Prior experience informs ensemble encoding

L. Elizabeth Crawford¹ · Jonathan C. Corbin¹ · David Landy²

Published online: 16 November 2018
© Psychonomic Society, Inc. 2018

Abstract

People quickly form summary representations that capture the statistical structure in a set of simultaneously-presented objects. We present evidence that such ensemble encoding is informed not only by the presented set of objects, but also by a meta-ensemble, or prototype, that captures the structure of previously viewed stimuli. Participants viewed four objects (shaded squares in Experiment 1; emotional expressions in Experiment 2) and estimated their average by adjusting a response object. Estimates were biased toward the central value of previous stimuli, consistent with Bayesian models of how people combine hierarchical sources of information. The results suggest that an inductively learned prototype may serve as a source of prior information to adjust ensemble estimates. To the extent that real environments present statistical structure in a given moment as well as consistently over time, ensemble encoding in real-world situations ought to take advantage of both kinds of regularity.

Keywords Ensemble encoding · Category learning · Statistical learning

Introduction

Visual ensemble encoding is the ability to form summary representations that capture the statistical structure inherent in visual experience. For example, when people view simultaneously-presented objects that vary along continuous dimensions such as size, orientation, or color, or a group of faces that vary in emotional expression, they accurately judge the set's average value on the relevant dimension even when many (16) objects are shown (Haberman & Whitney, 2009), when presentation is as brief as 50 ms and a memory delay is introduced (Chong & Treisman, 2003), or when objects are presented outside of focal attention (Alvarez & Oliva, 2009). Several studies suggest that people can report summary information even when they can report little about any individual object that was shown (Ariely, 2001; Corbett & Oriet, 2011). Because ensembles allow us to represent multiple objects with

summary information rather than with separate representations for each object, they achieve a form of informational compression (Alvarez, 2011). This compression can support performance on tasks that demand high visual processing capacity, such as visual search (e.g., Chetverikov, Campana, & Kristjánsson, 2016) or remembering individual objects (Brady & Alvarez, 2011), and can conserve working memory resources by producing efficient memory representations (Brady, Konkle, & Alvarez, 2009).

Many studies in this line of work examine how summary statistics are gleaned during a brief glimpse of several objects, i.e., they focus on averaging that occurs *during perception* rather than across longer-run experience. Because of this, studies of perceptual averaging are often designed to so that any contaminating influence of previous exposure to similar sets of objects is averaged out or eliminated (e.g., Chong & Treisman, 2003; Haberman & Whitney, 2009). However, doing so may leave out of consideration a valuable component of ensemble encoding – namely, what influence such prior experience may have on perceptual averaging. It is well established that people acquire statistical information about objects encountered over time and Bayesian theory shows how using that information could contribute to accuracy in perceptual averaging. Drawing on theories of ensemble encoding and category effects on memory, we suggest that averaging during perception *should be* informed by averages of the past.

✉ L. Elizabeth Crawford
lcrawfor@richmond.edu

¹ Department of Psychology, University of Richmond, 28 Westhampton Way, Richmond, VA 23173, USA

² Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA

Evidence that people store summary representations of objects presented over time (i.e., across several trials) comes from several studies of learning, memory, and perception. Some studies probe these summaries directly by presenting a series of objects and then asking participants to report the average of the set (e.g., Duffy & Crawford, 2008; Oriet & Hozempa, 2016). Others take a more indirect approach by assessing how statistical information about previous stimuli affects performance on other tasks, such as visual search (e.g., Chetverikov et al., 2016; Corbett & Melcher, 2014). Central to the current work, studies of object memory in the context of inductive category learning have shown how information about a distribution of objects informs memory for an individual object from that set (e.g., Huttenlocher, Hedges, & Vevea, 2000).

