

# Learning features in a complex and changing environment: A distribution-based framework for visual attention and vision in general

Andrey Chetverikov<sup>\*,†,‡,1</sup>, Gianluca Campana<sup>§,¶</sup>, Árni Kristjánsson<sup>\*</sup>

<sup>\*</sup>Laboratory for Visual Perception and Visuomotor Control, Faculty of Psychology, School of Health Sciences, University of Iceland, Reykjavik, Iceland

<sup>†</sup>Cognitive Research Lab, Russian Academy of National Economy and Public Administration, Moscow, Russia

<sup>‡</sup>Saint Petersburg State University, Saint Petersburg, Russia

<sup>§</sup>Department of General Psychology, University of Padova, Padova, Italy

<sup>¶</sup>Human Inspired Technology Research Centre, University of Padova, Padova, Italy

<sup>1</sup>Corresponding author: Tel.: +354-7851301, e-mail address: andrey@hi.is

---

## Abstract

What are the building blocks of our visual representations? Whatever we look at, the things we see will have some feature variability: even snow is not purely white but has a range of shades of white. However, in most studies investigating visual perception, *homogeneous* displays with all stimuli having a very limited range of features have been used. In contrast, recent studies using *heterogeneous* displays have shown that our perceptual system encodes surprisingly detailed information about stimuli, representing parameters such as the mean, variance, and most importantly the probability density functions of feature distributions. Learning the parameters of the distributions takes time as distribution representations are continuously updated with incoming information. However, the mechanisms guiding this process are not yet known. We will review current knowledge about the sampling and updating of representations of feature distributions in heterogeneous displays and will present new findings providing further insights into this process. Overall, the results show that representations of distributions can be remarkably detailed and shed light on how the information provided affects the learning of feature distributions. Observers' ability to quickly encode the probability density function of distributions in the environment may potentially provide novel interpretations of a number of well-known phenomena in visual perception.

---

## Keywords

Feature distributions, Ensemble perception, Summary statistics, Visual search, Visual attention, Perceptual learning, Texture perception, Probabilistic perception, Attention

---

## 1 INTRODUCTION

When you look out your window, you might see the sky, some trees, maybe a road, or a walkway. How do we represent the things we see outside our window? The textbook answer is that there are *features* in the image that are bound together during perceptual processing to create *objects*. Features are globally independent, in the sense that processing of a feature at one location does not depend on features at another location except for possible local interactions. This view is undermined, however, by studies showing that summary statistics, the properties of features aggregated over sets of stimuli or spatial locations, are no less important than specific features in perceptual processing (Albrecht and Scholl, 2010; Alvarez, 2011; Alvarez and Oliva, 2008; Ariely, 2001; Chong and Treisman, 2003, 2005; Dakin and Watt, 1997; Parkes et al., 2001). Moreover, the gist of a scene based on the summary representation of elements in the ensemble can be grasped even with short presentation times and even if they are unattended, while individual features can be missed (Brand et al., 2012; Chong and Treisman, 2005; Emmanouil and Treisman, 2008; Whiting and Oriet, 2011). Recent findings also indicate that observers can encode *probability distributions* of features instead of indiscriminately using approximations with a limited set of summary statistics (Chetverikov et al., 2016, 2017c). Representations of distributions are then continuously updated when new information is obtained, with complex distributions requiring more information for accurate encoding than simpler ones (Chetverikov et al., 2017a). As a result, perceptual ensembles might play a key role in the richness of perceptual experience (Cohen et al., 2016).

The idea of perception as probabilistic *inference* is certainly not new but currently experiences a revival (Feldman, 2014; Feldman and Friston, 2010; Fiser et al., 2010; Ma, 2012; Ma et al., 2011, 2015; Rao et al., 2002; Seriès and Seitz, 2013; Yuille and Kersten, 2006). However, we argue that the implications of perception with probabilistic *representations*, that is, distribution encoding, are severely underappreciated and have far-reaching consequences for the scientific study of visual perception.

We suggest that it is time to step away from the idea of features as single values, and switch to a new perspective with probability distributions as a basic unit of analysis. We propose that the perceptual system accumulates information about features over and above local values (e.g., a color value of a single pixel), local feature contrasts (e.g., when orientation or brightness change abruptly), and simple statistics such as measures of central tendency or variability. Instead, probability density functions (PDFs) of feature distributions are assembled and represented in perception and the obtained information is used to guide vision. The PDF is a well-known concept in statistics, defined simply as a function that describes the relative likelihood that a

random variable will have a given value. A representation of features as a PDF implies that for a given stimulus a representation of its property is not a single value or a set of values but rather a function determining how likely it is that the stimulus has a given value. Knowledge about feature distributions allows observers to behave in a Bayes-optimal way by using probability densities for search, categorization, and other visual function (e.g., [Ma et al., 2011, 2015](#); [Vincent et al., 2009](#)). Unlike computer displays, however, where the exact color value of a pixel can be measured by using a photometer, there is always some uncertainty in representations. Thus, Bayes-optimal computing does not mean that observers are ideal: they might have to rely on approximations to represent distributions (see Section 2).

We further assume that such representational PDFs can either be bound to specific locations in the visual scene or to entities not bound to a single location. Imagine that you are looking for your cat in the backyard ([Fig. 1B](#)). The trees and bushes have a specific distribution of colors and a specific location, while your cat has a specific distribution of colors but can be anywhere. Global, nonbound PDFs are more difficult to learn, but the evidence for their existence is abundant in the literature on natural environment statistics ([Coppola et al., 1998](#); [Girshick et al., 2011](#); [Long et al., 2006](#); [van Bergen et al., 2015](#); [Yang and Purves, 2004](#)). Locally bound PDFs, on the other hand, are easier to learn but are also more likely to dissipate easily (Section 3).

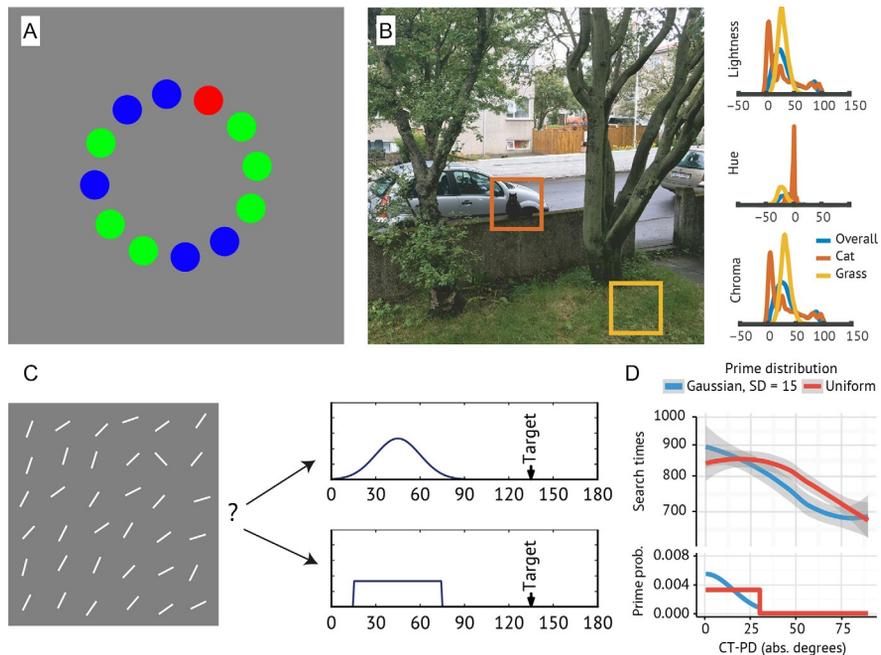
We further argue that the distribution-based framework advances our understanding of visual processing by providing simple explanations for some previously well-known phenomena and by predicting new findings, some of which we test in the experiments described later. Although we focus mainly on visual search, we believe that the present framework has important implications for studies of vision in general. Indeed, we also describe some potential applications of the framework to other topics within the visual perception literature (Section 4).

---

## 2 WHAT IS A “DISTRIBUTION” OF FEATURES AND HOW IS IT REPRESENTED?

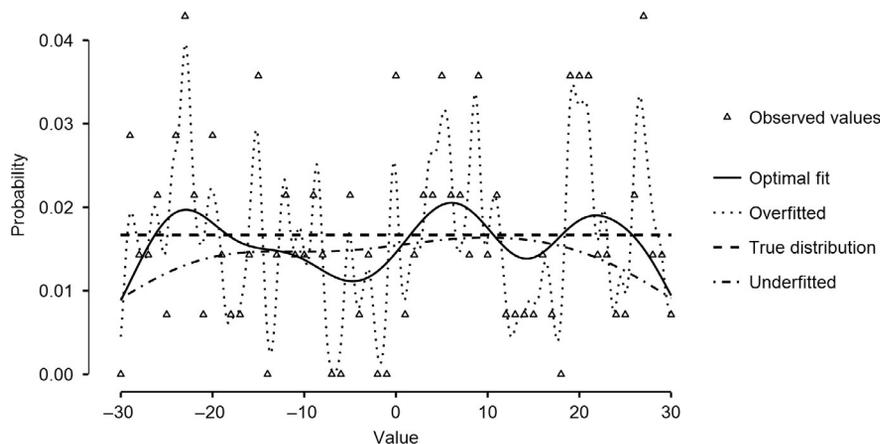
It is easy to define a distribution when generating stimuli: one can select from a range of well-defined PDFs (e.g., Gaussian or uniform) and generate random numbers conforming to that function. However, what is a distribution of features when we try to assess it from empirical data? Imagine that a researcher randomly picks one hundred colored patches from a distribution of colors with hues varying uniformly from  $-30$  to  $+30$  on some arbitrary dimension. The observed probabilities of each value will be different from the true distribution ([Fig. 2](#)). These observed values are noisy because the values are picked randomly. In real-life scenes there can be many more observed values, but they still would not represent the true distribution due to inherent external and internal noise. In general, a sample can never be a true representation of the underlying distribution, and a representation of a distribution will therefore always involve approximations to some degree in order to reduce noise.

