Rapid learning of visual ensembles

Andrey Chetverikov
Laboratory for Visual Perception and Visuomotor Control, Faculty of Psychology, School of Health Sciences, University of Iceland, Reykjavik, Iceland
Cognitive Research Lab, Russian Academy of National Economy and Public Administration, Moscow, Russia
Department of Psychology, Saint Petersburg State University, Saint Petersburg, Russia

Gianluca Campana
Dipartimento di Psicologia Generale, University of Padova, Padova, Italy
Human Inspired Technology Research Centre, University of Padova, Padova, Italy

Árni Kristjánsson
Laboratory for Visual Perception and Visuomotor Control, Faculty of Psychology, School of Health Sciences, University of Iceland, Reykjavik, Iceland

We recently demonstrated that observers are capable of encoding not only summary statistics, such as mean and variance of stimulus ensembles, but also the shape of the ensembles. Here, for the first time, we show the learning dynamics of this process, investigate the possible priors for the distribution shape, and demonstrate that observers are able to learn more complex distributions, such as bimodal ones. We used speeding and slowing of response times between trials (intertrial priming) in visual search for an oddly oriented line to assess internal models of distractor distributions. Experiment 1 demonstrates that two repetitions are sufficient for enabling learning of the shape of uniform distractor distributions. In Experiment 2, we compared Gaussian and uniform distractor distributions, finding that following only two repetitions Gaussian distributions are represented differently than uniform ones. Experiment 3 further showed that when distractor distributions are bimodal (with a 30° distance between two uniform intervals), observers initially treat them as uniform, and only with further repetitions do they begin to treat the distributions as bimodal. In sum, observers do not have strong initial priors for distribution shapes and quickly learn simple ones but have the ability to adjust their representations to more complex feature distributions as information accumulates with further repetitions of the same distractor distribution.

Introduction

The world around us consists of objects: spatially constrained entities with relatively consistent features. It is therefore tempting to assume that information processing is based on discrete objects or features. However, when one looks at the sky, the trees, or the sea, objects tend to disappear and features blend into one another, calling for another perspective on perception that can be called a distribution-based view. From this perspective, single features are less important than their probability distributions because the latter are more stable and less prone to noise than the former.

Observers' ability to make inferences regarding the ensembles of visual stimuli has been studied for a long time (see reviews in Bauer, 2015; Peterson & Beach, 1967; Pollard, 1975), but recently such views have gained renewed popularity through the studies of judgments of *summary statistics*, such as the mean or variance of color, motion direction, size, orientation, etc. (Alvarez, 2011; Alvarez & Oliva, 2008; Ariely, 2001; Chong & Treisman, 2003; Haberman, Brady, & Alvarez, 2015; Haberman & Whitney, 2012; Rosenholtz, 2001; Utochkin, 2015). Even though stimulus sets used in these experiments usually consist of discrete

Citation: Chetverikov, A., Campana, G., & Kristjánsson, Á. (2017). Rapid learning of visual ensembles. *Journal of Vision*, *17*(2):21, 1–15, doi:10.1167/17.2.21.

doi: 10.1167/17.2.21 Received September 29, 2016; published February 28, 2017 ISSN 1534-7362 Copyright 2017 The Authors

objects, observers' ability to extract their aggregate properties suggests that, at some level, these objects are integrated into a unified representation. Moreover, the stability of summary statistics is no less important for efficient processing than the stability of individual features. For example, changes in mean or variance of stimulus features over a sequence of trials can increase response times (RTs) on visual tasks (Corbett & Melcher, 2014; Michael, de Gardelle, & Summerfield, 2014).

However, attempts to show that observers are able to use other information about distributions of features than their mean and standard deviation have been unsuccessful. For example, Dakin and Watt (1997) concluded that observers cannot access the skew of orientation distributions. However, other studies find that observers utilize information about distributions of orientation statistics in natural images to make perceptual judgments (Girshick, Landy, & Simoncelli, 2011; Zhang, Kwon, & Tadin, 2013). So, although observers may not be able to assess complex distributions from single presentations, they could still be able to learn such distributions with repeated exposure. Moreover, Utochkin and Yurevich (2016) have recently demonstrated that two distractor distributions with the same range can lead to different visual search efficiency when they have different shape. However, this does not imply that observers encode the shape of the distribution. Instead, the authors suggest that this reflects Gaussian approximations with a different number of subsets.

Recently, we demonstrated for the first time that not only can people encode summary statistics of stimulus sets, but they can also encode probability density functions corresponding to distribution shape with repeated presentation (Chetverikov, Campana, & Kristjánsson, 2016). We used a novel experimental paradigm based on priming in visual search to compare effects of different distractor distributions on RTs following a change in distribution parameters (Figure 1). Observers searched for an oddly oriented line in a matrix of 36 lines. In each trial, distractor orientations were drawn randomly from a distribution with preset parameters. These parameters were kept constant during streaks of five to seven trials. It is well known that observers use information acquired from previous trials so that when target or distractors repeat, search is fast, but when they change, search is slower and less accurate (Kristjánsson & Campana, 2010; Lamy & Kristjánsson, 2013; Maljkovic & Nakayama, 1994; Meeter & Olivers, 2006). Importantly, different aspects of the task-relevant stimulus properties, such as distractor or target parameters and their relationship, can be primed independently (Kristjánsson, 2006; Kristjánsson & Driver, 2008; Kristjánsson, Ingvarsdóttir, & Teitsdóttir, 2008; Lamy, Antebi, Aviani, & Carmel, 2008; Maljkovic & Nakayama, 1994; Wang, Kristjánsson, & Nakayama, 2005). Intertrial priming reveals implicit expectations: The less a distractor or target with specific features is expected, the slower the RTs if they actually do appear (cf. Gekas, Seitz, & Seriès, 2015). The expectations, in turn, reveal the representations of the stimuli presented during the trials.

We utilized this priming effect and measured changes in RTs when the parameters of distractor distributions changed and when targets had features previously pertaining to the distractor distribution. This allowed us to infer how the distributions are encoded. As expected, we found that RTs are affected by changes in the mean and variance of distractors. More surprisingly, however, we found that RTs as a function of the distance in feature space between the current target and previous distractor distribution mean (current target – previous distractors, CT – PD) generally follow the shape of the preceding distributions. Following a streak of trials with a Gaussian distractor distribution, RTs monotonically decreased as CT – PD increased. But following trials with a uniform distribution, the RT $\propto$ CT – PD functions had two parts: The first had a flat slope when a current target was within the range of the preceding distribution, and the second part had a negative slope, monotonically decreasing as CT – PD increased. Similarly, we found that skewed distributions resulted in skewed functions of RT dependent on CT – PD. Several alternative explanations, such as priming from changes in target orientation or simple range encoding were rejected. These and other results allowed us to conclude that observers treat the part of feature space previously occupied by distractors differently, depending on the shape of the distractor distributions. In other words, they have an internal representation[1] corresponding to the shape of the distractor distribution.

Although our results demonstrated that observers encode and learn distribution shape, the dynamics of this learning process are still unknown. Studies of the processing of summary statistics show that a single trial suffices for observers to grasp the mean and variance of a distribution (e.g., Ariely, 2001; Utochkin & Tiurina, 2014). Yet psychophysical studies using single exposure have not revealed any encoding of other statistics (Atchley & Andersen, 1995; Dakin & Watt, 1997). In Chetverikov et al. (2016) five to seven trials in each streak were enough to reveal encoding of distribution shape. Here, by using test streaks of different lengths, we assess for the first time the amount of information needed for encoding distribution shape.

A second, tightly related and still unresolved question concerns observers' prior expectations and their role in ensemble perception. From a Bayesian perspective, observers may have priors that differ from

