



# Variability of dot spread is overestimated

Jessica K. Witt<sup>1</sup> · Mengzhu Fu<sup>2</sup> · Michael D. Dodd<sup>2</sup>

Accepted: 8 June 2022 / Published online: 16 June 2022

© The Psychonomic Society, Inc. 2022

## Abstract

Previous research has demonstrated that individuals exhibit a tendency to overestimate the variability of both low-level features (e.g., color, orientation) and mid-level features (e.g., size) when items are presented dynamically in a sequential order, a finding we will refer to as the variability overestimation effect. Because previous research on this bias used sequential displays, an open question is whether the effect is due to a memory-related bias or a vision-related bias. To assess whether the bias would also be apparent with static, simultaneous displays, and to examine whether the bias generalizes to spatial properties, we tested participants' perception of the variability of a cluster of dots. Results showed a consistent overestimation bias: Participants judged the dots as being more spread than they actually were. The variability overestimation effect was observed when there were 10 or 20 dots but not when there were 50 dots. Taken together, the results of the current study contribute to the ensemble perception literature by providing evidence that simultaneously presented stimuli are also susceptible to the variability overestimation effect. The use of static displays further demonstrates that this bias is present in both dynamic and static contexts, suggesting an inherent bias existent in the human visual system. A potential theoretical account—boundary effect—is discussed as a potential underlying mechanism. Moreover, the present study has implications for common visual tasks carried out in real-world scenarios, such as a radiologist making judgments about distribution of calcification in breast cancer diagnoses.

**Keywords** Ensemble perception · Visual biases · Perceived spread · Variability

A critical function of human perception is the ability to accurately process and perceive variability in our environment. People are continually confronted by variability of all forms, as objects can differ along several dimensions including color, size, weight, speed, and orientation. As a consequence, humans have developed the ability to rapidly extract summary statistical information from groups of items, a process known as ensemble perception (for review, see Whitney & Yamanashi Leib, 2018). For example, people can see the mean orientation, color, and size of objects, and even the mean emotional expression of a group of faces. People can also detect other summary statistics such as the variability of a set of items (Dakin & Watt, 1997; Haberman et al., 2015; Morgan et al., 2008; Norman et al., 2015; Solomon et al., 2011).

And yet, despite the incredible capacity for ensemble perception, biases also exist. For example, perceivers

underestimate variability when stimuli are presented in an ordered fashion (e.g., smallest to largest) relative to when the order is randomized (Lathrop, 1967). Perceivers also underestimate the variability when line-drawings of cylinders are colored to various heights (Kareev et al., 2002). Increased variability also leads to a bias to overestimate the mean size of circles and mean length of lines (Semizer & Boduroglu, 2021). Moreover, Kanaya et al. (2018) have reported amplification effects wherein viewers' perception of the mean size and temporal frequency of array are biased by the most salient items in the display, and that this effect increases as the number of items in the salient group increase.

Most relevant to the current work, a recent bias to overestimate variability has been observed in a series of experiments in which items were presented dynamically (Witt, 2019). In one experiment, participants viewed a set of nine lines, each presented at a different orientation and in a rapid sequence. Participants estimated the variability of the set of lines by performing a bisection task for which they selected one of several response options with some options being low in variability and other options being high in variability. Bisection tasks have been used to measure time perception in pigeons, mice, humans, and rats (Penney et al., 2008; Raslear, 1985) and speed perception in humans (Witt & Sugovic, 2010).

✉ Jessica K. Witt  
Jessica.Witt@colostate.edu

<sup>1</sup> Colorado State University, Fort Collins, CO, USA

<sup>2</sup> University of Nebraska–Lincoln, Lincoln, NE, USA

These tasks require observers to form an implicit midpoint, and the measurement of this midpoint is compared to the objective midpoint to identify the direction and magnitude of any bias. When judging the variability of line orientation, participants overestimated the variability of the lines by approximately 50% (Witt, 2019). We refer to this bias to overestimate the variability of the ensemble as the *variability overestimation effect*.