## 4 Distribution-based framework for visual attention

**FIG. 1**

Unlike the artificial displays used in many laboratory experiments (A), the *colors* in the natural scenes contain a multitude of hues and shapes (A and B). A target and distractors in a real-world visual search task, for example (such as the cat and the grass in (B)), do not have homogeneous *colors*, but their *colors* are not random either. Rather, they come from specific probability distributions. Given a stimulus display (C), it is hard to match the depicted stimuli to the correct distributions. If you look at the *oriented lines*, can you tell whether they come from a uniform distribution of distractor orientations or from a Gaussian one? This seems like a difficult problem. Nevertheless, only a few examples of such displays, search times provide evidence that our visual systems can tell such distributions apart. Panel (D) shows the correspondence between the internal representation of a distribution as revealed by response times and the physical distribution of stimuli (see details in text). Note that while the response time functions show remarkable similarity to the actual PDF's (there is a monotonic relationship between the two), there is nevertheless not a one-to-one correspondence between the physical PDF of stimuli (*lower plot*) and the PDF estimated using search times (*upper plot*).

Panel (D) adapted from experiment 3C, Chetverikov, A., Campana, G., Kristjánsson, Á., 2016. *Building ensemble representations: how the shape of preceding distractor distributions affects visual search.* *Cognition* 153, 196–210. <http://dx.doi.org/10.1016/j.cognition.2016.04.018>.

**FIG. 2**

Examples of the difference between a true distribution of an arbitrary feature distribution (ranging from  $-30$  to  $+30$  in this example), empirically observed probabilities and estimated distributions. *Dots* show mean empirical probabilities, while *lines* show the true distribution (which is in this case uniform), and different fits. While the optimal fit (as determined by Sheather and Jones (1991) algorithm implemented in *bw.SJ* in R) provides a balance between the observed data and a true distribution, the underfitted curve is closer to the true distribution, and the overfitted curve is closer to the observed data.

Such approximations can provide optimal fits to empirical data, yet they will always be different from the true distribution, since they will unavoidably reflect sampling error in addition to the distribution itself. On the other hand, overfitting or underfitting can provide closer fits to the true distribution while being further away from the data. In computer algorithms the degree of approximation can be controlled, for example, through bandwidth parameters when kernel density estimation is used to determine the shape of some function. In the brain, the resulting approximation might depend, among other parameters, on the width of neuron tuning curves or on the strength of lateral inhibitory connections. In sum, there is a difference between the true distribution of features and a representation of that distribution, and the representation that fits the data is not necessarily the best approximation of the true distribution.

In our studies, we have used visual search tasks to estimate observers' internal representations of distributions in the environment. In a typical experiment, observers search for an odd-one-out target among a setlines differing in orientation serving as distractors (Chetverikov et al., 2016, 2017a) or search for an oddly colored diamond (Chetverikov et al., 2017c). The distractors on each trial were randomly drawn from a specific distribution (see an example of orientation search display in Fig. 1C). During several adjacent trials (depending on the experiment; the number of repetitions varied from 1 to 11, with 5 to 7 repetitions being the standard),

parameters of the distractor distribution were kept constant and target orientation was either kept constant as well or varied randomly. Then, test trials followed where distractor distributions and the target had different distances in feature space (e.g., orientation or color) from the mean of a previous distractor distribution. It is well known that when observers expect distractors to have specific features, they respond slower when these features instead belong to the target (referred to as “role reversals,” [Kristjánsson and Driver, 2005, 2008](#); [Lamy et al., 2008](#); [Maljkovic and Nakayama, 1994](#); [Wang et al., 2005](#)). This means that measuring response times (RTs) as a function of the distance in feature space between the target and the mean of previous distractor distributions (CT-PD, current target to previous distractor distance) allows us to assess observers’ representations of previously encountered distractor distributions, and therefore expectations about successive ones. If the presented target causes a role-reversal effect, we can assume that it has a feature that belongs to the set of expected feature values for the distractor set. Looking at where in the preceding distractor distribution this target belongs (CT-PD), and then measuring the size of the role-reversal effect allows us to assess the representation of the distribution.

[Fig. 1D](#) shows the results from one of our experiments where we compared representations of uniform and Gaussian distributions using role-reversal effects to assess distributions (experiment 3C, [Chetverikov et al., 2016](#)). The topline shows RTs to targets depending on where they fell with respect to the previous distractor distribution, presented for five to seven consecutive search trials that preceded the test trial. While there is a correspondence between the physical distribution of stimuli and their perceptual representations (namely, there is a monotonic relationship between the two functions), it is not a simple one-to-one match. For example, a smooth decrease is observed for the uniform distribution instead of a sharp drop in the right part of the function. Comparisons of natural statistics and observers’ priors also reveal similarities without an exact match (e.g., [Girshick et al., 2011](#)). These data show that the perceptual system uses approximations when processing and representing empirical PDFs. It is probably not surprising given that there are many processing stages involved, and the ultimate goal of perception is to guide our interaction with the world, rather than to provide a precise representation of the environment.

---

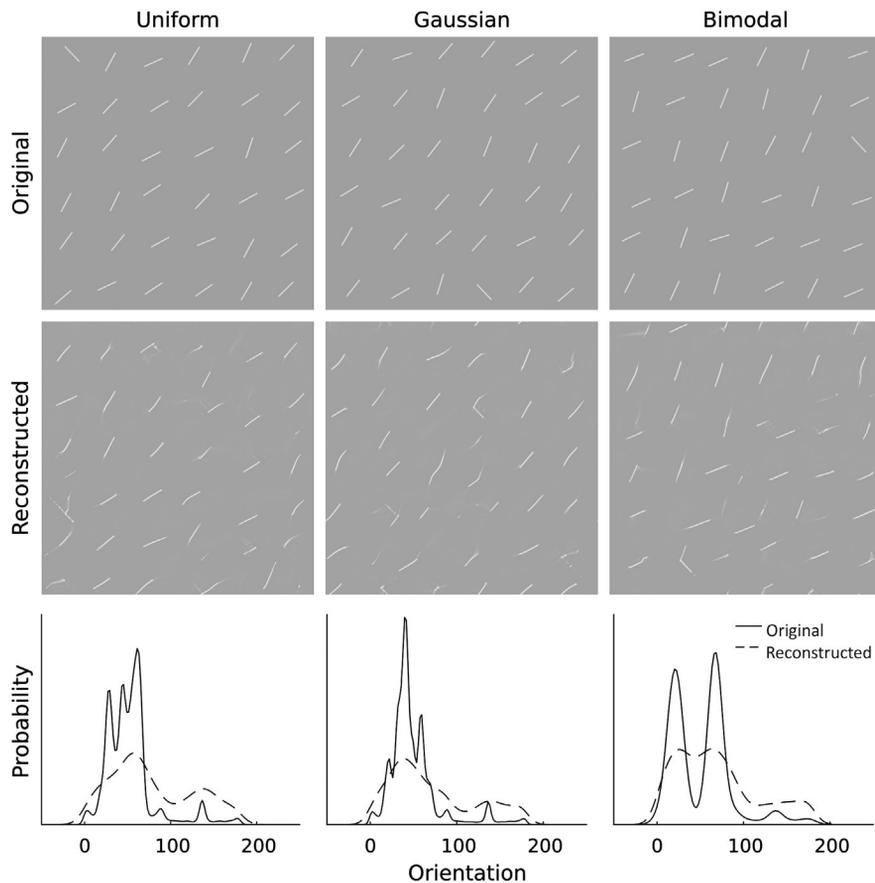
### 3 WHEN ARE STIMULI TREATED AS DISTRIBUTIONS?

Our perceptual system may engage different mechanisms depending on whether stimuli are treated as a bunch of isolated elements or as a set of exemplars from the same distribution ([Bacon and Egeth, 1991](#); [Meinecke and Donk, 2002](#); [Nothdurft, 2000](#); [Pöder, 2006](#); [Sagi and Julesz, 1987](#)). For example, [Sagi and Julesz \(1987\)](#) found nonmonotonic changes in search efficiency with increasing stimulus density and suggested that preattentive processing of target–distractor contrasts is range limited and cannot be used with low-density displays. [Bacon and Egeth \(1991\)](#), however, demonstrated that it is not target–distractor contrasts but

distractor–distractor groupings that become easier with increasing display densities. Thus, displays that allow grouping might be treated as single distributions, while sparse displays might lead either to separation into subgroups with independently estimated features or even to a lack of distribution processing at all. This can explain our results from experiments with small set sizes where no distribution learning was observed (Chetverikov et al., 2017b). Interestingly, by this logic it is possible that two displays traditionally used in a variety of visual tasks, namely, annular displays (with stimuli equidistant to fixation point) and field-like displays (with stimuli uniformly distributed over a circular or square field), might result in different kinds of processing, complicating comparisons between them.