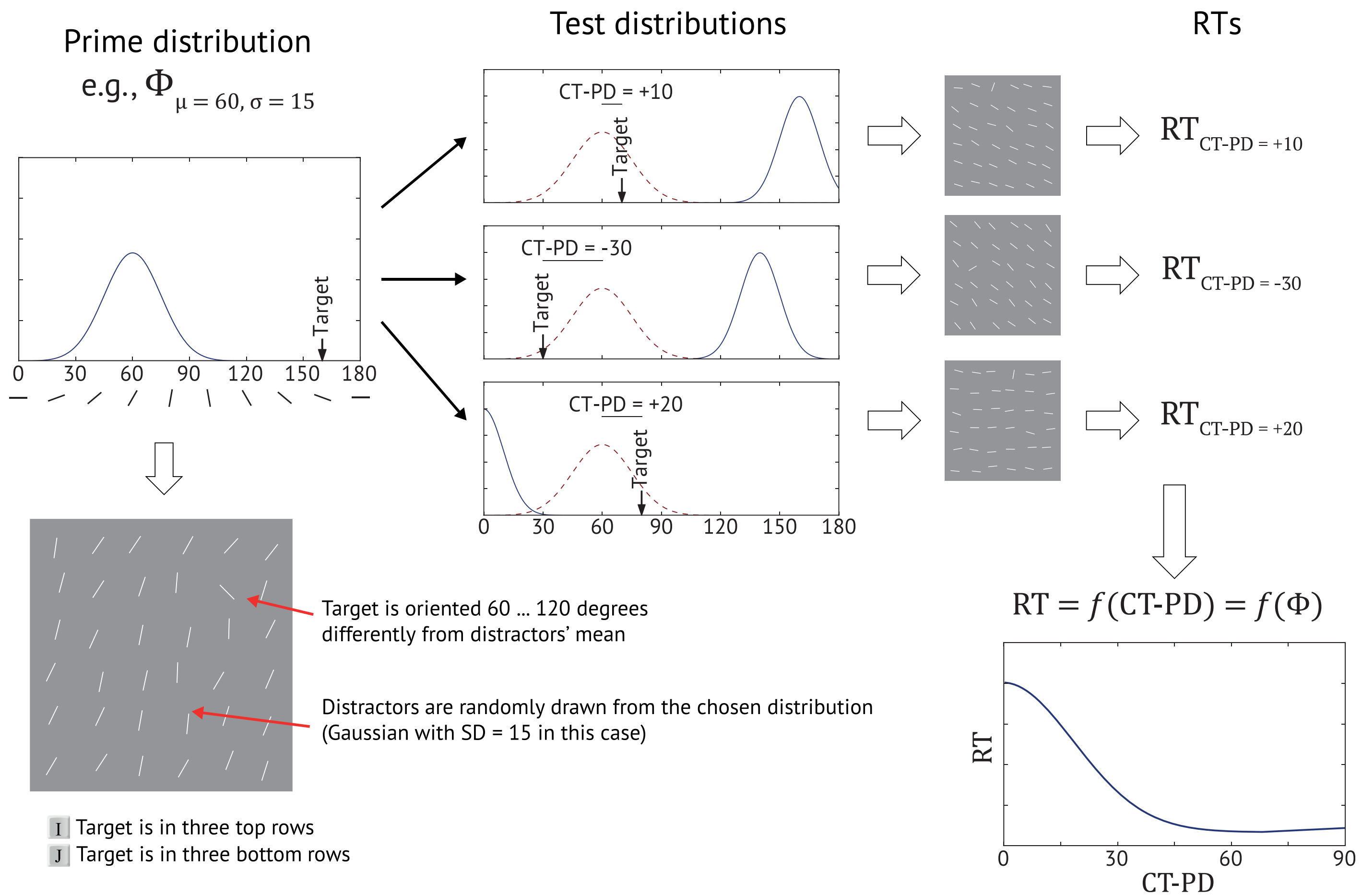

Figure 1. Schematic explanation of the method introduced by Chetverikov et al. (2016) and used in the present experiments. The aim is to estimate the internal representation corresponding to the prime distribution with a given probability density function (upper left panel; in this example, a Gaussian distribution with a mean of 60° and *SD* of 15° shown in blue). One or more displays for odd-one-out search are created from this distribution (lower left panel). The target is oriented 60° to 120° away from the distractor distribution mean and could be the same or change randomly on each trial. These displays are shown one after another (prime distribution streak). The prime distribution streak of consecutive trials from the same distribution is then followed by one or more test trials (middle panel: previous prime distribution is shown with the dashed red line; current distractor distribution is shown with the solid line). On test trials, the key variable is CT − PD, the distance between the mean of the previous distractor distribution and the current target orientation. The distractor distribution (Gaussian, *SD* = 10) in test trials is set randomly so that the difference between its mean and target orientation is the same as during the prime streak, 60° to 120°. RTs are then measured as a function of CT − PD (right panel). The whole procedure is then repeated with different prime distribution means. The resulting RTs are then used to assess observer expectations, revealing the internal representation of the prime distribution (the bottom-right panel; CT − PD for symmetric distributions is shown here and later in absolute degrees).

uniform, which influence distribution encoding. Only when new information is obtained will the priors be gradually updated to reflect the distribution's true shape. One of the candidates for such priors is a Gaussian distribution. Being defined by only two parameters, it could be quite useful for different tasks, such as categorization or scene segmentation (Utochkin, 2015). Moreover, a Gaussian approximation can be useful (although not optimal) for describing effects of distractor heterogeneity in visual search (Rosenholtz, 2001). A model with Gaussian priors predicts that with increasing streak length observers will gradually adjust their perception of the distractor distribution away from a Gaussian one and closer to the observed one.

In Experiment 1, we used a uniform distribution of distractors, varying streak length of search for targets within the same distractor distribution to test the dynamics of distractor distribution processing. We assessed the underlying distributions by measuring changes in RTs as target and distractors change.[2] More specifically, the critical measure was RT as a function of the distance in feature space between the current target and the mean of the previous distractor

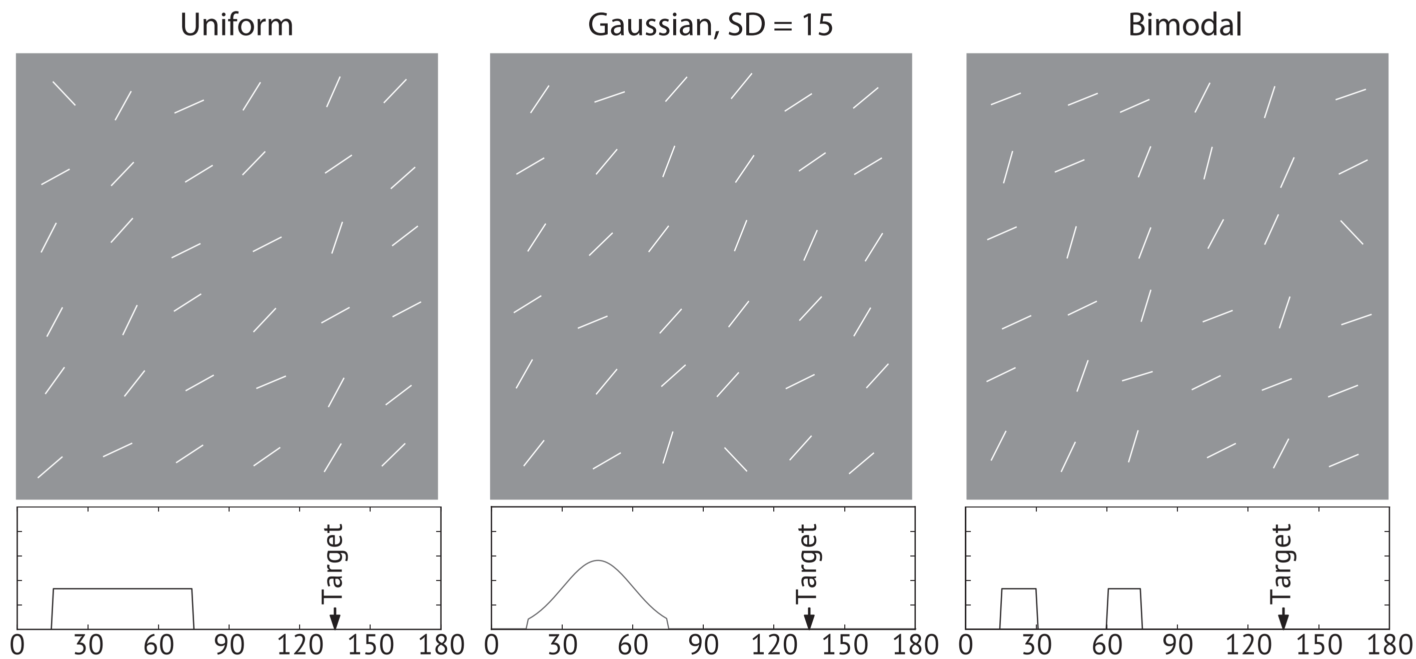

Figure 2. Example prime distributions in Experiments 1, 2, and 3. The top row shows example stimuli, and the bottom row shows corresponding probability density functions of distractor distributions and target position in feature space.

distribution (CT – PD). If there were strong Gaussian priors for the distribution shape, we would expect that with short streaks the representation of the distribution would still reflect these priors—that is, RT as a function of CT – PD would change monotonically with no breaking points. In contrast, if the priors were nonexistent or weak compared to the empirical evidence gained by observers, even short streaks would result in a two-segment function with a flat first segment and a second segment with a negative slope.

Experiment 1

Method

Participants

Ten observers (four female, age $M = 24.80$) at St. Petersburg State University took part in two experimental sessions taking approximately 20–30 min each. This and the following experiments were run in accordance with the Declaration of Helsinki and were approved by the ethics committee of St. Petersburg State University.

Equipment

The stimuli were presented on an iiyama ProLite T2250MTS display (1920 × 1080 pixel resolution) using PsychoPy (Peirce, 2007). Viewing distance was ~57 cm.