The variability overestimation effect is not limited to line orientation. Follow-up studies have documented a similar bias to overestimate variability of circle size and color value (Warden et al., 2022; Witt et al., 2019). In one set of studies, participants saw a set of circles that varied in size and estimated variability by performing the bisection task for which they judged whether the variability of the circle sizes was “less” or “more” variable. In another set of studies, participants saw circles that varied in color value (such as shades of green) and judged whether the variability of the circles’ color values were “less” or “more” variable. Participants overestimated the variability for both features. This reveals the variability overestimation effect generalizes to multiple features including low-level features of orientation and color value and the mid-level feature of size.

To date, however, the research on the variability overestimation effect has exclusively used dynamic, sequentially presented stimuli. Each item in a set was presented sequentially with only one object in the set visible at a time and each item presented in the same spatial location. This raises the issue of whether the variability overestimation effect is specific to sequential displays or whether the bias would also be present when all items in the set are presented simultaneously. This issue is relevant for mechanistic and for applied reasons. If the variability overestimation effect is observed only with sequential displays, this could be indicative that the effect is due to a memory-related bias. In contrast, if the variability overestimation effect is also observed with simultaneous displays, this could be indicative of a general bias inherent in the visual system that could impact many if not all forms of ensemble perception.

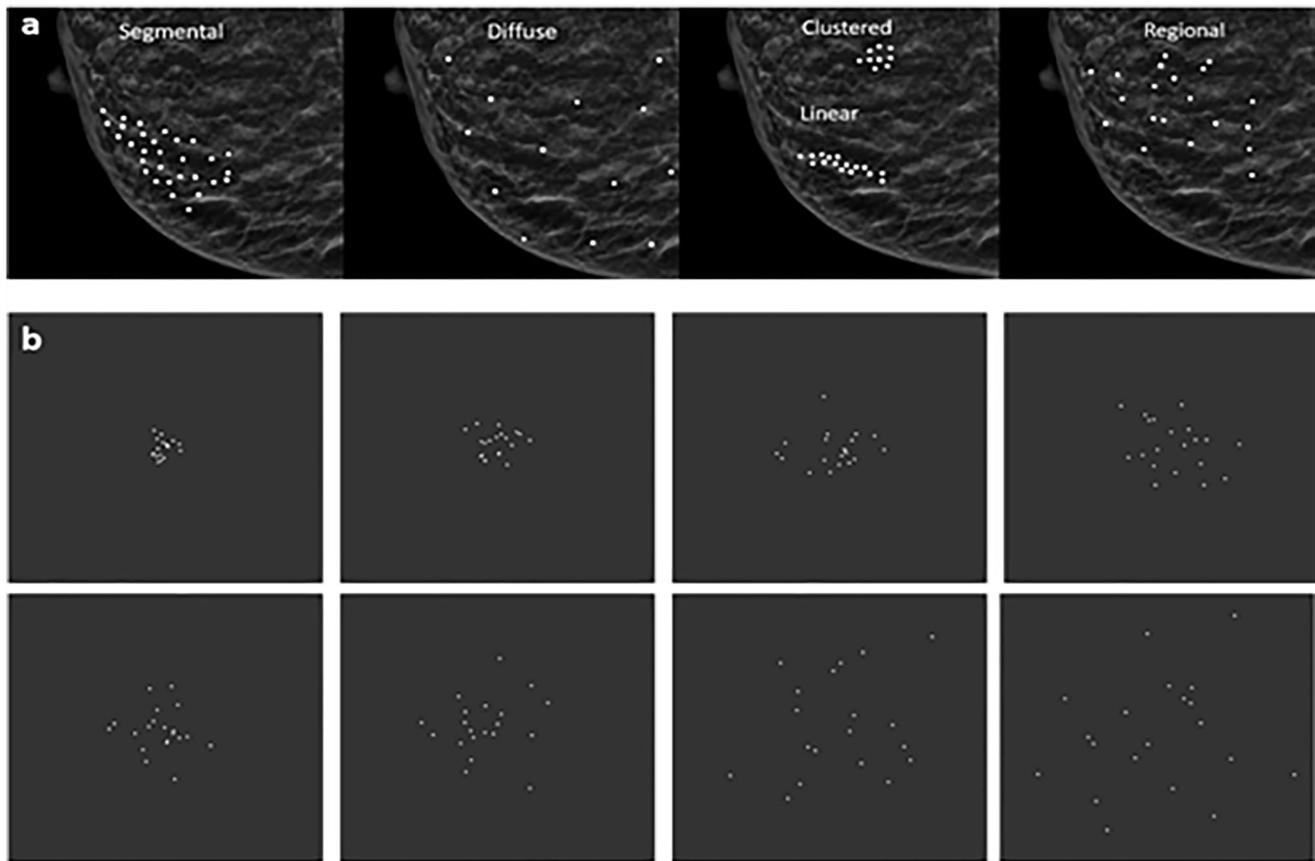
With simultaneous displays, it is also possible to determine whether the effect is observed when variability is defined in terms of the spatial properties between items. In the same way that we can rapidly extract information about the mean size, color, or orientation of items, there is considerable evidence that ensemble information can be rapidly extracted from spatial arrays that are not necessarily defined by featural differences. For example, using the centroid paradigm (Drew et al., 2010; Sun et al., 2016, 2018), it has been shown that individuals are able to accurately determine the centroid (e.g., center of mass) of a briefly presented array of dots. Accurate centroid perception has been reported with a variety of different display types and tasks (e.g., Friedenber & Liby, 2002; McGowan et al., 1998).