Spatial proximity is not the only factor that determines grouping of stimuli into distributions. A typical finding when heterogeneous (bimodal) and homogeneous search displays are compared is that it is easier to find a target among homogeneous distractors than heterogeneous ones (Feldmann-Wüstefeld and Schubö, 2013, 2015; Mazza et al., 2009; Schubö et al., 2007). This can be linked to distractor–distractor grouping and a consequent shift to distribution processing. Further evidence for this is provided by findings on “segmentability” that show more efficient visual search for uniformly distributed distractors than for bimodal (consisting of two homogeneous distributions) or trimodal (three homogeneous) distributions with the same range (Utochkin and Yurevich, 2016). In this study, the variance of uniform distributions was lower than the variance of bimodal or trimodal ones. When distribution parameters, such as the number of modes or its mean, change randomly on each trial as in the experiments of Utochkin and Yurevich (2016), observers are likely to approximate bimodal or trimodal distributions as unimodal (see also Chetverikov et al., 2017a, on representations of bimodal distributions) with higher variance or range compared to the actual unimodal distribution. Search will therefore take longer in the former than the latter case. As for spatial grouping discussed earlier, in the extreme case, a lack of grouping by similarity might result in one-by-one stimulus processing during search.

In general, the idea of grouping by proximity and similarity links distribution processing to texture perception (Julesz, 1981; Wolfe, 1992; Wolfson and Landy, 1998; see Landy, 2013; Rosenholtz, 2014 for recent reviews). For example, the Texture Tiling Model utilizes a texture synthesis algorithm developed by Portilla and Simoncelli (2000) to show that visual search performance, scene perception, or crowding effects can be explained by the use of statistical summaries by peripheral vision (Balas et al., 2009; Chang and Rosenholtz, 2016; Ehinger and Rosenholtz, 2016; Rosenholtz et al., 2012). Consider the toy example shown in Fig. 3. The original stimulus displays (top row) from Chetverikov et al. (2017a) were reconstructed using the texture synthesis model (Portilla and Simoncelli, 2000). The resulting textures (middle row) are remarkably similar to the original displays, and the statistics of orientation distributions are largely intact although some finer details might be lost. The critical point here is that this shows that a statistics-based representation potentially used in texture perception might suffice for distribution encoding. So it should be relatively easy for the visual system to represent the distribution even without

**FIG. 3**

Resynthesis of the displays used for distribution learning (Chetverikov et al., 2017a) with the Portilla and Simoncelli (2000) texture synthesis algorithm. The original displays (*top row*) look similar to the displays reconstructed by iteratively adjusting a number of statistics on five spatial scales (*middle row*). Distribution of orientations in particular (*bottom row*) is relatively intact after the reconstruction: the target (item oriented at approximately 120 degrees) is still present, the number of modes is the same as in the original display. This shows that statistical representations potentially used by our perceptual system even in analysis of peripheral stimuli may suffice to obtain an accurate representation of a distribution.

applying attention to the periphery. The particular texture synthesis algorithm used in this example operates on different scales and does not involve any assumptions about the size of feature pooling regions. However, the Texture Tiling Model assumes that peripheral vision uses small local samples (Rosenholtz et al., 2012). As argued earlier, distribution encoding probably relies more on global processing. Feature

integration with larger pooling regions (as, for example, in the GIST model, [Oliva and Torralba, 2001](#); [Ross and Oliva, 2010](#)) might therefore better describe the mechanisms of distribution encoding.

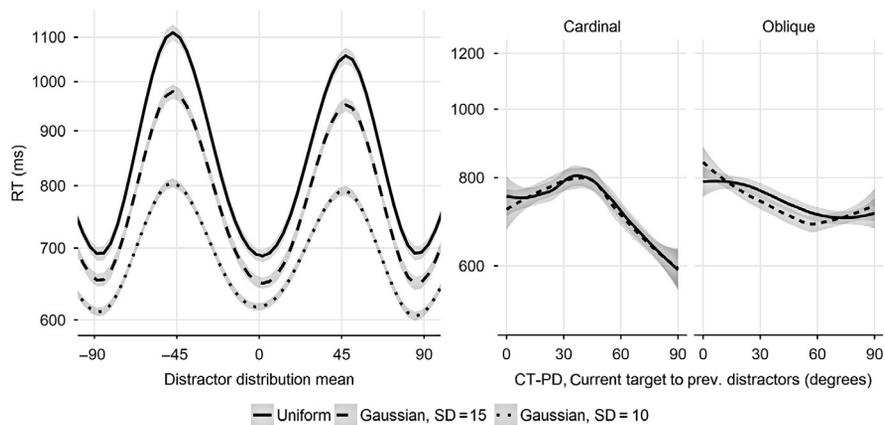
---

## 4 SAMPLING AND UPDATING INFORMATION ABOUT DISTRIBUTION

Observers can obtain information about feature distributions by sampling features from different locations or by aggregating information over whole stimulus sets. [Dakin \(2001, 2015\)](#) suggested that for orientation *averaging*, spatial arrangement (e.g., the display density) does not matter and observers use  $\sqrt{N}$  samples (where  $N$  is the number of stimuli) in their estimates. Other studies gave even lower estimates for averaging in different feature domains ([Maule and Franklin, 2016](#); [Myczek and Simons, 2008](#); [Solomon et al., 2016](#); [Tibber et al., 2015](#)). These estimates often rely on variance summation models where sampling strategies are assumed to be independent of noise (internal, i.e., from stimulus processing by the perceptual system and external, i.e., from the stimuli themselves). The assumptions of such models have recently been called into question, however ([Allard and Cavanagh, 2012](#); [Bocheva et al., 2015](#); but see [Dakin, 2015](#)). Moreover, the estimates of sampling efficiency change depending on the task (e.g., estimates of variance against estimates of mean; [Solomon, 2010](#)) and training ([Moerel et al., 2016](#)). It is not surprising then that studies not relying on the variance summation model generally agree that stimulus aggregation within a single ensemble occurs in parallel without strong processing limits ([Attarha and Moore, 2015](#); [Attarha et al., 2014](#); [Dakin, 2001](#); [Im and Halberda, 2013](#); [Robitaille and Harris, 2011](#); [Tokita et al., 2016](#); [Utochkin and Tiurina, 2014](#)).

Our own data suggest that information about distributions is unlikely to be obtained by sampling only a few elements ([Chetverikov et al., 2017b](#)). Moreover, the precision of distribution estimates in visual search grows with increased stimulus set sizes. With only a few elements ( $N=8$  or  $14$ ), there is little evidence that observers have any representations of distributions at all. With larger set sizes ( $N=16$  or  $24$ ), however, the distribution is encoded, but this encoding is imprecise as the information about shape (uniform vs Gaussian) is lost. Only with the largest set size ( $N=36$ ) did we find evidence of relatively precise distribution encoding ([Chetverikov et al., 2017b](#)).

It is likely that other factors, such as attention, noise levels, and task relevance, would affect the degree of approximation or precision of distribution encoding. Moreover, the shape of the distribution can be heavily influenced by existing priors. Humans are prone to biases in different domains, such as orientation, color, or motion perception, that can be explained by prior expectations arising from natural world statistics ([Girshick et al., 2011](#); [Long et al., 2006](#); [Purves and Lotto, 2003](#); [Seriès and Seitz, 2013](#); [Sotiropoulos et al., 2011](#); [Yang and Purves, 2004](#)). Such priors are likely to affect representations of distributions and their effects. For

**FIG. 4**

The effects of distractor distribution mean on search time (*left panel*) demonstrate the typical “oblique” effect. Search is faster when the distractor distribution is centered at cardinal orientations (around 0/180 and 90 degrees) compared to oblique ones. Previous studies suggest that this effect can be explained by decreased sensory uncertainty at cardinal compared to oblique orientations ([van Bergen et al., 2015](#)). In accordance with this logic, increasing stimulus heterogeneity increases the size of the oblique effect (cf. [Girshick et al., 2011](#)). Sensory uncertainty also affects distribution encoding (*right panel*; see [Fig. 1D](#) for comparison). The shape of the distribution is encoded with higher precision when it was centered at oblique orientations compared to cardinals (based on reanalysis of the data from experiment 2, 3A, and 3C, [Chetverikov et al., 2016](#)). Shaded areas show 95% confidence intervals.

example, sensory uncertainty is higher for oblique than cardinal orientations due to lower prior expectations ([van Bergen et al., 2015](#)). Accordingly, the heterogeneity of distributions has stronger effects at cardinal orientations ([Fig. 4](#), left panel; see also [Girshick et al., 2011](#)). Intuitively, this shows that priors have stronger effects when evidence is weak. Similarly, the effect of evidence should be stronger for weaker priors. Hence, the actual distribution shape will affect observers’ perception less strongly near-cardinal orientations than in between them as is indeed observed in our data ([Fig. 4](#), right panel).