Procedure

Observers searched for an oddly oriented line in a 6 × 6 grid of 36 lines subtending 16° × 16° at the center of a display. The length of each line was 1.41°. Line positions were jittered by randomly adding a value ±0.5° to both vertical and horizontal coordinates. The trials were organized in streaks of one to 11 trials of the prime distractor distribution (uniform; orientation range = 60°, Figure 2) with a constant mean. The mean distractor orientation for each prime streak was set randomly. Each prime streak was followed by a test streak with a Gaussian distribution ($SD = 10$, outliers with values outside of the ±20° range were removed; although it is technically a "truncated Gaussian," we refer to it as Gaussian for the sake of brevity) with a streak length of one or two trials (Figure 3). Searching for targets among distractors from such a Gaussian distribution is of intermediate difficulty for observers and was found to be a good candidate for testing the effects of priming in previous studies (Chetverikov et al., 2016). The probability of each streak length within each streak type (prime or test) was the same. Each observer participated in two sessions of 180 prime streaks lasting approximately 20–30 min each (with a total of 2,703 trials per observer on average). For each test streak, the mean of the prime distractor distributions differed from the mean of the test distractor distribution by –80° to +80° in 20° steps. The target in each test and prime streak was set randomly so that target-to-distractor distance ranged from 60° to 120° and was constant within the streak (previous studies

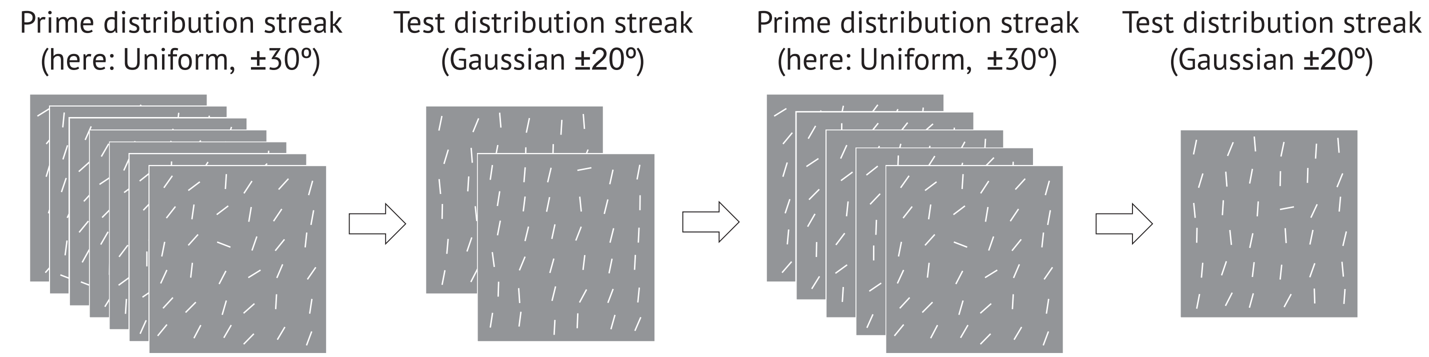

Figure 3. Example sequence of prime and test streaks in Experiment 1. The prime distribution streaks could be one to 11 trials long, and the test distribution streaks were one or two trials long.

have shown that the results are similar regardless of whether target orientation in each trial is randomized; Chetverikov et al., 2016).

In each trial during both prime and test streaks, observers pressed the "i" key if the target was in the upper half of the stimulus (upper three rows) and the "j" key if it was in the lower half of the stimulus (lower three rows). Target position was randomized. If their response was incorrect, the word "ERROR!" appeared in red for 1 s. Observers were encouraged to respond as quickly and accurately as possible and were told that their performance would be scored. The score from the previous trial was shown in the top left corner of the display (green if positive, red if negative, computed as $score = 10 + (1 - RT) \times 10$ for correct answers, where RT is measured in seconds on previous trial, and as $score = - \mid 10 + (1 - RT) \times 10 \mid -10$ for errors) along with the trial number and the total number of trials. Observers received positive scores for fast (faster than 1 s) and accurate responses and negative scores otherwise. This score is purely arbitrary and was used only to motivate observers. The total score was shown during resting periods following 90 prime streaks (half of a session). 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.

Data analysis

We used the same data analysis approach as in Chetverikov et al. (2016). The main focus of the analysis was RTs on test trials as a function of the distance in feature space between the current target and the mean of the distractor distribution used on preceding prime trials (CT − PD). Only correct responses were analyzed. Following the approach taken in the previously reported experiments, we also excluded posterror trials to avoid the effects of posterror slowing. The probability of each distractor in the uniform distribution was equal within the range of the distribution and zero elsewhere. Hence, in test trials similarly slow responses are expected when the target is within the range of the preceding distractor distribution (absolute values of $CT - PD < 30$) whereas a monotonic decrease in RTs is expected with increasing the distance from the range of the preceding distractor distribution.[3] In other words, the resulting $RT \propto CT - PD$ function should also have two components. The presence of two components was tested with segmented regression, also known as piece-wise or broken-line regression, using the *segmented* package in *R* (Muggeo, 2003, 2008). To test whether the two-segment model fits better than a one-segment model, we used the Davies test (provided in the same package) that tests for nonzero difference-in-slope parameters of the segments.

We report the results of linear mixed effects regression (LMER) as regression coefficients (with standard deviations in parentheses) and appropriate Student's t tests in case of simple linear regression and Wald's Z for binomial regression. The *lme4* package in *R* (Bates, Maechler, Bolker, & Walker, 2014) does not report p values for t from mixed effects regression because of indeterminacy surrounding the choice of appropriate degrees of freedom. However, values of t and Z above two can be treated similarly to $p < 0.005$.

Results

Average performance

First, we analyzed search times and accuracy as function of the distribution within prime and test streaks. Search was more difficult for targets among distractors drawn from the uniform distribution than the Gaussian with $SD = 10$ (Table 1), both in RT, $t(9.0) = 3.96$, $p = 0.003$, and accuracy, $t(9.0) = 5.27$, $p < 0.001$.

Repetition effects

We next assessed repetition effects on prime streaks (the test streaks only had one or two trials, analyzed in

	Accuracy		RT_{corr}, ms	
Distractors	*M*	*SD*	*M*	*SD*
Gaussian, $SD = 10$ (test streaks)	0.95	0.02	674	123
Uniform (prime streaks)	0.89	0.05	743	168

Table 1. Search times and accuracy. *Notes*: The uniform distribution had longer streaks, and thus the comparisons between distributions should be made with caution.

the next session). Search times decreased during the first few repetitions reaching a plateau approximately following the second trial (Figure 4). Accuracy also increased after the first trial in a streak. LMER with Helmert contrasts (comparing each trial with the average of the following trials within the streak) confirmed these observations, indicating that only RTs on the first trial were slower, $B = 0.12$ (0.01), $t = 17.07$, and less accurate, $B = -0.34$ (0.06), $Z = 5.88$, than on later trials. Search on the second trial was neither slower, $B = -0.01$ (0.01), $t = -1.78$, nor more accurate, $B = -0.03$ (0.07), $Z = 0.41$, than on later trials.[4]

RTs as function of CT − PD

We analyzed RTs on test streaks as a function of the distance between the current target orientation and distractor mean orientation on preceding prime trials (CT − PD). Replicating the results from our previous studies, we found a nonlinear dependency of RTs on preceding distractor distributions. Specifically, a segmented regression showed that RTs are described by a two-part linear function with a breaking point at 25.8° distance (Figure 5). The slope of the first part did not significantly differ from zero, $B = 0.67$, 95% $CI =$ [–1.36, 2.71] (values represent slope and CI for untransformed search times; the actual analysis was done using log-transformed RT data and yielded the same results), and for the second part, the slope was significantly negative, $B = -1.86$, 95% $CI =$ [–2.37, –1.36]. The Davies test confirmed that the difference in slopes for the two parts was indeed significant, $p =$ 0.001.

The most novel and important question here was how quickly representations of distributions develop. To test whether observers can represent uniform distributions and to analyze effects of streak length, we used LMER including CT − PD and its interaction with previous streak length for segments within (0° to 30° CT − PD) and outside the range of the previous distractor distribution (30° to 90° CT − PD). We also controlled for the effect of distance between the previous target and the mean of the current distractor distribution and the distance between means of previous and current distractor distributions. Within the range of the previous distractor distributions the slope of CT − PD was flat, $B = 0.01$ (0.02), $t = 0.70$, and did not interact with previous streak length, $B = -0.001$ (0.003), $t = 0.48$. Outside the previous distractor distribution range, it was negative, $B = -0.01$ (0.01), $t =$ –2.29, but also did not interact with previous streak length, $B = -0.001$ (0.001), $t = -0.68$.