Work on memory of individual objects forms the theoretical basis of our predictions about perceptual averaging. Huttenlocher and colleagues (e.g., Crawford, Huttenlocher & Hedges, 2006; Huttenlocher et al., 2000) characterized object memory as a reconstruction that combines a noisy, unbiased memory trace of the object with prior knowledge about the distribution that generated it. They modeled each of these information sources as normal distributions that are combined according to Bayes' theorem to create a posterior distribution from which estimates are drawn. This results in responses that are a weighted average of the prior distribution and the memory trace, with the weights determined by the relative precision of the prior and the trace (i.e., all else being equal, the more precise the memory trace, the less influence of the prior). Because of this, estimates of an object are biased toward the mean of the overall distribution, a phenomenon known as the Central Tendency Bias. Bias patterns consistent with this model have been shown in memory for object size (e.g., Crawford, Huttenlocher, and Engebretson, 2000; Hemmer & Steyvers, 2009), object shade (Huttenlocher et al., 2000), hue (Olkkonen & Allred, 2014; Persaud & Hemmer, 2016), and facial expression of emotion (Corbin, Crawford, & Vavra, 2017).

Here we highlight an alignment between this model – developed in the literature on memory and categorization – with theories of perceptual averaging, but first we acknowledge the differences between these research literatures. In the memory research described above, a typical experiment presents a series of items that vary on a continuous dimension (e.g., lines varying in length) one at a time for 1–2 s and, after a delay, has participants reproduce each item before moving on to the next. Perceptual averaging studies also use stimuli that vary on continuous dimensions, but they typically present several objects simultaneously and briefly, and then immediately instruct participants to estimate the average of the items shown. Whereas both lines of research claim that people grasp

summary statistics from a set of objects, the memory studies examine that aggregation across time while the perceptual averaging studies examine it across space. In addition, the memory studies assess averaging only indirectly (by examining its impact on object memory), while the perceptual averaging studies explicitly ask participants to estimate the average. Finally, memory tasks use longer presentation times and post-stimulus delays than perceptual averaging tasks.

These tasks have emerged independently from studies that rarely refer to each other, and they exemplify the general difference in approach between memory and perception research. However, from a computational perspective, there is substantial overlap between perceptual ensembles and inductively learned categories. Specifically, both address how summary information about groups of objects enables us to make good use of the inherently limited resources available for cognitive processing. Here we take an integrated approach to ensemble encoding and inductive category learning, coordinating both phenomena in the same paradigm. We propose that the impressive accuracy of perceptual averaging may be due, at least in part, to how it takes advantage of inductively formed categories (i.e., knowledge about the distribution of ensembles presented on previous trials).

This integrated perspective generates new predictions for how perceptual averaging of a set of objects is affected by sets seen previously. The current work differs from earlier studies in that rather than focus on the impressive accuracy of ensemble mean estimation, we analyze the kind of errors people make. We suggest that, to the extent that there is some uncertainty about the true mean of an ensemble (and there is always some), the optimal strategy for estimating the mean should combine the perceptible structure in the presented ensemble with the distribution acquired during previous trials. This hierarchical combination predicts that estimates of ensemble means should show the same pattern of bias found in estimates of individual objects: they should be biased toward the central value of previous ensemble means.

In two experiments, participants viewed four objects at once and then estimated the mean of those objects along a continuous dimension. In Experiment 1 the objects were squares that varied in shades of gray; in Experiment 2 they were morphed faces that varied in expression from sad to happy. In each experiment, participants were assigned to one of two overlapping distributions of ensembles. We predicted that within each distribution, estimates of ensemble means would be biased toward the central value of the distribution that participants were shown. We further predicted that estimates of the single ensemble that was common to both distribution conditions would be judged differently depending on the distribution in which it was embedded, as it would be biased toward different distribution means.

Experiment 1

Method¹

Participants

In accordance with our preregistration, 60 participants were recruited from Mechanical Turk. Six were removed because they failed to complete the study. The sample was 61.1% male, 84% White, 7% African American, 4% Asian, 2% American Indian or Alaska Native, and 4% preferred not to answer; 89% Non-Hispanic or Latino, 11% Hispanic or Latino; and the average age was 37.3 years ($SD = 10.49$, $min = 19$, $max = 63$).

Materials

We created a sequence of 41 images of squares that ranged in shade from very dark gray (RGB values equal to 10) to very light gray (RGB values equal to 214 (increments were approximately 5 units)). By convention, we treat this sequence as a continuous dimension centered on zero with the darkest value at -20 and the lightest value at +20, with these values corresponding to the numbered image frames from the sequence. The actual shades viewed depended on participants' own computer settings, making grayscale values imprecise and variable, but adequate for the goals of this study.