## 5 SHORT-TERM AND LONG-TERM LEARNING OF DISTRIBUTIONS

Once observers obtain a representation of the distribution, they continue updating it as new information arrives. Studies of averaging over time indicate that observers integrate information over time, but the estimates may be biased toward more recent stimuli ([Albrecht and Scholl, 2010](#); [Attarha et al., 2016](#); [Corbett and Oriet, 2011](#);

Haberman et al., 2009; Hubert-Wallander and Boynton, 2015; Oriet and Hozempa, 2016). The aggregation happens even when it is irrelevant to the main task (Dubé and Sekuler, 2015; Oriet and Hozempa, 2016). Similarly, changes in distributions parameters (such as means and variance) from one trial to another increase processing time (Corbett and Melcher, 2014; Michael et al., 2014). The impact of previously perceived stimuli on the perception of the current one decreases with increased spatial distance or time between them (Fischer and Whitney, 2014).

We found that the amount of information about a distribution picked up on previous exposures to it (on preceding search trials in this case) is an important parameter for complex distributions (Chetverikov et al., 2017a). That is, the precision of estimates for a bimodal distribution increases with the number of repetitions in a visual search task, gradually changing from a unimodal to a bimodal representation. But for simpler unimodal distributions (such as Gaussian or uniform ones), the distribution shape can be determined with relative ease following only one or two trials (Chetverikov et al., 2017a). In other words, when the distribution is complex so that their initial representation does not have enough precision, observers update the existing representation with new data as it arrives.

It is not yet known whether and how such short-term changes relate to long-term ones. To the best of our knowledge, perceptual learning studies come closest to answering that question. Perceptual learning in its simplest variant is a special case of distribution learning when observers are presented with a distribution with very low variance (usually, a single reference value in each block). Observers' task is to learn to differentiate between this distribution and some other distribution with a gradually decreasing difference between them.<sup>a</sup> From a distribution-based point of view, the main difficulty of the task is to decrease the range within which the approximated PDF of the reference distribution has nonzero values; in other words, to shrink the represented distribution. The drawback of such learning is that values that are close to the reference value will gradually fall out the range of the approximated PDF, resulting in a decreased ability to discriminate between them.

Such an approach to perceptual learning is not new. Edelman and Intrator (2002) discuss different models of how observers can estimate the probability density associated with the underlying generator of data (e.g., the repeatedly presented stimulus). For example, one can assume that the density is smooth and then estimate it by adjusting the shape of underlying radial functions and/or the weights connecting these functions with output variables (Poggio and Girosi, 1990). When considered in this way, perceptual learning becomes essentially equivalent to learning probability distributions of features. In line with this idea, Chalk et al. (2010) found that observers trained to estimate motion direction on stimuli from a bimodal distribution (i.e., with two motion directions more probable than the rest) quickly begin to exhibit biases in accordance with biases in the stimulus distributions. Their perception

---

<sup>a</sup>Although there are different forms of perceptual learning, such as category learning, or learning within different modalities, here we only consider visual perceptual learning.

shifted toward more probable directions. Moreover, when no motion was present but observers thought that they had seen movement, their responses reproduced the probability distributions in the data. Importantly, using the same method, [Gekas et al. \(2013\)](#) demonstrated that with a bimodal distribution observers overestimate the probability of stimuli between the two modes of motion directions. That is, similar to what was found by [Chetverikov et al. \(2017a\)](#) observers' representations show evidence of interpolation. In these experiments, the probability of a stimulus with features between the two modes of a bimodal distribution was either very low ([Gekas et al., 2013](#)) or no stimuli with such features were presented ([Chetverikov et al., 2017a](#)). The perceptual system nevertheless represents the feature space between the modes as if some of the presented stimuli had feature values from that region. That is, rather than simply encoding highly probable features, it approximates the data (Section 2) with a probability distribution that might be less precise but potentially more useful when the true feature distribution is not known.

A potential challenge, however, to distribution-based explanations of perceptual learning is that perceptual learning has long been considered to be retinotopically specific ([Karni and Sagi, 1993](#)) or specific to the trained eye ([Fahle et al., 1995](#)), arguing for an early locus for these effects such as the V1. But according to other more recent findings, this alleged specificity may be a consequence of the actual paradigms used rather than a principle of perceptual learning ([Xiao et al., 2008](#); [Zhang et al., 2010](#)). For example, [Zhang et al. \(2010\)](#) showed that brief pretests at peripheral locations enabled transfer of foveal perceptual learning to the peripheral sites. This finding clearly argues against a retinotopic locus for perceptual learning, and these authors argue that the observed learning reflected activity modulation at higher-level sites in visual processing. This suggests that perceptual learning reflects rule-based learning where higher-level decision units reweigh V1 input.

Further evidence linking perceptual learning with probability distributions comes from investigations of the role of existing priors in perception. For example, most of the objects in the world are stationary or move at slow speeds, and motion speed perception is accordingly biased in favor of slow speeds ([Stocker and Simoncelli, 2006](#)). [Sotiropoulos et al. \(2011\)](#) showed that initial priors for slow motion change when observers repeatedly perceive high-speed movement. In other words, observers' representation of motion speed can be described as a probability distribution that is determined both by temporary influences of the momentary input, as well as priors from long-term learning.

An interesting new avenue of perceptual learning research is task-irrelevant perceptual learning of unattended features. [Seitz and Watanabe \(2005\)](#) found that a sub-threshold motion signal (5% coherence of an array of moving dots) presented in the background as observers performed a central RSVP task led to improved performance on a subsequent suprathreshold (10% coherence) motion direction discrimination task. A distribution-based view of perceptual learning accounts for this by assuming that observers can learn the characteristics of the distribution of the dots and that this influences subsequent motion direction discrimination.

---

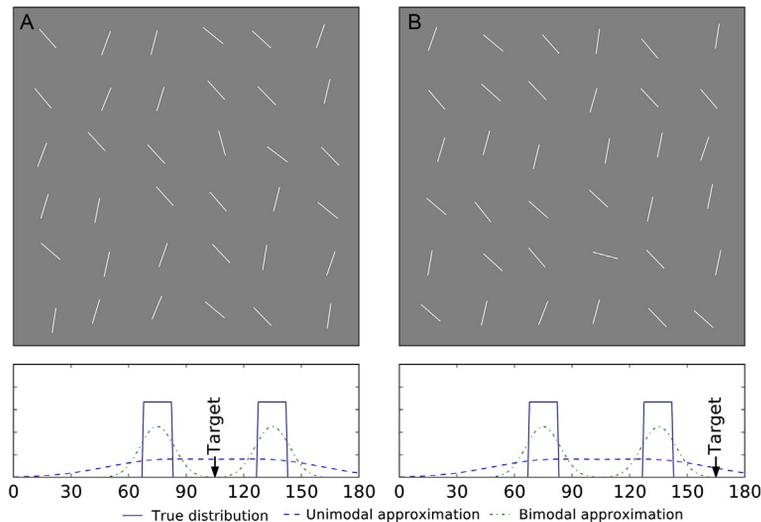
## 6 IMPLICATIONS FOR VISUAL ATTENTION AND VISUAL SEARCH

Visual search is one of the classic tasks used to investigate attention (Wolfe and Horowitz, 2017). It is well known that the heterogeneity of distractors is one of the major factors affecting search efficiency (Avraham et al., 2008; Banks and Prinzmetal, 1976; Bundesen and Pedersen, 1983; Duncan and Humphreys, 1989; Farmer and Taylor, 1980; Nagy and Thomas, 2003; Nagy et al., 2005; Utochkin and Yurevich, 2016). The more heterogeneous the distractor sets, the slower the search. Several models have been proposed to explain this, based on the idea of distractor grouping by similarity. For example, with the farthest-labeled nearest-neighbor algorithm (Avraham et al., 2008) the difficulty of search depends on the ratio of the distractor range to “target to nearest distractor distance” in feature space. That is, if distractors are similar (small distractor range) and the target significantly differs from all the distractors, search would be faster than if distractors are either different from each other or the target is similar to any of the distractors. Similarly, the best-normal approximation model suggested by Rosenholtz (2001) depends on the ratio of distractor variance and target distance to the mean of the distractors. In general, predictions within the proposed distribution-based framework are similar to the ones of existing models: the more similar the distractors are to one another and the less similar they are to targets, the faster the search. However, there is also a fundamental difference in the explanations of this rule. The distribution-based framework suggests that the difficulty of search depends on the estimated shape of the distribution and the probability of observing the target features in that distribution. This is akin to an extended signal-detection theory model also suggested by Rosenholtz (2001). The distribution-based model does, however, not assume that observers represent the distribution precisely according to the observed feature probabilities or use a specific model (e.g., Gaussian) to make inferences. In contrast, the distribution PDF is an approximate estimation with the possibility of under- or overfitting as discussed earlier (see Fig. 2). Recent results demonstrate that search based on approximations of the distribution can in fact be very efficient (Ma et al., 2011, 2015).

Search seems to be particularly difficult when a target cannot be easily separated from distractors in perceptual space. Consider the case of a bimodal distractor distribution with a target in between the modes, such as the one shown in Fig. 3. Search of this type is particularly difficult for observers in a number of different feature domains, such as for color, size, or orientation (Bauer et al., 1996b; D’Zmura, 1991; Hodson and Humphreys, 2001; Rosenholtz, 2001; Wolfe et al., 1992). Linear separability of target and distractor features is a common explanation of such effects: if a target can be separated from distractors by a single line in a feature space, the search will be easy, but otherwise hard (Arguin and Saumier, 2000; Bauer et al., 1996a,b; Nakayama and Martini, 2011; but see Vighneshvel and Arun, 2013). But an alternative interpretation provided by our distribution-based framework is based

on the finding that observers seem to approximate a bimodal distribution as unimodal (cf. [Utochkin and Yurevich, 2016](#)). In [Chetverikov et al. \(2017a, exp. 3\)](#), observers “filled in” the gap between modes and treated a bimodal distribution initially as uniform ([Fig. 5](#)). However, with increased trial numbers with distractors from the same bimodal distribution, where observers had more information about the distribution, their estimates became more precise and separation between the modes (in the CT-PD RT function) increased. We can then expect the exact approximation to also depend on the distractor heterogeneity so that the further away the two modes actually are from each other, the easier it will be to separate them.