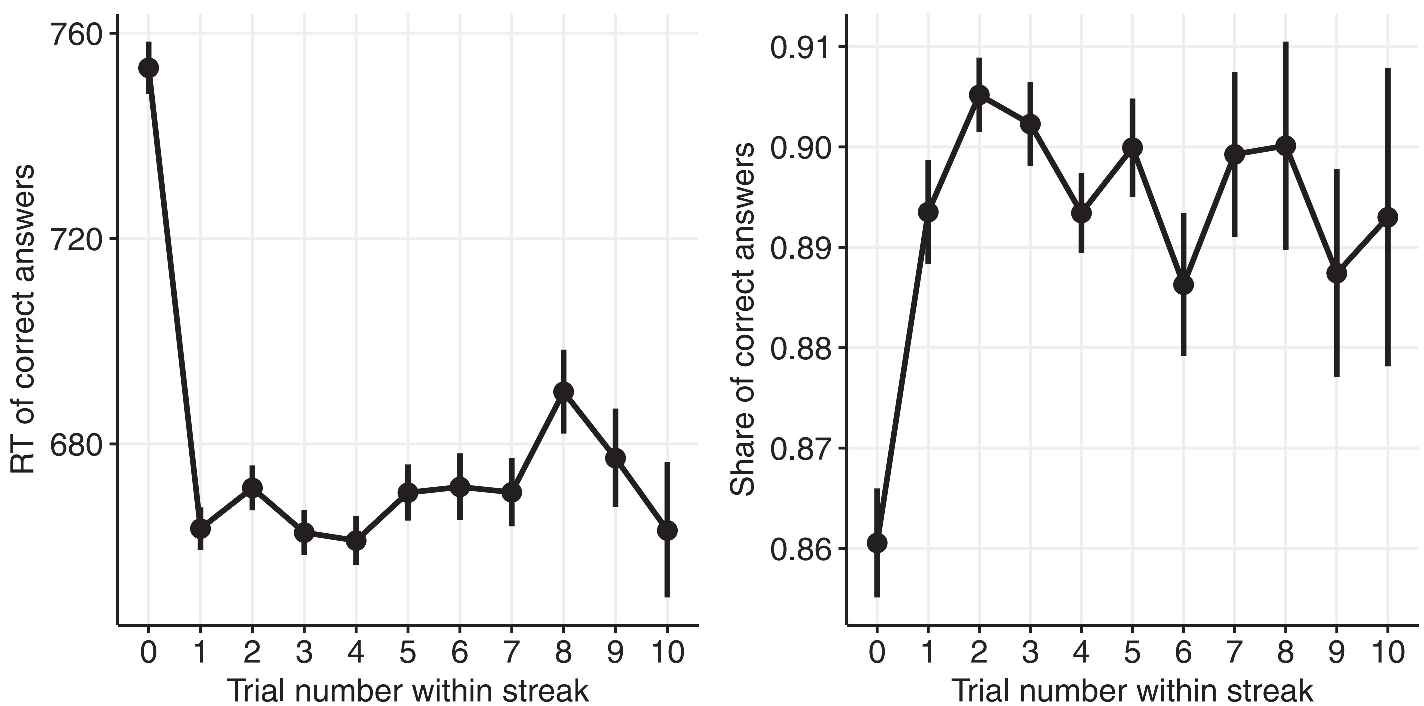

Figure 4. Repetition effects within streaks in Experiment 1 (prime distractor distributions). Bars show ±1 *SEM*.

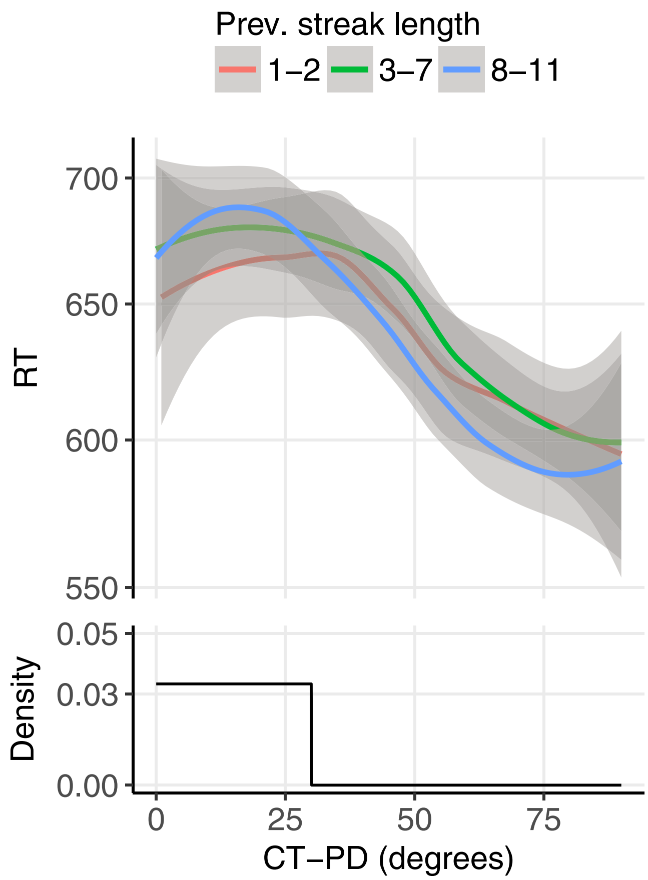

Figure 5. Search times as a function of the distance between the current target and the mean of the previous prime distractor distribution (CT – PD) in Experiment 1. The top panel shows the fitted loess functions for the streaks of different lengths; the bottom panel shows the probability density function (PDF) of the preceding distribution. Note that units for the PDF do not have physical meaning, but the PDF can be intuitively understood as a description of the "shape" of the distribution. Gray areas show 95% confidence intervals of the fitted functions. In this and the following plots, the absolute value of CT – PD is used because the analyzed distributions were symmetric. Hence, a uniform distribution with a range of 60° is shown to range from 0° to 30°.

Discussion

In Experiment 1, we replicated the effects of uniform distractor distributions on RTs for the consequent test distribution (Chetverikov et al., 2016). The RT $\propto$ CT – PD function consisted of two segments: The first had zero slope, and the second had a negative slope. In essence, search was slow as long as the target was inside the range of the previous uniform distractor distribution and fast when the target was outside of it. This result replicates previous findings that used the same two-segmented function and demonstrates the validity of the modeling approach we use.

The new finding here is that the number of repetitions did not influence the shape of the RT $\propto$ CT – PD function. Observers' responses on test trials indicate that they treated the preceding distribution as uniform following only a few repetitions. Further repetitions neither helped nor hindered performance.

The results of Experiment 1 demonstrate that observers do not have any strong nonuniform priors when they encode distractor distributions. Given that the mean value for the distribution shape changes for every prime streak, the results cannot be explained by carryover effects from preceding streaks. Moreover, our previous experiments show that neither learning target orientation nor learning the range of distributions can explain the correspondence between the shape of the RT as a function of CT – PD and the shape of the distribution (Chetverikov et al., 2016, see experiment 3C for results with random target orientation and experiments 2 and 3A for a comparison of the distributions with the same range). However, the effects obtained with short streaks in the present study might partially be explained by the fact that after several streaks observers begin to expect a uniform distribution. Moreover, the results leave open the question of the strength of uniform priors themselves. In other words, the fact that observers can easily learn uniform distributions of distractors does not imply that there are uniform priors or that nonuniform distributions cannot be learned with ease.

In Experiment 2, we tried to disentangle this by comparing the learning dynamics for uniform and Gaussian distributions. Experiment 1 did not show any effect of the number of repetitions on the shape of the RT $\propto$ CT – PD function. Thus, in Experiment 2, we contrasted short streaks (one to two trials) with relatively long streaks (six to seven trials), allowing us to increase the number of trials for each streak length used and, in combination with increased number of sessions, increase statistical power. We expected that if there was a strong prior for a uniform distractor distribution, then following a short prime streak of Gaussian trials, observers would still represent it as uniform, and a two-segment RT $\propto$ CT – PD function would be observed.

Experiment 2

Participants

Ten observers (four female, age $M = 25.10$) took part in three experimental sessions taking approximately 20–30 min each.

Distractors	Accuracy *M*	Accuracy *SD*	RT_{corr}, ms *M*	RT_{corr}, ms *SD*
Gaussian ±20 (test streaks)	0.96	0.02	606	41
Uniform ±20 (test streaks)	0.93	0.03	636	53
Gaussian ±30 (prime streaks)	0.93	0.03	609	55
Uniform ±30 (prime streaks)	0.91	0.04	663	72

Table 2. Search times and accuracy. *Notes*: Prime streaks were on average longer than test streaks, and comparisons between them should be made with caution.

Method

The procedure was similar to Experiment 1 with some modifications. The prime distribution streaks could be either long (six to seven trials) or short (one to trials). Each streak length had the same probability. To keep overall probabilities of Gaussian and uniform distributions equal, both prime and test distributions could be uniform or Gaussian except that prime distributions had a maximum possible range of 60° ($SD = 15$ for a Gaussian with outliers outside of $\pm 2 \times SD$ removed) whereas test distributions had a 40° range ($SD = 10$ for a Gaussian with similarly removed outliers). Test distribution type was counterbalanced with prime distribution and prime streak length. Each observer participated in three sessions of 208 prime and test streaks, each lasting approximately 20–30 min.

