Vision Research 140 (2017) 144–156

Contents lists available at ScienceDirect

Vision Research

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

Set size manipulations reveal the boundary conditions of perceptual ensemble learning

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

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

ARTICLE INFO

Keywords:
Feature distributions
Summary statistics
Ensemble perception
Visual search
Priming of pop-out

ABSTRACT

Recent evidence suggests that observers can grasp patterns of feature variations in the environment with surprising efficiency. During visual search tasks where all distractors are randomly drawn from a certain distribution rather than all being homogeneous, observers are capable of learning highly complex statistical properties of distractor sets. After only a few trials (learning phase), the statistical properties of distributions - mean, variance and crucially, shape - can be learned, and these representations affect search during a subsequent test phase (Chetverikov, Campana, & Kristjánsson, 2016). To assess the limits of such distribution learning, we varied the information available to observers about the underlying distractor distributions by manipulating set size during the learning phase in two experiments. We found that robust distribution learning only occurred for large set sizes. We also used set size to assess whether the learning of distribution properties makes search more efficient. The results reveal how a certain minimum of information is required for learning to occur, thereby delineating the boundary conditions of learning of statistical variation in the environment. However, the benefits of distribution learning for search efficiency remain unclear.

How do observers represent the variation in the environment such as the colors in a moss-covered lava field or the brightness distribution in snow covered landscapes? Although we may think of moss as "green" and snow as "white", we clearly perceive more than a single feature value. On the other hand, encoding every feature at every location along with their conjunctions will require a lot of resources. The question is then how feature variation in the external world is translated into a representation, and the answer will likely be somewhere between the two extremes outlined above. Processing of such heterogeneous perceptual ensembles has been studied with texture segregation tasks (Julesz, 1981) but natural sets are typically not as regular as those studied by Julesz. Take color – color variation in natural environments is rarely uniform – and neither are the oriented edges available in natural statistical distributions. There is accumulating evidence that human observers can extract summary statistics such as the mean and standard deviation of a number of features, such as color, size, orientation and brightness, from stimulus sets having a certain variability (Alvarez, 2011; Ariely, 2001; Corbett & Melcher, 2014; Haberman & Whitney, 2012; Michael, de Gardelle, & Summerfield, 2014; Rosenholtz, Huang, Raj, Balas, & Ilie, 2012; Utochkin, 2015). Summary statistics provide a concise way of representing feature variation but they are still relatively coarse because two different ensembles might have the same statistics while coming from different distributions.

Our recent experiments have revealed that observers can represent more intricate feature variation than studies of simple statistical parameters have suggested. Chetverikov, Campana, and Kristjánsson (2017b) showed that after only a few trials observers can learn the properties of feature distributions of colored distractors in an odd-one-out visual search task over and above the mean and standard deviations, and in Chetverikov, Campana, and Kristjánsson (2016, 2017a), we found similar results for orientation. In those studies, assessed observers' representations by measuring their implicit expectations of upcoming stimulus distributions with response times (RTs) instead of explicit judgements of distribution properties. Namely, we measured effects of 'role-reversals' between targets and distractors on visual search performance (Kristjánsson & Driver, 2008). A role-reversal occurs when a target on a preceding trial becomes a distractor on the next trial, or vice-versa, which typically slows search (Becker, 2010). This effect is not limited to specific feature domains and seems to reflect

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

http://dx.doi.org/10.1016/j.visres.2017.08.003
Received 12 April 2017; Received in revised form 11 August 2017; Accepted 20 August 2017

encoding of distractors in implicit short-term visual working memory (Carlisle & Kristjánsson, 2017; Lamy, Antebi, Aviani, & Carmel, 2008; Maljkovic & Nakayama, 1994). In a typical role-reversal study, the distractors are homogeneous. For example, in a color search observers would look for a red target among green distractors. After a few trials with repeated distractor colors observers encode the distractor features, and when the targets become green (among distractors of some other color), search is slowed. The key difference in our manipulation relative to previous studies was that distractors were heterogeneous and on a single trial their features were randomly drawn from a specific probability distribution. The distractors, in other words, formed a perceptual ensemble. Continuing with the example above, instead of a red target among green distractors, observers had to search for a red target among distractors of varying degrees of "greenness", akin to searching for a red berry within moss patches. Then, as these conditions were repeated for a few trials, a role-reversal to a greenish target resulted in slowed search. Importantly, the degree of slowing depended on the correspondence between target hue and the probability of that particular hue among previous distractors. This allowed us to assess observers' representations of ensembles encoded on previous search trials.

We assumed, in other words, that if a target falls within observers' representations of preceding distractor distributions it would cause role-reversal effects, that is, search would be slower because the features of the odd-one-out target would clash with representations of distractor distributions from previous trials (Chetverikov et al., 2016, 2017a,b). Using targets corresponding to different parts of previously learned distractor distributions allows us to infer the probabilistic representation of that distribution by assessing how much search is slowed. For example, following several odd-one-out search trials in the orientation domain with distractors drawn from a truncated Gaussian distribution with an orientation $\mu = 45^\circ$ and $\sigma = 15^\circ$ (range restricted to $45 \pm 30^\circ$), observers respond more slowly when a 45° odd-one-out target suddenly appears than when a 40° target appears, which, in turn, will yield longer response times than a 35° target, and search will be fastest for targets that fall outside the range of the previous distractor distribution. The search RTs will therefore be slow if observers expected this orientation to be from the distractor distribution of immediately preceding trials. The degree of slowing reflects encoded feature probability. By repeating blocks of learning and test trials with different test targets, we were able to "probe" observers' representations of feature distributions along the whole range of possible feature values and obtain detailed *continuous* estimates of these representations.