Observers typically need a longer time to find a target as the number of distractors increases. But if a search task is very easy, the slope of the  $RT \times$  set size function should be flat instead of positive, according to the predictions of many conceptions of visual search (e.g., [Wolfe and Horowitz, 2017](#)). However, when target color is unpredictable, negative slopes are often observed ([Bacon and Egeth, 1991](#); [Bravo](#)



**FIG. 5**

(A) Linearly nonseparable (LNS) and (B) linearly separable (LS) displays. *Top rows* show example stimuli displays, while *bottom rows* show schematic representation of stimuli distributions. With LNS displays the target (column 4, row 4 from the *bottom-left*) is located between the two peaks of the distractor distribution. In contrast, with LS displays the target (column 4, row 2 from the *bottom-left*) is further away from distribution mean than either mode. If observers use a crude unimodal approximation, the target feature can be represented as part of distractor distribution with higher probability in the case of LNS displays compared to LS displays. However, a more accurate approximation (that could be achieved with repetitions of the same distractor distribution) will result in decreased distractor probability density both between and outside the modes. This could potentially make separable and nonseparable displays more comparable in difficulty.

and Nakayama, 1992; Kempgens et al., 2013; Kristjánsson, 2015; Santhi and Reeves, 2004; Schubö et al., 2007; Utochkin, 2013) providing a bit of a puzzle for traditional visual search models. Notably, despite suggestions that this effect is determined by local saliency (Wolfe, 2016), studies, where stimulus density was controlled for, suggest that local saliency is only partly responsible (Bacon and Egeth, 1991; Utochkin, 2013). Utochkin (2013) argued that increased efficiency with larger set sizes reflects increased “statistical power” of a comparison between the distractors subset size and a target subset that has a size equal to one (that is, a singleton subset). A variance reduction for estimates of subset summary statistics was suggested as a complementary mechanism. A distribution-based explanation makes predictions similar to the ones of a variance reduction explanation: with more data, the estimate of the distribution shape becomes more precise. However, in contrast to a variance reduction explanation, we expect that not only will RTs decrease with increased set size, but distribution shapes will also be estimated more precisely. This will in turn lead to differences in search times depending on distribution shape. Even when two distributions have the same range or variance, their representation will differ depending on their shape and hence the difficulty of finding an odd-one-out target will vary as well. Partial support for these predictions was provided by recent findings where we found that not only the precision of distribution shape estimates increases with set size but also that depending on distribution shape set size can sometime have different effects on search efficiency (Chetverikov et al., 2017b).

Attentional guidance in visual search also benefits from repetition. There seems to be a growing consensus that the dramatic effects seen in attentional priming in visual search tasks reflect activity modulations at various stages of the perceptual hierarchy (Kristjánsson and Campana, 2010). In the empirical work that supports the distribution-based framework that we propose here (Chetverikov et al., 2016, 2017a,b,c) such priming is used as evidence of distribution-based encoding of ensembles. The typical priming paradigm that involves a single target among identical distractors can be thought of as a truncated case of distribution-based encoding. Interestingly, in experiments with varying set size (Chetverikov et al., 2017b) we found that both priming effects (effects of changes in target and mean distractor orientations on search times) and the learning of distribution shape become more pronounced with larger set sizes (cf. Becker and Ansorge, 2013). Importantly, though, priming effects could occur even without distribution shape learning, suggesting that the two effects partly involve different distribution approximation levels.

The distributional account also poses the question of where attention will be guided within heterogeneous distributions of features. The idea of using distribution parameters for guiding vision has repeatedly been suggested in the summary statistics literature (Alvarez, 2011; Ariely, 2001; Chong and Treisman, 2003; Haberman and Whitney, 2012; Rosenholtz et al., 2012; Utochkin, 2015). Some results suggest, for example, that memory is biased toward the mean of an ensemble representation or that the mean has higher fidelity than the individual items (Alvarez and Oliva, 2008; Brady and Alvarez, 2011; Dubé and Sekuler, 2015; Oriet and Brand, 2013). Attention may therefore be biased toward the mean as well. But we suggest that

approximated distribution PDFs are represented rather than only means or standard deviations. According to our results, there is no preferential representation of the mean within distractor distributions. For example, when observers encoded uniform distributions, RTs on test trials were similar within the range of the encoded distribution (Chetverikov et al., 2016). That is, when a target feature was previously the mean of the distractor distribution, RTs were similar to when it had any other value within that distribution range. Crucially, this shows that observers did not particularly expect to encounter the mean above any other value within the distribution. If the mean were always automatically and preferentially encoded, role-reversal effects should be strongest for the mean, which was not the case. The main difference between the summary statistics approach and the distribution-based framework proposed here is that the latter does not suggest that summary properties of a distribution (e.g., its mean) are given any special treatment in perceptual processing. In contrast, we propose that those are properties that could be estimated if needed, but that otherwise vision is guided by distributions rather than their summary. If observers are explicitly asked about the mean, they can provide an answer, but it is not *specifically* represented. Accordingly, attention is unlikely to be specifically biased toward the means.

Apart from mean values, heterogeneous distributions, such as a Gaussian one, can include low-probability items—outliers, in other words. Studies on summary statistics show that during aggregation tasks (i.e., estimation of mean) observers discount the outliers (de Gardelle and Summerfield, 2011; Haberman and Whitney, 2010). On the other hand, it is well known that outliers attract visual attention (Treisman and Gelade, 1980). A question that could be addressed in future research involves whether observers (a) initially discount the outliers when encoding the heterogeneous distribution without an explicit aggregation task and (b) whether such outliers gradually stop attracting attention when observers learn the distractor distribution.

---

## 7 CONCLUSIONS

We propose a distribution-based framework that is based on one main idea: the visual system uses representations of probabilistic feature distributions to guide vision. This simple idea has a wide range of consequences, however. Most importantly, distributions need to be approximated. The same visual input can therefore easily give rise to different representations, and the estimated probability of one feature depends on probabilities of other features both in the past and in the present. Feature distributions seem to be sampled globally but might depend on segregation by location or other features. Here, studies on texture perception that assess global statistics or long-range interactions seem to be highly relevant for understanding distribution encoding. Once sampled, the distribution representation is updated with new information. While there are studies of short-term learning, little is known about long-term learning of complex distributions. However, perceptual learning studies provide examples

of long-term learning with simple distributions and show that such learning can change existing biases in different feature domains.

This framework has implications also for visual attention and visual search. Several phenomena—the effect of distractor heterogeneity and linear separability, negative search slopes, and attentional priming—can more easily be understood within a feature distribution-based framework than with more traditional approaches, such as guidance by local features and their contrasts. This approach also allows for novel predictions that can be tested in future research.

There are a number of open questions related to representations of probability distributions. For instance, to what extent does the precision of observers' representations depend on top-down attention and observers' goals? One might assume that the more relevant the particular features are, the higher the precision will be. On the other hand, it might be that additional attention will lead to segmentation of features into separate distributions. Additionally, does binding of features into objects hinder their perception as parts of the distribution? Do distributions of different features bind together? Can observers have different precision of distribution estimates in different regions of single feature domain? Surprisingly, little is known about representations of distributions as opposed to single features or their conjunctions. In general, most studies in domains such as visual search use simple homogenous displays. We believe that while useful, reliance on such displays is not a particularly ecologically valid approach. The visual world consists of feature distributions and visual representations reflect this.

---

## ACKNOWLEDGMENTS

This chapter was supported by the Russian Foundation for Humanities (#15-36-01358A2) and the Icelandic Research Fund (#152427).

---

## REFERENCES

- Albrecht, A.R., Scholl, B.J., 2010. Perceptually averaging in a continuous visual world: extracting statistical summary representations over time. *Psychol. Sci.* 21 (4), 560–567. <http://dx.doi.org/10.1177/0956797610363543>.
- Allard, R., Cavanagh, P., 2012. Different processing strategies underlie voluntary averaging in low and high noise. *J. Vis.* 12 (11), 1–12. <http://dx.doi.org/10.1167/12.11.6>.
- Alvarez, G.A., 2011. Representing multiple objects as an ensemble enhances visual cognition. *Trends Cogn. Sci.* 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. *Psychol. Sci.* 19 (4), 392–398. <http://dx.doi.org/10.1111/j.1467-9280.2008.02098.x>.
- Arguin, M., Saumier, D., 2000. Conjunction and linear non-separability effects in visual shape encoding. *Vision Res.* 40 (22), 3099–3115. [http://dx.doi.org/10.1016/S0042-6989\(00\)00155-3](http://dx.doi.org/10.1016/S0042-6989(00)00155-3).