Biases have also been reported, however, which influence the accuracy of spatial perception in these tasks. As one example, dots clustered more closely together exerted less weight on estimated mean position (i.e., centroid) compared with more isolated dots (Rashid & Chubb, 2021). As another example of bias, observers were not as accurate in determining the centroid of an array of target stimuli when presented simultaneously with distracting stimuli that were to-be-ignored (Drew et al., 2010) and were even worse when the target stimuli are comprised of two or more levels of a single feature dimension (Lu et al., 2019). Adaptation effects also lead to biases: adaptation to dense visual texture affects accurate perception of both the numerosity and spatial distribution of stimuli comprising a texture (Durgin, 1995).

In the current experiments, we tested whether the variability overestimation effect is also present when viewing a cluster of dots. These studies test the generalizability of the variability overestimation effect by testing its presence with simultaneous displays and with an ensemble that varies along the spatial property of location or position, rather than object properties such as orientation, color value, and size. Testing the generalizability of the bias helps determine the boundary conditions of the effect, which has implications for the underlying mechanisms such as whether there is likely to be a memory-related component.

From a practical standpoint, the variability overestimation effect has real-world implications. There are many instances for which it is necessary to judge the variability of an ensemble presented simultaneously. Within the realm of data visualization, the variability of the data can be an important part of the message being communicated, so biases in perception of the variability could impact data interpretation. Szafir and colleagues (Albers et al., 2014; Szafir et al., 2016) have shown that individuals can make judgments about the aggregate properties of data presented to them in single images and that their ability to extract certain types of information is a function of the visualization design: different display types lead to differential efficiency in ensemble processing.

A real-world example with life-or-death consequences is mammography. To diagnose breast cancer, radiologists must perform the visual task of identifying cancer in a scan. One of these visual tasks is to classify the distribution of calcifications in the breast into one of five categories (see Fig. 1a). As this is primarily a visual task, biases within the visual system could impact diagnostic accuracy. In the present study, we examined whether the variability overestimation effect is also present in a task in which all relevant stimuli are presented simultaneously. To give the task real-world relevance, the displays were designed to mimic one aspect of the task of a radiologist: Participants viewed an ensemble of dots that were either more clustered or more dispersed and judged the spread of the dots (see Fig. 1b).



**Fig. 1** **a** Examples of the five categories of calcification distributions for diagnosing breast cancer (image courtesy of Dr. Elizabeth Edney; Witt et al., 2022), and **(b)** examples of our experimental stimuli. In **(b)**, the first

image shows the lowest level of spread with increasing levels for each subsequent image, and the last image shows the highest level of spread

If the variability overestimation effect (such as with line orientation) is also found in the perceived variability of simultaneously presented dots, this could have implications for accurately diagnosing breast cancer. A bias to overestimate variability in calcification spread could lead to distributions that should be classified as “clustered” as being “diffuse.” Given that diffuse distributions portend a benign diagnosis, a visual bias to overestimate variability could lead to a higher rate of false negative diagnoses (i.e., “misses”). These misdiagnoses would lead to delays and missed opportunities for easier, more effective early treatments. To the extent that any pathological condition depends on a visual judgment of a distribution, the bias to overestimate variability in the distribution could have adverse consequences.