Importantly, we also found that even when two distributions have the same range or variance, observers' representations differ (Chetverikov et al., 2016). So in contrast to a Gaussian distribution, following learning of a uniform distribution with the same 45° mean and ± 30° range, response times (RTs) for any target within $45 \pm 30^\circ$ degrees will be approximately the same. That is, even the shapes of the distributions (e.g., whether they are Gaussian, uniform, skewed or even bimodal) are encoded (Chetverikov et al., 2016, 2017a). Differences in the estimates for differently shaped distributions suggest that the precision of ensemble perception is much higher than was thought before.

1. Mechanisms of ensemble perception

How do observers obtain such precise ensemble representations from the stimuli presented on the screen? Recent studies involving explicit summary statistic judgements indicate that the aggregation is limited by the number of stimulus subsets rather than the number of stimuli within a subset (Attarha, Moore, & Vecera, 2014; Im & Halberda, 2013; Maule & Franklin, 2015; Utochkin & Tiurina, 2014; Utochkin & Yurevich, 2016). But the exact mechanisms of aggregation *within* subsets remain controversial. Several studies support the idea of limited sampling (Maule & Franklin, 2016; Myczek & Simons, 2008; Solomon, May, & Tyler, 2016; Tibber et al., 2015) with the number of sampled stimuli being below four. That is, observers can respond accurately when asked about summary statistics even if they analyse only a few exemplars from the stimulus set. Others have argued against this, however (Attarha & Moore, 2015; Attarha et al., 2014; Dakin, 2001; Im & Halberda, 2013; Tokita, Ueda, & Ishiguchi, 2016; Utochkin & Tiurina, 2014). Moreover, approximations involved in *explicit* averaging may differ from tasks where the use of statistics is not explicitly required but might nevertheless be useful or even necessary. Such tasks may include visual search (Rosenholtz et al., 2012), visual categorization (Utochkin, 2015), attentional selection (Im, Park, & Chong, 2015), or texture perception (Dakin, 2015). In particular, distribution learning in visual search (Chetverikov et al. (2016) is not required by the task and therefore allows the study of mechanisms involved in incidental use of summary statistics.

The use of explicit judgments about the properties of feature distributions in previous studies limits our understanding of the mechanisms leading to ensemble representations. It is possible that potential bottlenecks on the precision of such explicit judgments have little to do with distribution representations *per se*. There are a number of ways in which even if observers have highly precise representations of distributions, explicit judgments will still rely on only a few samples. For example, observers might use their representation to generate a limited sample for explicit judgements. That is, when asked to judge the mean, observers might simply sample the distributions they saw. Another option is that, during an averaging task, observers' might try to hold in working memory only the stimuli useful for the averaging they are asked to perform. Using tasks with incidental encoding would be helpful to understand whether limitations found in some studies for explicit averaging are related to ensemble encoding or simply reflect the use of explicit judgements.

Regardless of the mechanisms underlying explicit averaging, incidental distribution encoding within the present paradigm is of interest by itself. Previous results indicate that distribution representations that observers use in visual search are more precise than, for example, those that can be derived from forced-choice judgements (see review in Chetverikov et al., 2016). How this higher precision is obtained is an interesting question in and of itself, one we investigate here.

Our previous results indicate that distribution learning in visual search can occur rapidly (Chetverikov et al., 2017a). Sometimes only two trials seem to be needed to learn simpler distributions, while learning a more complex (bimodal) distribution required a larger number of search trials and involved a gradual change from a unimodal to a bimodal representation. This shows that distribution representations can be based on the accumulation of information coming from multiple samples – otherwise the representation would be the same regardless of trial number. But how many samples are needed from a single display is unknown. For example, on a given trial observers may sample three or four items (Maule & Franklin, 2016; Myczek & Simons, 2008; Solomon et al., 2016) and then integrate the samples from different trials.

Here we used set size manipulations to investigate the limits on processing of simultaneously presented information during feature distribution learning. If learning of distribution parameters is based on a few stimuli sampled from each trial, the learning should be equally efficient with small and large set sizes. On the other hand, if the learning is based on an aggregation (possibly, in parallel) of a large number of stimuli, larger set sizes would result in better learning.

2. Role of set size for search efficiency and inter-trial priming

Set size manipulations have played a key role in theories of visual attention. For easier searches where the target is easily found among distractors, RTs are constant with set size, or can even decrease (Bravo & Nakayama, 1992; Kristjánsson, 2015; Wang, Kristjánsson, & Nakayama, 2005; Wolfe & Horowitz, 2017). Using this classic manipulation may therefore also reveal whether and how distractor distribution learning affects search performance more generally.