Procedure

At the start of the experiment, participants were asked to maximize their browser window. Regardless of participants' screen dimensions, the relevant task area always appeared as a light blue (RGB 46, 208, 243) rectangle (1,024 × 768 pixels) and areas of the browser window outside the task area were filled black. As shown in Fig. 1, each trial presented four squares for 1,000 ms, followed by a 500-ms blank screen, followed by a response square in the center of the screen. Participants could adjust the response square to make it match the average shade of the set they had just seen by pressing one key to cycle through the images in the darker direction and another key to cycle in the lighter direction. This created an animated effect of a square that changed shade seamlessly. Participants judged 14 unique ensembles six times each for a total of 84 randomly ordered trials.

The between-groups manipulation was the distribution of ensembles shown. In the Dark condition, the ensembles had average frame values that ranged from -13 to 0; in the

Light condition, these values that ranged from 0 to +13. Within each ensemble, the four squares differed from each other by four frame units (thus spanning 20% of the range that participant viewed during the experiment). The starting value for the response square was randomly assigned to be -20 (Darkest), 0 (Midpoint), or 20 (Lightest) on each trial and balanced so that each ensemble was estimated from each starting value two times.

Results

Response bias is calculated as the response frame value minus the true average value of the ensemble, such that higher scores indicate a bias to judge the ensemble as lighter than it was. Error is the absolute value of bias. In accordance with preregistered culling procedures, two participants were removed for having an average error of greater than 2.5 SD s above the overall by-participant error mean. Individual trials were then removed if they had an error greater than 3 SD s above the overall mean error. These procedures removed 225 (5%) out of 4,536 total observations collected from those who finished the experiment.

Figure 2 shows bias in estimates as a function of the true ensemble mean shown. The overall pattern shows that, within each condition, darker ensembles were judged to be lighter than they were (i.e. positive bias) and lighter ensembles were judged to be darker than they were (negative bias), producing negatively sloping bias patterns. We tested for an intercept difference between the two conditions with a linear mixed model in which bias was predicted by condition (Dark vs. Light) and average shade, including participant as a random effect. As shown in Table 1, the intercept was 2.69 units ($t(81) = 6.05$, $p < .001$) lighter in the Light condition than in the Dark condition, indicating that the discontinuity in bias apparent in Fig. 2 is unlikely to be due to chance. Noting that the intercept differences are driven by the whole range of ensemble means which differed between conditions, we also examined estimates of the only ensemble mean that was common to both conditions: the one centered on zero. Estimates of that ensemble were significantly lighter in the Light condition ($M = 2.45$, $SD = 2.23$) than in the Dark condition ($M = -0.33$, $SD = 2.2$), Welch's $t(47.61) = 4.49$, $p < .001$, $d = 1.26$. Finally, we assessed whether the bias pattern was negatively sloped in each condition by running the above linear model (without the condition variable) separately for each condition. The slope in the Dark condition was $B = -0.39$, $SE = .02$, $t(1,871.10) = -16.92$, $p < .001$ and the slope in the Light condition was $B = -0.41$, $SE = .02$, $t(2,388.10) = -19.80$, $p < .001$.²

¹ All materials, data files, and preregistrations for both experiments are available at: <https://osf.io/ndfhv/>. Both experiments were approved by the Institutional Review Board and run in accordance with the American Psychological Association's ethics code.

² Including the starting value of the response in the model did not change reported results in either experiment. Consistent with prior work (Allred, Crawford, Duffy, & Smith, 2016; Corbin et al., 2017), estimates were biased in the direction of each starting value. Those interested in these effects can upload the data from <https://osf.io/ndfhv/>.

Fig. 1 A visualization of a single trial of Experiment 1. The left image represents study and the right, response. Not present in the photo on the left is a green square, which was placed in the bottom left corner, which

Discussion

Participants were sensitive to the distribution of stimuli they were shown and used that information when estimating ensemble means. The negatively sloped bias patterns show that within each condition, estimates were biased inward, away from the darkest and lightest stimuli. We also note the unpredicted result that estimates in the dark condition are generally shifted to be lighter than would be expected if prior ensembles were the only source of bias in responses. Comparisons between conditions show that estimates of the ensemble that was centered on medium gray depended on the distribution in which it was embedded: it was judged to be darker when embedded in a dark distribution than when in a lighter distribution. These results mirror findings of inductive category effects on memory of objects (e.g., Huttenlocher et al., 2000), which show distributions learned across trials bias estimates.