Results

Average performance

A two-way repeated-measures ANOVA showed a significant effect of distractor distribution type on search times, $F(1, 9) = 57.98$, $p < 0.001$, $\eta^2_G = 0.11$, qualified by an interaction with distribution range, $F(1, 9) = 18.03$, $p = 0.002$, $\eta^2_G = 0.01$. For accuracy, the effects of distribution type, $F(1, 9) = 74.33$, $p < 0.001$, $\eta^2_G = 0.17$, and range, $F(1, 9) = 18.45$, $p = 0.002$, $\eta^2_G = 0.18$, did not interact. As Table 2 shows, the interaction stems from the fact that RTs were similar for Gaussian distributions with a different range, and for the uniform distributions with a different range, the RTs differed. A separate analysis of the Gaussian and uniform distributions with larger range showed differences both in RT, $t(9.0) = 7.06$, $p < 0.001$, and accuracy, $t(9.0) = 5.17$, $p < 0.001$. Similarly, with smaller range, the uniform distribution resulted in slower responses, $t(9.0) = 6.54$, $p < 0.001$, and lower accuracy, $t(9.0) = 6.54$, $p < 0.001$, than the Gaussian.

Observers responded more slowly, $B = 0.14$ (0.01), $t = 19.84$ for uniform and $B = 0.14$ (0.01), $t = 25.38$ for Gaussian distribution, and were less accurate, $B = -0.39$ (0.07), $Z = -5.82$ and $B = -0.43$ (0.08), $Z = -5.48$, respectively, on the first trial than on later trials as well as on the second trial for the Gaussian distribution, $B = 0.02$ (0.01), $t = 3.69$, and marginally slower for the uniform distribution, $B = 0.01$ (0.01), $t = 1.88$, but not more accurately, $B = 0.12$ (0.09), $Z = 1.35$ and $B = -0.07$ (0.10), $Z = -0.65$, respectively, than on later trials (Figure 6).

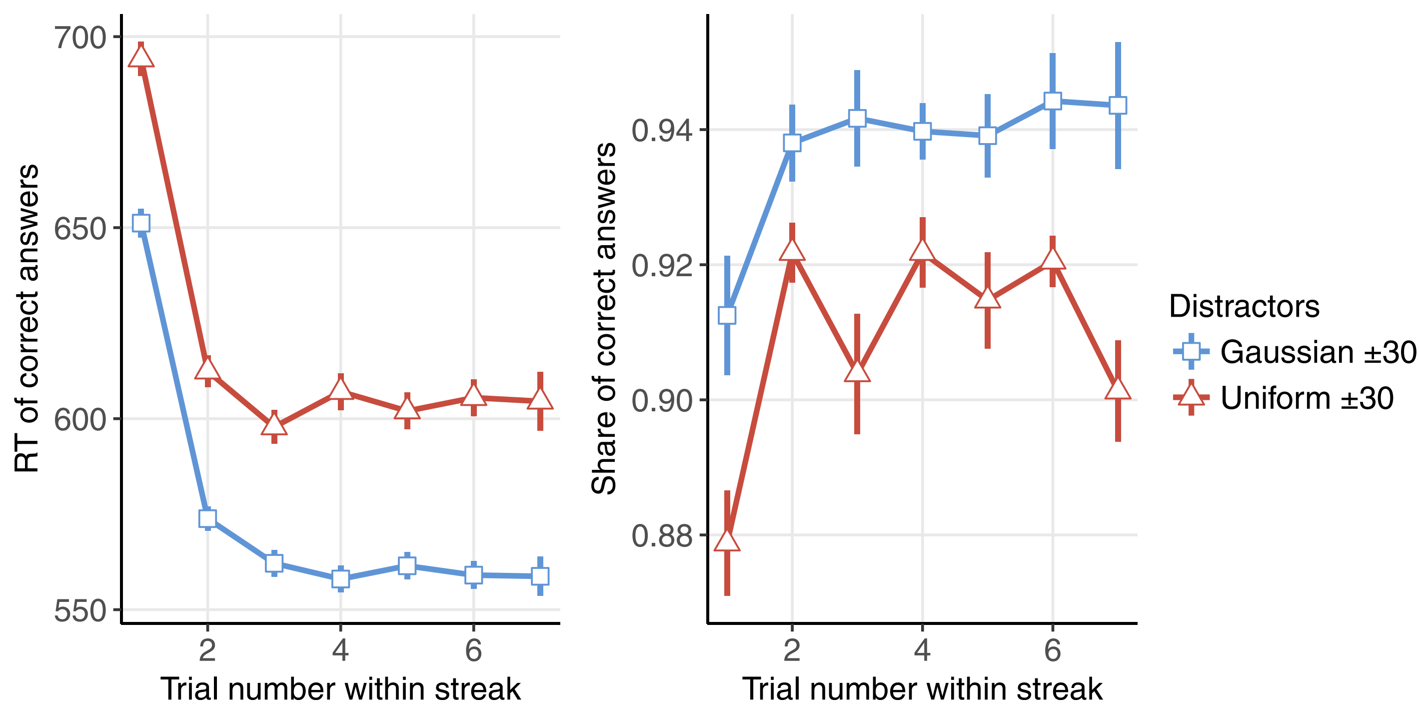

Figure 6. Repetition effects in Experiment 2 (prime distractor distributions). Bars show ±1 *SEM*.

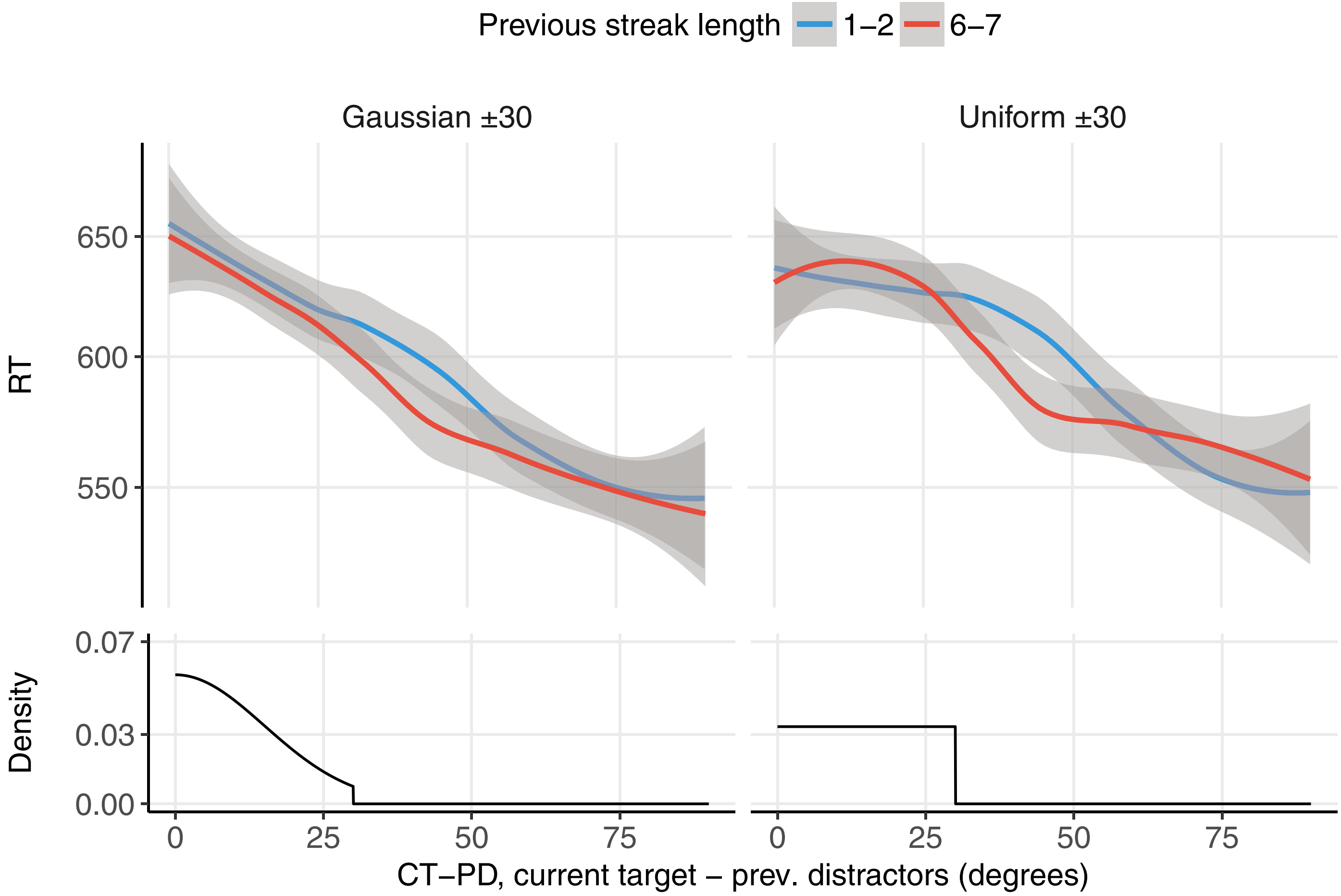

Figure 7. Search times as a function of the distance between the current target and the mean of previous distractor distribution (CT − PD) in Experiment 2. The top panel shows the fitted loess functions for the streaks of different lengths, and the bottom panel shows the probability density of the preceding distribution. Gray areas show 95% confidence intervals of the fitted functions.