## Experiment 1

The stimuli and design of the Experiment were intended to roughly mimic the task of a radiologist who needs to determine the distribution of calcifications in a scan. Participants saw displays of 20 or 50 dots with various levels of spread and classified their distribution as less or more spread out.

## Method

**Participants** The participants were 33 students from the University of Nebraska–Lincoln, who volunteered in exchange for course credit. All participants had normal or corrected-to-normal vision and were naïve as to the purpose of the study.

**Stimuli** The stimuli were a cluster of light gray dots presented on a dark-gray background (see Fig. 1b). There were either 20 or 50 dots presented on each trial. To manipulate the spread of the dots, their  $x$  and  $y$  positions were determined by randomly sampling from normal distributions with a mean of zero and a standard deviation ( $SD$ ) ranging from 0.25 to 2 in 8 evenly spaced increments (.25, .5, .75, 1, 1.25, 1.5, 1.75, 2). So that participants would have to focus on variability rather than use a proxy or heuristic (such as the vertical position of the highest dot), we systematically manipulated the mean position of the cluster of dots by offsetting the cluster to the left, right, or center and up, down, or in the middle for 9 possible mean positions. Thus, the design was a 2 (number of dots: 20, 50)  $\times$  8 (spread of dots; implemented as the  $SD$  of distribution)  $\times$  9 (mean position) for a total of 144 images. All displays were

created in R (R Core Team, 2019) and are available at <https://osf.io/xgsjc/>.

**Procedure** Participants were instructed as follows:

You will see a group of dots. Your task is to determine how spread out the dots are. Less spread out means all the dots are clustered in the same place. More spread out means the dots are scattered around the entire screen. Most groups of dots will be somewhere in between these two options. Your task is to determine whether the dots tend to be MORE spread out or LESS spread out. Ready? Press Enter.

On each trial, the dots were displayed, and participants indicated whether their spread was “less” or “more” by pressing the corresponding button (1 for less, 2 for more). The dots were visible until participants made their judgment, and time was unlimited. No feedback was provided. Unbeknownst to the participants, the first four trials were practice trials and showed the highest and lowest levels of spread for each of the two possible numbers of dots. Importantly, this ensured that all participants had the exact same context across the first four trials as the more/less spread decision is likely influenced by memory for spread on previous trials and this ensures an equivalent starting point for everyone. Order of practice trials was randomized. Participants then completed four blocks of experimental trials. Each block consisted of 144 trials with each image presented once. Trial order within block was randomized. Pilot testing indicated that the four blocks of trials could be completed within 30 minutes, and it was our intent to have all participants go through all four blocks twice. There was individual variation, however, in terms of how long participants took to respond on each trial and a subset of our participants only completed the initial four block sequence as they were not left with sufficient time to complete all blocks again. Of the 33 participants, nine completed half of the blocks and 24 completed all blocks. For those who completed all blocks, the instructions were repeated at the midway point of the experiment. All methods and procedures were approved by both the Colorado State University and University of Nebraska–Lincoln Institutional review boards.

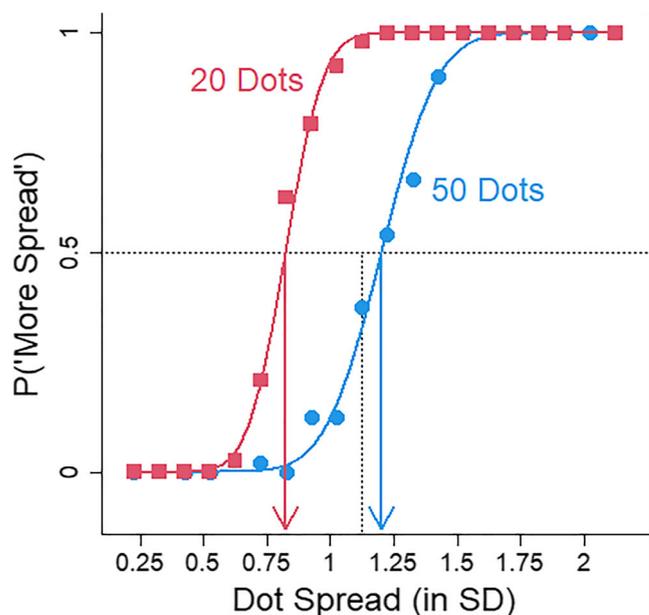
## Results

We first explored the data for outliers by running a general linear mixed model (GLMM) using the *lme4* and *lmerTest* packages (Bates et al., 2015; Kuznetsova et al., 2017). The dependent measure was response, which was coded as 0 for “less” and 1 for “more.” The fixed effect was the spread of the dots, which was mean-centered. We used the *SD* of the dots as shown, rather than the intended *SD*, as our measure of spread.