- Ariely, D., 2001. Seeing sets: representation by statistical properties. *Psychol. Sci.* 12 (2), 157–162.
- Attarha, M., Moore, C.M., 2015. The capacity limitations of orientation summary statistics. *Atten. Percept. Psychophys.* 77 (4), 1116–1131. <http://dx.doi.org/10.3758/s13414-015-0870-0>.
- Attarha, M., Moore, C.M., Vecera, S.P., 2014. Summary statistics of size: fixed processing capacity for multiple ensembles but unlimited processing capacity for single ensembles. *J. Exp. Psychol. Hum. Percept. Perform.* 40 (4), 1440–1449. <http://dx.doi.org/10.1037/a0036206>.
- Attarha, M., Moore, C.M., Vecera, S.P., 2016. The time-limited visual statistician. *J. Exp. Psychol. Hum. Percept. Perform.* 42, 1497–1504. <http://dx.doi.org/10.1037/xhp0000255>.
- Avraham, T., Yeshurun, Y., Lindenbaum, M., Yeshurun, Y., Lindenbaum, M., 2008. Predicting visual search performance by quantifying stimuli similarities. *J. Vis.* 8, 9.1–22. <http://dx.doi.org/10.1167/8.4.9>.
- Bacon, W.F., Egeth, H.E., 1991. Local processes in preattentive feature detection. *J. Exp. Psychol. Hum. Percept. Perform.* 17 (1), 77–90. <http://dx.doi.org/10.1037/0096-1523.17.1.77>.
- Balas, B.J., Nakano, L., Rosenholtz, R., 2009. A summary-statistic representation in peripheral vision explains visual crowding. *J. Vis.* 9 (12), 13.1–18. <http://dx.doi.org/10.1167/9.12.13>.
- Banks, W.P., Prinzmetal, W., 1976. Configurational effects in visual information processing. *Percept. Psychophys.* 19 (4), 361–367. [http://dx.doi.org/10.1016/0009-2614\(82\)83587-2](http://dx.doi.org/10.1016/0009-2614(82)83587-2).
- Bauer, B., Jolicoeur, P., Cowan, W.B., 1996a. Distractor heterogeneity versus linear separability in colour visual search. *Perception* 25 (11), 1281–1293. <http://dx.doi.org/10.1068/p251281>.
- Bauer, B., Jolicoeur, P., Cowan, W.B., 1996b. Visual search for colour targets that are or are not linearly separable from distractors. *Vision Res.* 36 (10), 1439–1465. [http://dx.doi.org/10.1016/0042-6989\(95\)00207-3](http://dx.doi.org/10.1016/0042-6989(95)00207-3).
- Becker, S.I., Ansorge, U., 2013. Higher set sizes in pop-out search displays do not eliminate priming or enhance target selection. *Vision Res.* 81, 18–28. <http://dx.doi.org/10.1016/j.visres.2013.01.009>.
- Bocheva, N., Stefanov, S., Stefanova, M., Genova, B., 2015. Global orientation estimation in noisy conditions. *Acta Neurobiol. Exp.* 75 (4), 412–433.
- Brady, T.F., Alvarez, G.A., 2011. Hierarchical encoding in visual working memory ensemble statistics bias memory for individual items. *Psychol. Sci.* 22 (3), 384–392. <http://dx.doi.org/10.1177/0956797610397956>.
- Brand, J., Oriet, C., Tottenham, L.S., 2012. Size and emotion averaging: costs of dividing attention after all. *Can. J. Exp. Psychol.* 66 (1), 63–69. <http://dx.doi.org/10.1037/a0026950>.
- Bravo, M.J., Nakayama, K., 1992. The role of attention in different visual-search tasks. *Percept. Psychophys.* 51 (5), 465–472. <http://dx.doi.org/10.3758/BF03211642>.
- Bundesen, C., Pedersen, L.F., 1983. Color segregation and visual search. *Percept. Psychophys.* 33 (5), 487–493. <http://dx.doi.org/10.3758/BF03202901>.
- Chalk, M., Seitz, A.R., Seriès, P., 2010. Rapidly learned stimulus expectations alter perception of motion. *J. Vis.* 10 (8), 2. <http://dx.doi.org/10.1167/10.8.2>.
- Chang, H., Rosenholtz, R., 2016. Search performance is better predicted by tileability than presence of a unique basic feature. *J. Vis.* 16, 1–18. <http://dx.doi.org/10.1167/16.10.13.doi>.
- Chetverikov, A., Campana, G., Kristjánsson, Á., 2016. Building ensemble representations: how the shape of preceding distractor distributions affects visual search. *Cognition* 153, 196–210. <http://dx.doi.org/10.1016/j.cognition.2016.04.018>.

- Chetverikov, A., Campana, G., Kristjánsson, Á., 2017a. Rapid learning of visual ensembles. *J. Vis.* 17 (21), 1–15. <http://dx.doi.org/10.1167/17.2.21>.
- Chetverikov, A., Campana, G., & Kristjánsson, Á. 2017b. Set size manipulations reveal the boundary conditions of distractor distribution learning. Submitted for Review.
- Chetverikov, A., Campana, G., Kristjánsson, Á., 2017c. Representing color ensembles. *Psychol. Sci.* in press, <http://dx.doi.org/10.1177/0956797617713787>.
- Chong, S.C., Treisman, A., 2003. Representation of statistical properties. *Vision Res.* 43 (4), 393–404. [http://dx.doi.org/10.1016/S0042-6989\(02\)00596-5](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 Res.* 45 (7), 891–900. <http://dx.doi.org/10.1016/j.visres.2004.10.004>.
- Cohen, M.A., Dennett, D.C., Kanwisher, N., 2016. What is the bandwidth of perceptual experience? *Trends Cogn. Sci.* 20 (5), 324–335. <http://dx.doi.org/10.1016/j.tics.2016.03.006>.
- Coppola, D.M., Purves, H.R., McCoy, A.N., Purves, D., 1998. The distribution of oriented contours in the real world. *Proc. Natl. Acad. Sci. U.S.A.* 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. *J. Exp. Psychol. Hum. Percept. Perform.* 40 (5), 1915–1925. <http://dx.doi.org/10.1037/a0037375>.
- Corbett, J.E., Oriet, C., 2011. The whole is indeed more than the sum of its parts: perceptual averaging in the absence of individual item representation. *Acta Psychol. (Amst)* 138 (2), 289–301. <http://dx.doi.org/10.1016/j.actpsy.2011.08.002>.
- Dakin, S.C., 2001. Information limit on the spatial integration of local orientation signals. *J. Opt. Soc. Am. A* 18 (5), 1016–1026. <http://dx.doi.org/10.1364/JOSAA.18.001016>.
- Dakin, S.C., 2015. Seeing statistical regularities: texture and pattern perception. In: Wagemans, J. (Ed.), *The Oxford Handbook of Perceptual Organization*. Oxford University Press, Oxford, UK, pp. 150–167.
- Dakin, S.C., Watt, R.J., 1997. The computation of orientation statistics from visual texture. *Vision Res.* 37 (22), 3181–3192. [http://dx.doi.org/10.1016/S0042-6989\(97\)00133-8](http://dx.doi.org/10.1016/S0042-6989(97)00133-8).
- de Gardelle, V., Summerfield, C., 2011. Robust averaging during perceptual judgment. *Proc. Natl. Acad. Sci. U.S.A.* 108 (32), 13341–13346. <http://dx.doi.org/10.1073/pnas.1104517108>.
- Dubé, C., Sekuler, R., 2015. Obligatory and adaptive averaging in visual short-term memory. *J. Vis.* 15 (4), 13. <http://dx.doi.org/10.1167/15.4.13>.
- Duncan, J., Humphreys, G.W., 1989. Visual search and stimulus similarity. *Psychol. Rev.* 96 (3), 433–458.
- D’Zmura, M., 1991. Color in visual search. *Vision Res.* 31 (6), 951–966. [http://dx.doi.org/10.1016/0042-6989\(91\)90203-H](http://dx.doi.org/10.1016/0042-6989(91)90203-H).
- Edelman, S., Intrator, N., 2002. Models of perceptual learning. In: Fahle, M., Poggio, T. (Eds.), *Perceptual Learning*. MIT Press, Cambridge, MA, pp. 337–354.
- Ehinger, K.A., Rosenholtz, R., 2016. A general account of peripheral encoding also predicts scene perception performance. *J. Vis.* 16 (2), 13. <http://dx.doi.org/10.1167/16.2.13>.
- Emmanouil, T.A., Treisman, A., 2008. Dividing attention across feature dimensions in statistical processing of perceptual groups. *Percept. Psychophys.* 70 (6), 946–954. <http://dx.doi.org/10.3758/PP.70.6.946>.
- Fahle, M., Edelman, S., Poggio, T., 1995. Fast perceptual learning in hyperacuity. *Vision Res.* 35 (21), 3003–3013. [http://dx.doi.org/10.1016/0042-6989\(95\)00044-Z](http://dx.doi.org/10.1016/0042-6989(95)00044-Z).