RTs as function of CT − PD

Figure 7 shows that the RT ∝ CT − PD function clearly differed between the uniform and Gaussian prime streaks. A segmented regression showed that following a uniform distribution RTs can be described by a two-part linear function with a breaking point at 22.4 distance (Davies' test $p = 0.040$). The slope of the first part did not significantly differ from zero, $B = 0.40$, 95% $CI = [-2.64, 3.44]$, and for the second part, the slope was significantly negative, $B = -1.43$, 95% $CI = [-1.73, -1.13]$. In contrast, following the Gaussian distribution, no significant breaking point was found (Davies' test $p = 0.732$). The results were similar for different test distribution shapes (Supplementary Figure S1).

LMER confirmed that the slopes within the distribution range (0° to 30°) were significantly more negative for the Gaussian than the uniform distributions, $B = 0.02$ (0.01), $t = 2.30$. However, streak length did not affect the slopes for the uniform, $B = 0.003$ (0.014), $t = 0.20$, or the Gaussian distributions, $B = 0.003$ (0.013), $t = 0.23$. Interestingly, the slopes outside the previous distractor distribution range were more shallow for longer streaks than for shorter streaks following the uniform, $B = 0.01$ (0.01), $t = 2.21$, but not the Gaussian distribution, $B = 0.01$ (0.01), $t = 1.81$.

Discussion

Experiment 2 shows that for both uniform and Gaussian distributions the shape of the RT ∝ CT − PD functions was not affected by the number of prime trial repetitions. When the previous distractor distribution was Gaussian, RTs monotonically decreased with increasing CT − PD distance already after one to two repetitions. However, when the previous distractor distribution was uniform, the RT ∝ CT − PD function had a two-segment shape with a flat slope in the first segment and a negative slope in the second segment, roughly following the probability density function of the previous distractor distribution. Note that both Figures 5 and 7 indicate that the length of the first segment is shorter and closer to the actual distribution

Distractors	Accuracy		RT_{corr}, ms	
	M	*SD*	*M*	*SD*
Gaussian, *SD* = 10 (test streaks)	0.95	0.02	637	62
Bimodal (prime streaks)	0.88	0.04	739	98

Table 3. Search time and accuracy in Experiment 3. *Notes*: The bimodal distribution had longer streaks, and thus the comparisons between distributions should be made with caution.

range with longer streaks. Thus, it is possible that additional repetitions allow for more precise estimation of distribution shape. This could explain the effect of previous streak length on slopes of the second segment of the RT $\propto$ CT – PD function for uniform distributions: Longer streaks might allow a more detailed representation of the probability outside the previous distractor distribution. Nevertheless, the main conclusion of Experiment 2 is that one or two repetitions are enough for observers to grasp the general shape of distractor distributions.

In Experiments 1 and 2, the prime distributions were relatively simple and may have been easy to learn, which might explain why further repetitions have little or no effect. In Experiment 3, we therefore used a bimodal distribution to measure how observers would represent this more complex distribution following short streaks and whether they would be able to learn the true shape of the distribution with increased repetitions.

Experiment 3

Method

Procedure

The procedure was the same as in Experiment 1 except that the prime distribution was bimodal (see Figure 2, right panel), consisting of two uniform distributions with orientations randomly drawn from (–30°, –15°) and (15°, 30°) segments of feature space (relative to the distribution center).

Participants

Eleven observers[5] (three female, age $M = 24.91$) took part in two experimental sessions taking approximately 20–30 min each.

Results

Average performance and repetition effects

Similarly to the uniform distribution in Experiment 2, the bimodal distribution search was more difficult than the Gaussian, both in search time, $t(10.0) = 6.83$, $p < 0.001$, and accuracy, $t(10.0) = 6.39$, $p < 0.001$ (Table 3). There was a significant quadratic effect of target-to-distractor distance for both Gaussian, $B = 0.76$ (0.25), $t = 3.02$, and bimodal, $B = 5.27$ (0.35), $t = 15.19$, distributions. The repetition effect was most noticeable after the first trial (Figure 8). LMER with Helmert

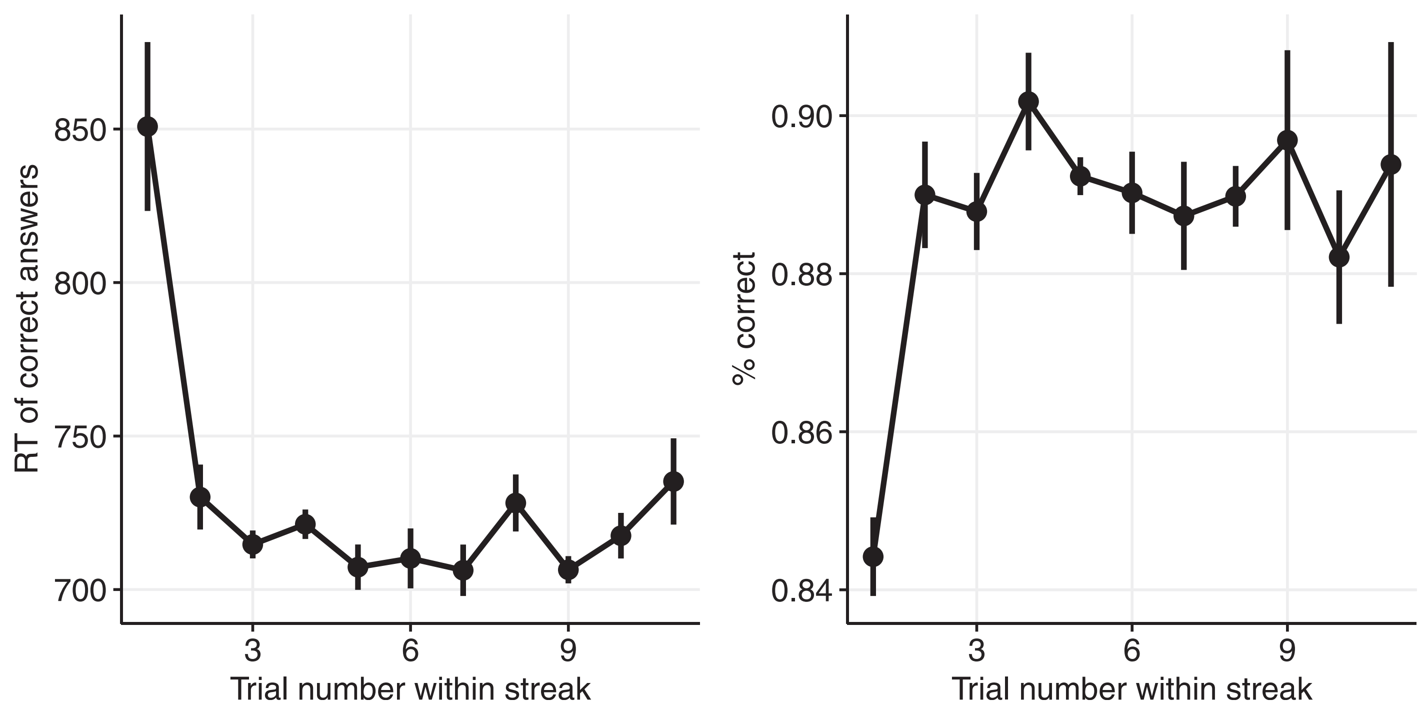

Figure 8. Repetition effect within streak in Experiment 3 (prime distractor distributions). Bars show ±1 *SEM*.

