human Perception and Performance, 29*(5), 1021–1035. http://dx.doi.org/10.1037/0096-1523.29.5.1021.
Nagy, A. L., Neriani, K. E., & Young, T. L. (2005). Effects of target and distractor heterogeneity on search for a color target. *Vision Research, 45*, 1885–1899. http://dx.doi.org/10.1016/j.visres.2005.01.007.
Norman, L. J., Heywood, C. A., & Kentridge, R. W. (2015). Direct encoding of orientation variance in the visual system. *Journal of Vision, 15*, 1–14. http://dx.doi.org/10.1167/15.4.3.
Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. J. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience, 4*(7), 739–744. http://dx.doi.org/10.1038/89532.
Peirce, J. W. (2007). PsychoPy – Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1–2), 8–13. http://dx.doi.org/10.1016/j.jneumeth.2006.11.017.
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*(4), 985–999. http://dx.doi.org/10.1037//0096-1523.27.4.985.
Rosenholtz, R., Huang, J., Raj, A., Balas, B. J., & Ilie, L. (2012). A summary statistic representation in peripheral vision explains visual search. *Journal of Vision, 12*(4), 14. http://dx.doi.org/10.1167/12.4.14.
Rosenholtz, R. (2014). Texture perception. In J. Wagemans (Ed.), *Oxford handbook of perceptual organization* (pp. 1–24). Oxford University Press. http://dx.doi.org/10.1093/oxfordhb/9780199686858.013.058.
Utochkin, I. S. (2013). Visual search with negative slopes: The statistical power of numerosity guides attention. *Journal of Vision, 13*(3), 1–14. http://dx.doi.org/10.1167/13.3.18.
Utochkin, I. S. (2015). Ensemble summary statistics as a basis for rapid visual categorization. *Journal of Vision, 15*(4), 8. http://dx.doi.org/10.1167/15.4.8.
Vincent, B. T., Baddeley, R. J., Troscianko, T., & Gilchrist, I. D. (2009). Optimal feature integration in visual search. *Journal of Vision, 9*(5), 1–11. http://dx.doi.org/10.1167/9.5.15.
Wang, D., Kristjánsson, A., & Nakayama, K. (2005). Efficient visual search without top-down or bottom-up guidance. *Perception & Psychophysics, 67*(2), 239–253. http://dx.doi.org/10.3758/BF03206488.