- Farmer, E.W., Taylor, R.M., 1980. Visual search through color displays: effects of target-background similarity and background uniformity. *Percept. Psychophys.* 27 (3), 267–272. <http://dx.doi.org/10.3758/BF03204265>.
- Feldman, J., 2014. Probabilistic models of perceptual features. In: Wagemans, J. (Ed.), *Oxford Handbook of Perceptual Organization*. Oxford University Press, Oxford, UK, pp. 933–947. <http://dx.doi.org/10.1093/oxfordhb/9780199686858.013.049>.
- Feldman, H., Friston, K.J., 2010. Attention, uncertainty, and free-energy. *Front. Hum. Neurosci.* 4, 215. <http://dx.doi.org/10.3389/fnhum.2010.00215>.
- Feldmann-Wüstefeld, T., Schubö, A., 2013. Context homogeneity facilitates both distractor inhibition and target enhancement. *J. Vis.* 13 (3), 11. <http://dx.doi.org/10.1167/13.3.11>.
- Feldmann-Wüstefeld, T., Schubö, A., 2015. Target discrimination delays attentional benefit for grouped contexts: an ERP study. *Brain Res.* 1629, 196–209. <http://dx.doi.org/10.1016/j.brainres.2015.10.018>.
- Fischer, J., Whitney, D., 2014. Serial dependence in visual perception. *Nat. Neurosci.* 17 (5), 738–743. <http://dx.doi.org/10.1038/nn.3689>.
- Fiser, J., Berkes, P., Orbán, G., Lengyel, M., 2010. Statistically optimal perception and learning: from behavior to neural representations. *Trends Cogn. Sci.* 14, 119–130. <http://dx.doi.org/10.1016/j.tics.2010.01.003>.
- Gekas, N., Chalk, M., Seitz, A.R., Seriès, P., 2013. Complexity and specificity of experimentally-induced expectations in motion perception. *J. Vis.* 13 (4), 8. <http://dx.doi.org/10.1167/13.4.8>.
- Girshick, A.R., Landy, M.S., Simoncelli, E.P., 2011. Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nat. Neurosci.* 14 (7), 926–932. <http://dx.doi.org/10.1038/nn.2831>.
- Haberman, J., Whitney, D., 2010. The visual system discounts emotional deviants when extracting average expression. *Atten. Percept. Psychophys.* 72 (7), 1825–1838. <http://dx.doi.org/10.3758/APP.72.7.1825>.
- Haberman, J., Whitney, D., 2012. Ensemble perception: summarizing the scene and broadening the limits of visual processing. In: Wolfe, J.M., Robertson, L. (Eds.), *From Perception to Consciousness: Searching With Anne Treisman*. Oxford University Press, Oxford, pp. 339–349.
- Haberman, J., Harp, T., Whitney, D., 2009. Averaging facial expression over time. *J. Vis.* 9 (11), 1.1–13. <http://dx.doi.org/10.1167/9.11.1>.
- Hodsoll, J.P., Humphreys, G.W., 2001. Driving attention with the top down: the relative contribution of target templates to the linear separability effect in the size dimension. *Percept. Psychophys.* 63 (5), 918–926. <http://dx.doi.org/10.3758/BF03194447>.
- Hubert-Wallander, B., Boynton, G.M., 2015. Not all summary statistics are made equal: evidence from extracting summaries across time. *J. Vis.* 15 (4), 1–12. <http://dx.doi.org/10.1167/15.4.5>.
- Im, H.Y., Halberda, J., 2013. The effects of sampling and internal noise on the representation of ensemble average size. *Atten. Percept. Psychophys.* 75 (2), 278–286. <http://dx.doi.org/10.3758/s13414-012-0399-4>.
- Julesz, B., 1981. Textons, the elements of texture perception, and their interactions. *Nature* 290 (5802), 91–97. <http://dx.doi.org/10.1038/290091a0>.
- Karni, A., Sagi, D., 1993. The time course of learning a visual skill. *Nature* 365 (6443), 250–252. <http://dx.doi.org/10.1038/365250a0>.

- Kempgens, C., Loffler, G., Orbach, H.S., 2013. Set-size effects for sampled shapes: experiments and model. *Front. Comput. Neurosci.* 7, 67. <http://dx.doi.org/10.3389/fncom.2013.00067>.
- Kristjánsson, Á., 2015. Reconsidering visual search. *I-Perception* 6 (6), 1–14. <http://dx.doi.org/10.1177/2041669515614670>.
- Kristjánsson, Á., Campana, G., 2010. Where perception meets memory: a review of repetition priming in visual search tasks. *Atten. Percept. Psychophys.* 72 (1), 5–18. <http://dx.doi.org/10.3758/APP.72.1.5>.
- Kristjánsson, Á., Driver, J., 2005. Priming in vision: target repetition effects, context effects and role-reversal effects. *Perception* 34 (Suppl), 40c.
- Kristjánsson, Á., Driver, J., 2008. Priming in visual search: separating the effects of target repetition, distractor repetition and role-reversal. *Vision Res.* 48 (10), 1217–1232. <http://dx.doi.org/10.1016/j.visres.2008.02.007>.
- 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 Res.* 48 (1), 30–41. <http://dx.doi.org/10.1016/j.visres.2007.10.009>.
- Landy, M.S., 2013. Texture analysis and perception. In: Werner, J.S., Chalupa, L.M. (Eds.), *The New Visual Neurosciences*. MIT Press, Cambridge, MA, pp. 639–652.
- Long, F., Yang, Z., Purves, D., 2006. Spectral statistics in natural scenes predict hue, saturation, and brightness. *Proc. Natl. Acad. Sci. U.S.A.* 103 (15), 6013–6018. <http://dx.doi.org/10.1073/pnas.0600890103>.
- Ma, W.J., 2012. Organizing probabilistic models of perception. *Trends Cogn. Sci.* 16 (10), 511–518. <http://dx.doi.org/10.1016/j.tics.2012.08.010>.
- Ma, W.J., Navalpakkam, V., Beck, J.M., Van Den Berg, R., Pouget, A., 2011. Behavior and neural basis of near-optimal visual search. *Nat. Neurosci.* 14 (6), 783–790. <http://dx.doi.org/10.1038/nn.2814>.
- Ma, W.J., Shen, S., Dziugaite, G., van den Berg, R., 2015. Requiem for the max rule? *Vision Res.* 116, 179–193. <http://dx.doi.org/10.1016/j.visres.2014.12.019>.
- Maljkovic, V., Nakayama, K., 1994. Priming of pop-out: I. Role of features. *Mem. Cognit.* 22 (6), 657–672.
- Maule, J., Franklin, A., 2016. Accurate rapid averaging of multihue ensembles is due to a limited capacity subsampling mechanism. *J. Opt. Soc. Am. A* 33 (3), A22–29. <http://dx.doi.org/10.1364/JOSAA.33.000A22>.
- Mazza, V., Turatto, M., Caramazza, A., 2009. Attention selection, distractor suppression and N2pc. *Cortex* 45 (7), 879–890. <http://dx.doi.org/10.1016/j.cortex.2008.10.009>.
- Meinecke, C., Donk, M., 2002. Detection performance in pop-out tasks: nonmonotonic changes with display size and eccentricity. *Perception* 31 (5), 591–602. <http://dx.doi.org/10.1068/p3201>.
- 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>.
- Moerel, D., Ling, S., Jehee, J.F.M., 2016. Perceptual learning increases orientation sampling efficiency. *J. Vis.* 16 (3), 36. 1–9, <http://dx.doi.org/10.1167/16.3.36>.
- Myczek, K., Simons, D.J., 2008. Better than average: alternatives to statistical summary representations for rapid judgments of average size. *Percept. Psychophys.* 70 (5), 772–788. <http://dx.doi.org/10.3758/PP.70.5.772>.
- Nagy, A.L., Thomas, G., 2003. Distractor heterogeneity, attention, and color in visual search. *Vision Res.* 43 (14), 1541–1552. [http://dx.doi.org/10.1016/S0042-6989\(03\)00234-7](http://dx.doi.org/10.1016/S0042-6989(03)00234-7).