Experiment 2

Experiment 2 examined whether the effects observed in Experiment 1 generalize to ensemble encoding of faces that vary in emotional expression. Faces are multidimensional, affectively charged, socially relevant, and they tend to be processed holistically (Maurer, Le Grand, & Mondloch, 2002). Despite this complexity, several studies have established that people can perceptually average facial features (e.g., Haberman, Brady, & Alvarez, 2015; Haberman & Whitney, 2009). We varied emotional expression by morphing images of an actor making sad, neutral, and happy facial expressions and treating the resulting image frames as points along a continuous dimension. The resulting faces necessarily differ in both affective significance and physical characteristics, such as mouth shape and brow orientation. Here we do not attempt to distinguish affective

participants pressed to move on to the next trial. In the actual experiment, the background was colored light blue in order to allow for the lightest-shaded square to be distinguished from the background

and perceptual processing of these stimuli and instead capitalize on previous work showing that continua generated in this way produce results that mirror those found in studies using simpler dimensions such as size, color, or shade (Corbin et al., 2017). That work suggests that, when shown a set of faces that vary on a morphed expression continuum, people are sensitive to the central tendency of the set along that dimension.

Method

Participants

We recruited 105 participants from Mechanical Turk and removed 13 for failing to complete the study. One participant did not complete the demographic questionnaire and is not included in the following statistics. The sample was 54% male, 88% White, 9% African American, and 3% Asian; 85% Non-Hispanic or Latino, 13% Hispanic or Latino, and 2% preferred not to answer. The average age of our sample was 37.9 years ($SD = 10.92$, $min = 20$, $max = 67$).

Materials

Images used in the present study were drawn from the NimStim³ face stimulus set. Photos of a woman (Female 1 from the Nimstim database) making sad, neutral, and happy expressions were used to create the stimuli. Using FantaMorph software (Abrosoft, 2002), two morphs were created: one changed from the model's sad expression to her neutral expression, the second from her neutral to her happy expression in 5% increments. From these morphs, we

³ Development of the MacBrain Face Stimulus Set was overseen by Nim Tottenham and supported by the John D. and Catherine T. MacArthur Foundation Research Network on Early Experience and Brain Development. Please contact Nim Tottenham at tott0006@tc.umn.edu for more information concerning the stimulus set.

Fig. 2 Mean bias (with 95% CI) and slopes across levels of ensemble average shade (-13 Darkest to 13 Lightest) for Dark and Light distribution conditions. Means at ensemble average shade 0 are highlighted to note the difference in means between the shade that was shared by both distributions. Figure created with ggplot2 R package (Wickham, 2009)

extracted 41 evenly distributed expressions of each model's face ranging from sad (expression -20), to neutral (expression 0), to happy (expression 20) (see Fig. 3 for example). We treat these morph units as a continuous dimension, analogous to the shade dimension used in Experiment 1.

Procedure

The procedure and design are analogous to Experiment 1, but the squares were replaced with images of faces showing expressions that ranged from sad to happy (faces were slightly smaller than the squares at 152×195 pixels and they were 232 pixels apart on the x-axis and 275 pixels apart in the y-axis.) The background color for the browser window in this experiment was grey (RGB 192,192,192). As in Experiment 1, the response face was initially set at either of the two most extreme values or the central value and could be adjusted with a

key press, and the adjustment created a smooth animation in which the face appeared to change expression continuously.

Results

Bias is calculated as the response value minus the true average value of the ensemble, such that higher scores indicate a bias to judge the ensemble as happier than it was. Five participants were removed for having an average error greater than 2.5 *SDs* above the overall by-participant error mean and then individual trials greater than 3 *SDs* above the overall mean error were also removed. These procedures removed 547 (7.1%) out of 7,728 total observations collected from those who finished the experiment.