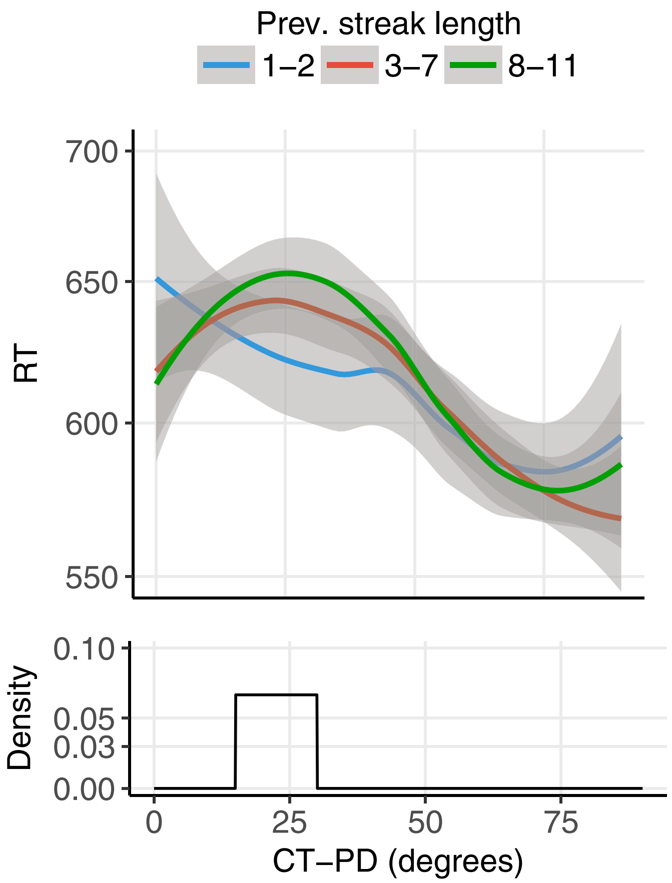

Figure 9. Search times as a function of the distance between the current target and the mean of the previous distractor distribution (CT − PD) in absolute degrees in Experiment 3. The top panel shows the fitted loess functions for the streaks of different lengths, and the bottom panel shows the probability density function of the preceding distribution. Gray areas show 95% confidence intervals of the fitted functions. As in previous plots, absolute values of CT − PD are analyzed, so the bimodal distribution with two segments at −15° to −30° and 15° to 30° is shown as a single-segment distribution.

contrasts indicated that search on the first trial was slower, $B = 0.15$ (0.01), $t = 22.52$, and less accurate, $B = -0.42$ (0.05), $Z = -7.97$, than on later trials.

RTs as function of CT − PD

Figure 9 shows that following a bimodal distribution the RT ∝ CT − PD function differed from those observed in previous studies. Given that a bimodal distribution with CT − PD analyzed in absolute degrees has three parts (0° to 15°, 15° to 30°, and 30° to 90°), we first ran a segmented regression with two break points. However, because the second breaking point did not correspond to any significant change ($p = 0.868$) in slope, we repeated the analysis assuming only one breaking point. The analysis revealed a breaking point at 14.0° ($p < 0.001$). There was a tendency-level positive

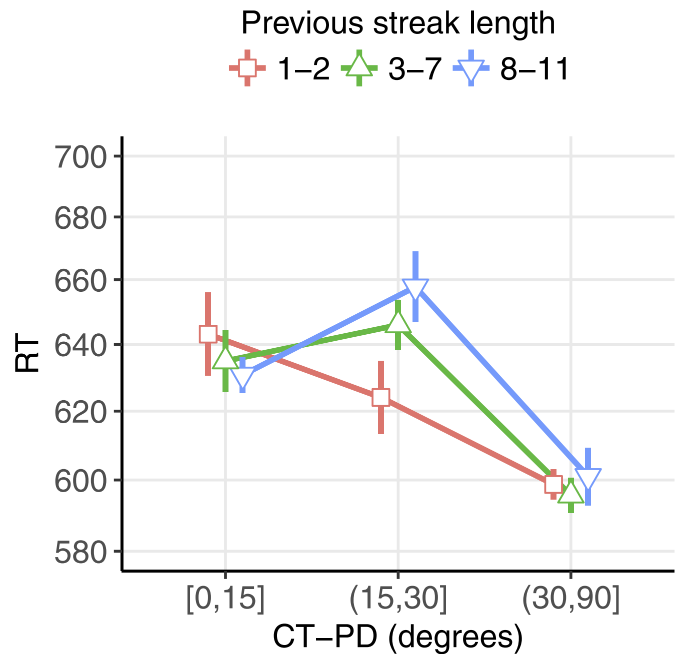

Figure 10. Search times as a function of the distance between targets and previous distractors' mean split by previous streak length in Experiment 3. Bars show ±1 *SEM*.

slope for the first part, $B = 3.33$, 95% $CI = [-0.05, 6.71]$, and for the second part, the slope was significantly negative, $B = -1.38$, 95% $CI = [-1.66, -1.10]$. We then ran a mixed effects regression that included the effect of the CT − PD split into three segments (0° to 15°, 15° to 30°, and 30° to 90°) controlling for distance between previous target and the current distractor and the accuracy of the preceding response. As in Experiment 1, targets far from the mean of the preceding distribution (>30°) were processed significantly faster, $B = -0.03$ (0.01), $t = -2.66$. Crucially, however, RTs in the 15° to 30° segment were slower than in the 0° to 15° segment, $B = 0.02$ (0.01), $t = 2.15$.

The effect of the previous distribution on RTs became more pronounced with more trials. To facilitate the comparisons, we contrasted short (one to two trials), medium (three to seven trials), and long streaks (eight to 11 trials). For the shortest streaks, RTs for the 0°–15°s CT − PD segment and the 15°–30° CT − PD segment did not differ, $B = -0.02$ (0.03), $t = -0.85$. However, a significant interaction between the effect of the segment and previous streak length indicated that for the longest streaks, RTs for the 15°–30° segment became slower than for the 0°–15° segment, $B = 0.07$ (0.03), $t = 2.03$ (see Figure 10, blue line). Pair-wise comparisons confirmed that the difference between the 0°–15° and 15°–30° segments was significant for the longest streaks, $B = 0.04$ (0.02), $p = 0.033$, but nonsignificant for the midlength, $B = 0.03$ (0.02), $p = 0.239$, and the shortest streaks, $B = -0.02$ (0.03), $p = 0.601$. The difference between the 30°–90° segment and the 15°–30° segment was already significant for the shortest streaks, $B = -0.05$ (0.02), $t = -2.17$, and did not interact with streak length, $B = 0.03$ (0.03), $t = 1.17$.

Discussion

Two important conclusions follow from Experiment 3. Initially, observers fill in the gaps between the modes of a bimodal distribution, seemingly assuming that the distribution is unimodal and uniform between the modes. In fact, with short streaks, RTs are longer when the current target falls in the region *between* the peaks of a previous bimodal distractor distribution (0°–15° in Figure 10) than when it falls outside its range (30°–90° in Figure 10). It is as if this part of the feature space between peaks is also occupied by distractors.

With more repetitions—and this is the second important finding—observers begin to differentiate between the region in between the modes and the regions at the modes. That is, they learn that the probability of seeing a particular distractor is not uniform. However, the effect of this additional learning is relatively small compared to the overall difference between the region between the modes and the region outside of them.

General discussion

In the three experiments reported here, we investigated for the first time the dynamics of stimulus distribution encoding over several trials. Using visual search for an oddly oriented line, we asked whether observers have any strong priors when they encode distractor features and whether the knowledge about distractor distributions improves with repetitions. To answer this question, we assessed how the shape of previously observed distractor distributions influences search times depending on the number of repetitions and the shape of the distribution.

In Experiment 1, we demonstrated that observers can represent uniform distributions already after one or two trials, and further trials do not change RT patterns. Experiment 2 showed that observers are equally fast at learning uniform and Gaussian distributions. The results from these experiments replicate and extend our previous findings (Chetverikov et al., 2016). As in previous experiments, we found that RTs as a function of the distance between the mean of the previous distractor distribution (CT – PD) correspond to the shape of distractor distribution. The RTs seemingly correspond to observers' (implicit) expectations. Following uniform distributions, the expectations are similarly uniform within the range of the preceding distribution, and following Gaussian distributions, they monotonically decrease even though this distribution has the same range.

However, in contrast to our previous experiments, we varied the number of repetitions and were able to study the prior expectations and dynamics of distribution encoding. We reasoned that observers' representation of distractor distribution would depend both on their prior expectations and the actual shape with the weight of the priors decreasing with repetitions. Hence, we expected that any strong priors would be visible after short streaks. However, our findings demonstrate that observers do not have any strong priors related to distribution shape when they process distractors in visual search. Both Gaussian and uniform distributions are encoded correctly after one or two repetitions, and further repetitions do not affect observers' representations. It might seem surprising, but on the other hand, each trial contains 35 distractors, and 70 distractors appear to be more than enough to obtain probability density estimates. The average RTs in Experiment 2 for the first two trials in the streak were 600–700 ms, providing a conservative estimate of 2 s for analyses of the distributions during two trials. It is possible that with shorter presentation times encoding might differ. At the very least, our findings convincingly indicate that, even if representations of distractor distributions are influenced by any prior expectation when observers first encounter the stimulus distribution, this prior dissipates amazingly quickly under the influence of incoming sensory data.

Experiment 3 provided important additional information about distribution shape learning. When confronted with a bimodal distractor distribution (consisting of two separate uniform segments), observers initially treated it as unimodal, interpolating the gap between the two modes as if there were distractors there. However, with further repetitions, their RTs gradually became different for the feature space that was occupied by distractors and the interpolated part between the modes of the distribution that was not occupied by distractors. Their RTs started to follow the shape of the preceding distribution more closely, indicating that they began to encode the bimodality of the distribution.