- Nagy, A.L., Neriani, K.E., Young, T.L., 2005. Effects of target and distractor heterogeneity on search for a color target. *Vision Res.* 45, 1885–1899. <http://dx.doi.org/10.1016/j.visres.2005.01.007>.
- Nakayama, K., Martini, P., 2011. Situating visual search. *Vision Res.* 51 (13), 1526–1537. <http://dx.doi.org/10.1016/j.visres.2010.09.003>.
- Nothdurft, H.C., 2000. Saliency from feature contrast: variations with texture density. *Vision Res.* 40 (23), 3181–3200. [http://dx.doi.org/10.1016/S0042-6989\(00\)00168-1](http://dx.doi.org/10.1016/S0042-6989(00)00168-1).
- Oliva, A., Torralba, A., 2001. Modeling the shape of the scene: a holistic representation of the spatial envelope. *Int. J. Comput. Vis.* 42 (3), 145–175. <http://dx.doi.org/10.1023/A:1011139631724>.
- Oriet, C., Brand, J., 2013. Size averaging of irrelevant stimuli cannot be prevented. *Vision Res.* 79, 8–16. <http://dx.doi.org/10.1016/j.visres.2012.12.004>.
- Oriet, C., Hozempa, K., 2016. Incidental statistical summary representation over time. *J. Vis.* 16 (3), 1–14. <http://dx.doi.org/10.1167/16.3.3.doi>.
- Parkes, L., Lund, J., Angelucci, A., Solomon, J.A., Morgan, M.J., 2001. Compulsory averaging of crowded orientation signals in human vision. *Nat. Neurosci.* 4 (7), 739–744. <http://dx.doi.org/10.1038/89532>.
- Pöder, E., 2006. Crowding, feature integration, and two kinds of “attention”. *J. Vis.* 6 (2006), 163–169. <http://dx.doi.org/10.1167/6.2.7>.
- Poggio, T., Girosi, F., 1990. Regularization algorithms for learning that are equivalent to multilayer networks. *Science* 247 (4945), 978–982. <http://dx.doi.org/10.1126/science.247.4945.978>.
- Portilla, J., Simoncelli, E.P., 2000. Parametric texture model based on joint statistics of complex wavelet coefficients. *Int. J. Comput. Vis.* 40 (1), 49–71. <http://dx.doi.org/10.1023/A:1026553619983>.
- Purves, D., Lotto, R.B., 2003. *Why We See What We Do: An Empirical Theory of Vision*. Sinauer Associates, Sunderland, MA.
- Rao, R.P., Olshausen, B.A., Lewicki, M.S., 2002. *Probabilistic Models of the Brain: Perception and Neural Function*. MIT Press, Cambridge, MA, UK.
- Robitaille, N., Harris, I.M., 2011. When more is less: extraction of summary statistics benefits from larger sets. *J. Vis.* 11 (12), 18. <http://dx.doi.org/10.1167/11.12.18>.
- Rosenholtz, R., 2001. Visual search for orientation among heterogeneous distractors: experimental results and implications for signal-detection theory models of search. *J. Exp. Psychol. Hum. Percept. Perform.* 27 (4), 985–999. <http://dx.doi.org/10.1037/0096-1523.27.4.985>.
- Rosenholtz, R., 2014. Texture perception. In: Wagemans, J. (Ed.), *Oxford Handbook of Perceptual Organization*. Oxford University Press, Oxford, UK, pp. 1–24. <http://dx.doi.org/10.1093/oxfordhb/9780199686858.013.058>.
- Rosenholtz, R., Huang, J., Raj, A., Balas, B.J., Ilie, L., 2012. A summary statistic representation in peripheral vision explains visual search. *J. Vis.* 12 (4), 14. <http://dx.doi.org/10.1167/12.4.14>.
- Ross, M.G., Oliva, A., 2010. Estimating perception of scene layout properties from global image features. *J. Vis.* 10 (1), 2.1–25. <http://dx.doi.org/10.1167/10.1.2>.
- Sagi, D., Julesz, B., 1987. Short-range limitation on detection of feature differences. *Spat. Vis.* 2 (1), 39–49. <http://dx.doi.org/10.1163/156856887X00042>.
- Santhi, N., Reeves, A., 2004. The roles of distractor noise and target certainty in search: a signal detection model. *Vision Res.* 44 (12), 1235–1256. <http://dx.doi.org/10.1016/j.visres.2003.11.011>.

- Schubö, A., Wykowska, A., Müller, H.J., 2007. Detecting pop-out targets in contexts of varying homogeneity: investigating homogeneity coding with event-related brain potentials (ERPs). *Brain Res.* 1138 (1), 136–147. <http://dx.doi.org/10.1016/j.brainres.2006.12.059>.
- Seitz, A.R., Watanabe, T., 2005. A unified model for perceptual learning. *Trends Cogn. Sci.* 9 (7), 329–334. <http://dx.doi.org/10.1016/j.tics.2005.05.010>.
- Seriès, P., Seitz, A.R., 2013. Learning what to expect (in visual perception). *Front. Hum. Neurosci.* 7, 668. <http://dx.doi.org/10.3389/fnhum.2013.00668>.
- Sheather, S.J., Jones, M.C., 1991. A reliable data based bandwidth selection method for kernel density estimation. *J. R. Stat. Soc. Ser. B Methodol.* 53, 683–690. <http://dx.doi.org/10.2307/2345597>.
- Solomon, J.A., 2010. Visual discrimination of orientation statistics in crowded and uncrowded arrays. *J. Vis.* 10 (14), 19. <http://dx.doi.org/10.1167/10.14.19>.
- Solomon, J.A., May, K.A., Tyler, C.W., 2016. Inefficiency of orientation averaging: evidence for hybrid serial/parallel temporal integration. *J. Vis.* 16 (1), 13. <http://dx.doi.org/10.1167/16.1.13>.
- Sotiropoulos, G., Seitz, A.R., Seriès, P., 2011. Changing expectations about speed alters perceived motion direction. *Curr. Biol.* 21 (21), R883–R884. <http://dx.doi.org/10.1016/j.cub.2011.09.013>.
- Stocker, A.A., Simoncelli, E.P., 2006. Noise characteristics and prior expectations in human visual speed perception. *Nat. Neurosci.* 9 (4), 578–585. <http://dx.doi.org/10.1038/nn1669>.
- Tibber, M.S., Anderson, E.J., Bobin, T., Carlin, P., Shergill, S.S., Dakin, S.C., 2015. Local and global limits on visual processing in schizophrenia. *PLoS One* 10 (2), 1–17. <http://dx.doi.org/10.1371/journal.pone.0117951>.
- Tokita, M., Ueda, S., Ishiguchi, A., 2016. Evidence for a global sampling process in extraction of summary statistics of item sizes in a set. *Front. Psychol.* 7, 711. <http://dx.doi.org/10.3389/fpsyg.2016.00711>.
- Treisman, A.M., Gelade, G., 1980. A feature-integration theory of attention. *Cogn. Psychol.* 12 (1), 97–136. [http://dx.doi.org/10.1016/0010-0285\(80\)90005-5](http://dx.doi.org/10.1016/0010-0285(80)90005-5).
- Utochkin, I.S., 2013. Visual search with negative slopes: the statistical power of numerosity guides attention. *J. Vis.* 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. *J. Vis.* 15 (4), 8. <http://dx.doi.org/10.1167/15.4.8>.
- 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 Psychol. (Amst)* 146, 7–18. <http://dx.doi.org/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”. *J. Exp. Psychol. Hum. Percept. Perform.* 42 (7), 995–1007. <http://dx.doi.org/10.1037/xhp0000203>.
- van Bergen, R.S., Ma, W.J., Pratte, M.S., Jehhee, J.F.M., 2015. Sensory uncertainty decoded from visual cortex predicts behavior. *Nat. Neurosci.* 18 (12), 1728–1730. <http://dx.doi.org/10.1038/nn.4150>.
- Vighneshvel, T., Arun, S.P., 2013. Does linear separability really matter? Complex visual search is explained by simple search. *J. Vis.* 13 (11), 1–24. <http://dx.doi.org/10.1167/13.11.10>.
- Vincent, B.T., Baddeley, R.J., Troscianko, T., Gilchrist, I.D., 2009. Optimal feature integration in visual search. *J. Vis.* 9 (5), 1–11. <http://dx.doi.org/10.1167/9.5.15>.

- Whiting, B.F., Oriet, C., 2011. Rapid averaging? Not so fast! *Psychon. Bull. Rev.* 18 (3), 484–489. <http://dx.doi.org/10.3758/s13423-011-0071-3>.
- Wolfe, J.M., 1992. “Effortless” texture segmentation and “parallel” visual search are not the same thing. *Vision Res.* 32 (4), 757–763. [http://dx.doi.org/10.1016/0042-6989\(92\)90190-T](http://dx.doi.org/10.1016/0042-6989(92)90190-T).
- Wolfe, J.M., 2016. Visual search revived: the slopes are not that slippery: a reply to Kristjansson (2015). *I-Perception* 7 (3), 1–6. <http://dx.doi.org/10.1177/2041669516643244>.
- Wolfe, J.M., Horowitz, T.S., 2017. Five factors that guide attention in visual search. *Nat. Hum. Behav.* 1 (3), 58. <http://dx.doi.org/10.1038/s41562-017-0058>.
- Wolfe, J.M., Friedman-Hill, S.R., Stewart, M.I., O’Connell, K.M., 1992. The role of categorization in visual search for orientation. *J. Exp. Psychol. Hum. Percept. Perform.* 18 (1), 34–49. <http://dx.doi.org/10.1037/0096-1523.18.1.34>.
- Wolfson, S.S., Landy, M.S., 1998. Examining edge- and region-based texture analysis mechanisms. *Vision Res.* 38 (3), 439–446. [http://dx.doi.org/10.1016/S0042-6989\(97\)00153-3](http://dx.doi.org/10.1016/S0042-6989(97)00153-3).
- Xiao, L.Q., Zhang, J.Y., Wang, R., Klein, S.A., Levi, D.M., Yu, C., 2008. Complete transfer of perceptual learning across retinal locations enabled by double training. *Curr. Biol.* 18 (24), 1922–1926. <http://dx.doi.org/10.1016/j.cub.2008.10.030>.
- Yang, Z., Purves, D., 2004. The statistical structure of natural light patterns determines perceived light intensity. *Proc. Natl. Acad. Sci. U.S.A.* 101 (23), 8745–8750. <http://dx.doi.org/10.1073/pnas.0402192101>.
- Yuille, A., Kersten, D., 2006. Vision as Bayesian Inference: Analysis by Synthesis? *Trends Cogn. Sci.* 10 (7), 301–308. <https://doi.org/10.1016/j.tics.2006.05.002>.
- Zhang, J.-Y., Zhang, G.-L., Xiao, L.-Q., Klein, S.A., Levi, D.M., Yu, C., 2010. Rule-based learning explains visual perceptual learning and its specificity and transfer. *J. Neurosci.* 30 (37), 12323–12328. <http://dx.doi.org/10.1523/JNEUROSCI.0704-10.2010>.