Figure 4 shows bias in estimates as a function of the true ensemble mean shown. A linear mixed model with bias predicted by condition (Happy vs. Sad) and average expression, including participant as a random effect (see Table 2) showed that the intercept difference was 2.99 morph units ($t(152) = 7.51, p < .001$). As

Table 1 Linear mixed models predicting bias

	Full model		Dark		Light	
	B (95% CI)	SE	B (95% CI)	SE	B (95% CI)	SE
Fixed effects						
(Intercept)	-0.33 (-0.94 – 0.28)	.31	-0.24 (-0.92 – 0.45)	.35	2.43*** (1.87 – 2.99)	.28
Average Shade	-0.40*** (-0.43 – -0.37)	.02	-0.39*** (-0.43 – -0.34)	.02	-0.41*** (-0.45 – -0.37)	.02
Light Condition	2.69*** (1.82 – 3.56)	.44				
Random effects						
$\sigma_{\text{Participant}}$	1.35		1.45		1.27	
σ_{Error}	4.06		3.99		4.11	
N	52		23		29	
R^2 / Ω_0^2	.270 / .269		.224 / .224		.213 / .213	

* $p < .05$, ** $p < .01$, *** $p < .001$

Fig. 3 Selected morphed expressions from Sad to Happy (note that in the study, a morph unit refers to a 5% interval)

in Experiment 1, we examined estimates of the ensemble common to both conditions: the one centered on neutral. Estimates of that ensemble were significantly happier in the Happy condition ($M = 1.65$, $SD = 3.15$) than in the Sad condition ($M = -1.3$, $SD = 2.74$), Welch's $t(72.41) = 4.59$, $p < .001$, $d = 1.00$. Finally, we tested whether the bias pattern was negatively sloped in each condition by running the above linear model (without the condition variable) separately for each condition. The slope in the Sad condition was $B = -0.31$, $SE = .03$, $t(2,627.2) = -11.20$, $p < .001$ and the slope in the Happy condition was $B = -0.32$, $SE = .02$, $t(4,468) = -17.98$, $p < .001$, showing that ensembles on the sadder end of the presented range were judged to be happier and ensembles on the happier end of the presented range were judged to be sadder within each range that was shown.

Discussion

The results show that the pattern of bias found with shade values in Experiment 1 generalizes to estimates of average facial expressions on an emotional continuum. Within each condition, estimates were biased toward the center of the presented distribution and away from its outer edges. Comparisons between conditions show that the ensemble that was centered on the neutral face was estimated to be happier when it was embedded in the Happy distribution than when it was embedded in the Sad distribution.

General discussion

Research on perceptual averaging suggests that people can quickly form summary representations that capture the statistical structure of simultaneously presented objects. We present initial evidence that these summary representations are informed not only

by the set of objects presented on a particular trial, but also by a meta-ensemble, or prototype, that reflects the statistical structure of previously shown ensembles. The results of two experiments showed that estimates of ensemble means are biased toward the overall central tendency of presented ensembles. This effect held regardless of which distribution was shown and generalized across different kinds of stimuli. As Bayesian models have noted elsewhere (Brady & Alvarez, 2011; Feldman, Griffiths, & Morgan, 2009; Huttenlocher et al., 2000), this type of bias is the expected outcome when people combine hierarchical sources of information in a manner that optimizes accuracy of estimates.

The findings extend previous studies of inductive category effects on estimates of individual objects. Whereas the Category Adjustment Model (CAM; e.g., Huttenlocher et al., 2000) was developed to address the central tendency bias in object estimation, the Bayesian combination can apply to a range of circumstances that involve combining sources of information that have varying degrees of uncertainty. By using CAM to predict a central tendency bias in ensemble mean estimation, we demonstrate the generality of its approach beyond memory of a single stimulus. The findings also extend research on ensemble encoding by examining the influence of prior knowledge. Many studies of perceptual averaging are designed to avoid or average out prior knowledge effects rather than to investigate them. One exception is work by Whiting and Oriet (2011), which assessed perceived means of object size with a two-alternative forced choice response and found that participants were biased to choose the incorrect response when it was closer to the cumulative mean than was the correct response. This type of error was especially pronounced at brief stimulus presentations and when stimuli were immediately masked. Although they interpreted their results as evidence, at least under certain conditions, for strategic responding rather than ensemble encoding, we see their

Fig. 4 Mean bias (with 95% CI) and slopes across levels of ensemble average expression (-13 Saddest to 13 Happiest) for Sad and Happy distribution conditions. Means at ensemble average expression 0 are highlighted to note the difference in means between expressions that were shared by both distributions. Figure created with ggplot2 R package (Wickham, 2009)