The results for the bimodal distribution show that observers are limited in their ability to approximate the shape of distributions. Although Gaussian or uniform distributions are easily approximated, a bimodal distribution is not. Bimodal distributions are only approximated following five to seven repetitions. This might indicate that observers initially expect unimodal distributions and only after obtaining more information start to separate the modes (Utochkin & Yurevich, 2016). The encoding of object ensembles is therefore flexible but not without limits.

These results are novel and surprising given the role that Gaussian distributions play in accounts of ensemble perception (Rosenholtz, 2001; Utochkin, 2015). However, the fact that the distribution is encoded as uniform does not mean that it cannot be further approximated as Gaussian if necessary. In fact, the results of our previous experiments (Chetverikov et al., 2016) and Experiment 2 here demonstrate that observers can correctly encode the shape of Gaussian distributions. But apparently they do not enforce such approximations on data when they do not fit. As we argued before (Chetverikov et al., 2016), encoding of the shape of feature distributions can be implemented in many ways, ranging from simple counting of specific orientations to an approximation with multiple simple functions (such as Gaussian functions) or using higher-order moments, such as skewness or kurtosis. We do not advocate a specific mechanism as the research into distribution representations has only just began. However, simple counting seems to be unlikely given the results of Experiment 3 as it does not explain the interpolation between distribution peaks with short streaks. It is then likely that more complex mechanisms are involved.

Our results further demonstrate the advantages of our new method for studying internal representations of ensembles. In a feature-based search task, targets and distractors cannot have the same features. The more probable it is that distractors have features from a particular region of feature space, the less probable it is that a target will have features from this region, and the slower observers respond when this actually happens. The mapping between features in the physical world, the corresponding probabilities in perceptual space, and the mapping between the latter and RTs is nonlinear. However, as our results demonstrate, the resulting RT patterns correspond to distributions of features in the physical world, and varying the parameters of the distributions of features in the physical world and analyzing RTs as a function of CT – PD enables the study of the characteristics of ensemble perception and the underlying mechanisms.

Keywords: visual search, learning, ensemble representations, summary statistics, priming of pop out, line orientation

Acknowledgments

This paper was supported by the Russian Foundation for Humanities (#15-36-01358A2) and a grant from the Icelandic Research Fund (IRF #152427). The data from the experiments reported in this paper is available at https://osf.io/3apcv/.

Commercial relationships: none.
Corresponding author: Andrey Chetverikov.
Email: andrey@hi.is.
Address: Department of Psychology, University of Iceland, Reykjavík, Iceland.

Footnotes

[1] We did not test for observers' awareness of the difference between distributions, but previous studies (Atchley & Andersen, 1995; Dakin & Watt, 1997) show that observers perform poorly on tasks requiring explicit discrimination between distributions based on skewness or kurtosis. Hence, it is highly unlikely that observers have conscious access to representations of distributions in our studies.

[2] How exactly physical feature space, perceptual domain, and RTs are mapped to each other is unclear. But we assume that the relationship between them is monotonic: The lower the probabilities in physical space are, the lower are the probabilities in perceptual space, and the RTs are subsequently longer.

[3] We expect a monotonic decrease rather than a sharp drop of RTs and then similarly fast RTs outside of the range of the previous distractor distribution because previous results show that the sharp drop in distractors' probability corresponds to a gradual decrease of response times (Chetverikov et al., 2016). Whether this is a result of encoding the borders of the distribution or simply a noisy representation is not yet known.

[4] The lack of the effect from third and later repetitions is unlikely to result from the different number of trials per repetition. Figure 4 shows that the confidence intervals become only slightly larger with increasing trial number in a streak. Moreover, Helmert contrasts compare each trial with the average of the following trials within the streak, and hence, the number of trials per comparison remains high. Only the latest comparison, e.g., ninth trial ($N = 879$) versus 10th and 11th ($N = 835$) may begin to suffer from the comparatively lower power. Hence, we do not think that our conclusions regarding the repetition effects can be explained by the differences in the number of trials within different streaks. Later comparisons related to CT – PD effects for different streak length are not affected by different numbers of trials in the streaks. Although there are more first trials in prime streaks than 11th trials, the number of streaks consisting of one trial and 11 trials is balanced in the design.

[5] We planned to enroll 10 observers as in previous experiments; however, data from an additional participant were collected due to an error in the schedule. The

results do not change if the data from the last observer is discarded.

References

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, *15*(3), 122–131, doi: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, doi:10.1111/j.1467-9280.2008.02098.x.

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*(2), 157–162. Retrieved from http://pss.sagepub.com/content/12/2/157.short

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

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. Retrieved from http://cran.r-project.org/package=lme4

Bauer, B. (2015). A selective summary of visual averaging research and issues up to 2000. *Journal of Vision*, *15*(4):14, 1–15, doi:10.1167/15.4.14. [PubMed] [Article]

Chetverikov, A., Campana, G., & Kristjánsson, Á. (2016). Building ensemble representations: How the shape of preceding distractor distributions affects visual search. *Cognition*, *153*, 196–210, doi:10.1016/j.cognition.2016.04.018.

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

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

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

Gekas, N., Seitz, A. R., & Seriès, P. (2015). Expectations developed over multiple timescales facilitate visual search performance. *Journal of Vision*, *15*(9):10, 1–22, doi:10.1167/15.9.10. [PubMed] [Article]

Girshick, A. R., Landy, M. S., & Simoncelli, E. P. (2011). Cardinal rules: Visual orientation perception reflects knowledge of environmental statistics. *Nature Neuroscience*, *14*(7), 926–932, doi:10.1038/nn.2831.

Haberman, J., Brady, T. F., & Alvarez, G. A. (2015). Individual differences in ensemble perception reveal multiple, independent levels of ensemble representation. *Journal of Experimental Psychology. General*, *144*(2), 432–446, doi:10.1037/xge0000053.

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, England: Oxford University Press.

Kristjánsson, Á. (2006). Rapid learning in attention shifts: A review. *Visual Cognition*, *13*(3), 324–362, doi:10.1080/13506280544000039.

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, doi: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, doi:10.1016/j.visres.2008.02.007.

Kristjánsson, Á., Ingvarsdóttir, Á., & Teitsdóttir, U. D. (2008). Object- and feature-based priming in visual search. *Psychonomic Bulletin & Review*, *15*(2), 378–384, doi:10.3758/PBR.15.2.378.

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, doi:10.1016/j.visres.2007.10.009.

Lamy, D. F., & Kristjánsson, Á. (2013). Is goal-directed attentional guidance just intertrial priming? A review. *Journal of Vision*, *13*(3):14, 1–19, doi:10.1167/13.3.14. [PubMed] [Article]

Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: I. Role of features. *Memory & Cognition*, *22*(6), 657–672. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7808275

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, doi:10.1080/13506280500277488.

Michael, E., de Gardelle, V., & Summerfield, C. (2014). Priming by the variability of visual information. *Proceedings of the National Academy of Sciences,* *USA*, *111*(21), 7873–7878, doi:10.1073/pnas.1308674111.

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. Retrieved from http://cran.r-project.org/doc/Rnews/

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13, doi:10.1016/j.jneumeth.2006.11.017.

Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin*, *68*(1), 29–46, doi:10.1037/h0024722.

Pollard, P. (1975). Intuitive judgments of proportions, means, and variances: A review. *The Review of Black Political Economy*, *5*(2), 5–18, doi:10.1007/BF02689368.

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, doi:10.1037//0096-1523.27.4.985.

Utochkin, I. S. (2015). Ensemble summary statistics as a basis for rapid visual categorization. *Journal of Vision*, *15*(4):8, 1–14, doi:10.1167/15.4.8. [PubMed] [Article]

Utochkin, I. S., & Tiurina, N. A. (2014). Parallel averaging of size is possible but range-limited: A reply to Marchant, Simons, and De Fockert. *Acta Psychologica*, *146*, 7–18, doi:10.1016/j.actpsy.2013.11.012.

Utochkin, I. S., & Yurevich, M. A. (2016). Similarity and heterogeneity effects in visual search are mediated by "segmentability." *Journal of Experimental Psychology*: *Human Perception and Performance*, *42*(7), 995–1007, doi.:10.1037/xhp0000203.

Wang, D., Kristjánsson, Á., & Nakayama, K. (2005). Efficient visual search without top-down or bottom-up guidance. *Perception & Psychophysics*, *67*(2), 239–253, doi:10.3758/BF03206488.

Zhang, R., Kwon, O.-S., & Tadin, D. (2013). Illusory movement of stationary stimuli in the visual periphery: Evidence for a strong centrifugal prior in motion processing. *Journal of Neuroscience*, *33*(10), 4415–4423, doi:10.1523/JNEUROSCI.4744-12.2013.