Table 2 Linear mixed models predicting bias

	Full model		Sad		Happy	
	B (95% CI)	SE	B (95% CI)	SE	B (95% CI)	SE
Fixed effects						
(Intercept)	-1.67*** (-2.24 – -1.10)	.29	-1.60*** (-2.33 – -0.87) **	.37	1.36*** (0.93 – 1.78)	.21
Average Expression	-0.32*** (-0.35 – -0.29)	.02	-0.31*** (-0.36 – -0.25)	.03	-0.32*** (-0.36 – -0.29)	.02
Happy Condition	2.99*** (2.21 – 3.76)	.40				
Random effects						
$\sigma_{\text{Participant}}$	1.45		1.76		1.23	
σ_{Error}	5.17		5.66		4.86	
N	87		32		55	
R^2 / Ω_0^2	.140 / .139		.136 / .135		.128 / .128	

* $p < .05$, ** $p < .01$, *** $p < .001$

results as consistent with a Bayesian perspective on ensemble encoding: Manipulations that increase uncertainty about a particular ensemble should increase the relative weight given to prior knowledge during estimation, thus leading to stronger bias. The optimal solution for estimating an ensemble mean will take advantage of statistical structure that is available not only across space but also across time.

Here we have provided initial evidence that people use prior information to adjust estimates of ensemble means. This work stems from computational-level considerations and does not specify the processing mechanisms that give rise to this effect. For instance, whereas Ariely (2001) argued that people average across items without storing information about individual objects, others have argued that people may focus on an individual object in the ensemble to generate their estimates of the average (de Fockert & Marchant, 2008; Muczek & Simons, 2008), and the present work does not distinguish these possibilities. In addition, there is debate about whether the priors used in such tasks reflect summaries of the distribution of all previous items equally weighted or only a recent subset of them (e.g., Duffy, Huttenlocher, Hedges, & Crawford, 2010; Duffy & Crawford, 2008; Duffy & Smith, 2017), and the current work is not designed to address that question. Regardless of the exact nature of the prior, these findings indicate that people adjust their perceptual averages with prior knowledge about the distribution from which the ensemble was drawn, and that they do so in a manner that is consistent with the predictions of a Bayesian model.

The degree of benefit achieved through this hierarchical combination depends on how accurately people's priors capture the true distribution of objects. We report evidence that people form and use priors that are at least somewhat sensitive to the central tendency of presented ensembles. Although this sensitivity is necessary, it is not sufficient to demonstrate that people are optimally tuned to the structure of their environments and it remains

an open question whether people weight prior information optimally during estimation. Our broader aim is to suggest a fundamental commonality between ensemble encoding and inductive categories. Despite coming from different research traditions, estimating an individual object from memory and estimating an ensemble mean have a common computational goal of combining multiple sources of information to achieve accuracy of estimates under conditions of uncertainty.

The work presented here suggests that memory of prior experience is not a contaminant of a separate perceptual-averaging process, but instead may be a central component. This approach generates clear, testable predictions about the shape of error patterns in ensemble mean estimation. The results suggest that that memory does not step in when perception fails – instead, judgments combine perceptual information with past statistics in specific, theoretically-motivated ways. Although it is possible to create circumstances that minimize the contribution of prior knowledge in ensemble encoding, doing so may limit our view of the phenomenon. Perceptual averaging research can be designed to explore the kinds of processes that are probably ubiquitous in real-world situations: those that integrate statistical structure gleaned in an instant and acquired over time.

Author Note Support for this project was provided by a fellowship from the University of Richmond's Arts and Sciences Faculty Research Committee.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Abrosoft. (2002). FantaMorph SE (Version 5.4.6) [Software]. Available from <http://www.fantamorph.com/>

- Allred, S.R., Crawford, L.E., Duffy, S., & Smith, J.B. (2016). Working memory and spatial judgments: Cognitive load increases the central tendency bias. *Psychonomic Bulletin & Review*, 23(6), 1825–1831.
- Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, 15(3), 122–131. <https://doi.org/10.1016/j.tics.2011.01.003>
- Alvarez, G. A., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proceedings of the National Academy of Sciences*, 106(18), 7345–7350. <https://doi.org/10.1073/pnas.0808981106>
- Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, 12, 157–162. <https://doi.org/10.1111/1467-9280.00327>
- Brady, T. F., & Alvarez, G.A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science*, 22(3), 384–392. <https://doi.org/10.1177/0956797610397956>
- Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: Using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General*, 138(4), 487–502. <https://doi.org/10.1037/a0016797>
- Chetverikov, A., Campana, G., & Kristjánsson, Á. (2016). Building ensemble representations: How the shape of preceding distractor distributions affects visual search. *Cognition*, 153, 196–210. <https://doi.org/10.1016/j.cognition.2016.04.018>
- Chong, S. C. & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, 43, 393–404. [https://doi.org/10.1016/S0042-6989\(02\)00596-5](https://doi.org/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.
- 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 psychologica*, 138(2), 289–301. <https://doi.org/10.1016/j.actpsy.2011.08.002>
- Corbin, J. C., Crawford, L. E., & Vavra, D. T. (2017). Misremembering emotion: Inductive category effects for complex emotional stimuli. *Memory & Cognition*, 45(5), 691–698. <https://doi.org/10.3758/s13421-017-0690-7>
- Crawford, L. E., Huttenlocher, J., & Engebretson, P. H. (2000). Category effects on estimates of stimuli: Perception or reconstruction? *Psychological Science*, 11, 280–284. <https://doi.org/10.1111/1467-9280.00256>
- Crawford, L.E., Huttenlocher, J. & Hedges, L.V. (2006). Within-category feature correlations and Bayesian adjustment strategies. *Psychonomic Bulletin & Review*, 13, 245–250. <https://doi.org/10.3758/BF03193838>
- de Fockert, J. W., & Marchant, A. P. (2008). Attention modulates set representation by statistical properties. *Perception & Psychophysics*, 70, 789–794. <https://doi.org/10.3758/PP.70.5.789>
- Duffy, S., & Crawford, L. E. (2008). Primacy or recency effects in the formation of inductive categories. *Memory & Cognition*, 36, 567–577. <https://doi.org/10.3758/MC.36.3.567>
- Duffy, S., Huttenlocher, J., Hedges, L. V., & Crawford, L. E. (2010). Category effects on stimulus estimation: Shifting and skewed frequency distributions. *Psychonomic Bulletin & Review*, 17, 224–230. <https://doi.org/10.3758/PBR.17.2.224>
- Duffy, S., & Smith, J. (2017). Category effects on stimulus estimation: Shifting and skewed frequency distributions – A reexamination. *Psychonomic Bulletin & Review*. <https://doi.org/10.3758/s13423-017-1392-7>
- Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009). The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review*, 116(4), 752–782. <https://doi.org/10.1037/a0017196>
- 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. <https://doi.org/10.1037/xge0000053>
- Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 718–734. <https://doi.org/10.1037/a0013899>
- Hemmer, P., & Steyvers, M. (2009). Integrating episodic memories and prior knowledge at multiple levels of abstraction. *Psychonomic Bulletin & Review*, 16, 80–87.
- Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General*, 129, 220–241. <https://doi.org/10.1037/0096-3445.129.2.220>
- Maurer D., Le Grand, R., Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, 6, 255–260. [https://doi.org/10.1016/S1364-6613\(02\)01903-4](https://doi.org/10.1016/S1364-6613(02)01903-4)
- Muczek, K., & Simons, D. J. (2008). Better than average: Alternatives to statistical summary representations for rapid judgments of average size. *Perception & Psychophysics*, 70, 772–788. <https://doi.org/10.3758/PP.70.5.772>
- Olkonen, M., & Allred, S. R. (2014). Short-term memory affects color perception in context. *PloS One*, 9, e86488. <https://doi.org/10.1371/journal.pone.0086488>
- Oriet, C., & Hozempa, K. (2016). Incidental statistical summary representation over time. *Journal of Vision*, 16(3), 1–14. <https://doi.org/10.1167/16.3.3>
- Persaud, K., & Hemmer, P. (2016). The Dynamics of Fidelity over the Time Course of Long-Term Memory. *Cognitive Psychology*, 88, 1–21. <https://doi.org/10.1016/j.cogpsych.2016.05.003>
- Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://doi.org/10.1007/978-0-387-98141-3>
- Whiting, B. F., & Oriet, C. (2011). Rapid averaging? Not so fast!. *Psychonomic Bulletin & Review*, 18(3), 484–489. <https://doi.org/10.3758/s13423-011-0071-3>