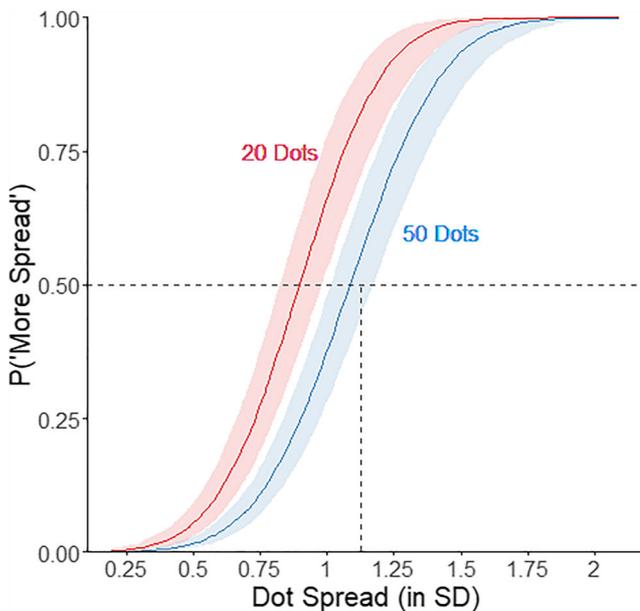
Due to random sampling, some dot locations were beyond the visible range of the plot (four images had 19 instead of 20 dots and 13 images had fewer than the intended 50 dots by 1–3 dots). Random effects for participant included the intercepts and slopes for dot spread. We explored the random effect coefficients for outliers. One participant had a PSE (see below) that was 1.5 times beyond the interquartile range (IQR) and was excluded.

To measure any bias in estimates of variability, we calculated the point of subjective equality (PSE). The PSE is the stimulus level at which the dots are judged as being equally less variable and more variable. The PSE is quantified as the stimulus level at which responses are 50%. The PSE can be contrasted with the point of objective equality (POE), which is the stimulus level that is equidistant between the lowest and highest levels of variability. The POE was 1.125. If participants overestimate variability, the PSE will be less than the POE. If there is no bias, the PSE will be similar to the POE. Data from a representative participant is shown in Fig. 2 (curves for all participants are available at <https://osf.io/xgsjc/>).

PSEs were calculated from a GLMM using the *MixedPsy* package (Moscatelli et al., 2012). For GLMM, the dependent measure was response. The fixed effects were dot spread, the number of dots, and their interaction. Random effects for participant included intercepts and slopes for each fixed effect. The model results are shown in Fig. 3. The package was also used to calculate 95% confidence intervals (CIs) via the



**Fig. 2** The mean proportion of “more spread” responses as a function of dot spread and the number of dots for one participant in Experiment 1. Lines represent binary regressions. The dashed vertical line is positioned at the POE. The arrows point to the PSEs for each number of dots. Note that overestimating variability (as shown in the red curve) leads to a PSE that is less than the POE



**Fig. 3** The estimated proportion of “more spread” responses as a function of the spread of the dots and the number of dots from the GLMM in Experiment 1. Shading represents 95% CIs. The vertical dashed line represents the POE

bootstrapping method. The package does not provide  $p$ -values, so only the 95% CIs are reported.

The mean PSEs are shown in Fig. 4. When there were 20 dots, the PSEs were less than the POE, indicating a bias to overestimate variability. The magnitude of the bias was 20%. When there were 50 dots, the PSEs were similar to the POE, indicating no bias in the estimates of variability. The 95% CIs of the difference between the PSEs did not include zero, [0.11, 0.25]. This suggests a difference in the estimation of variability as a function of the number of items in the display.

This is the first evidence for the variability overestimation effect for stimuli that were presented simultaneously instead of sequentially. It is important to note, however, that the magnitude of overestimation was smaller than has been reported for sequentially presented stimuli: 20% with 20 dots in current study versus 50% in Witt (2019), and that the effect was moderated by the amount of visual information in the display (number of dots).

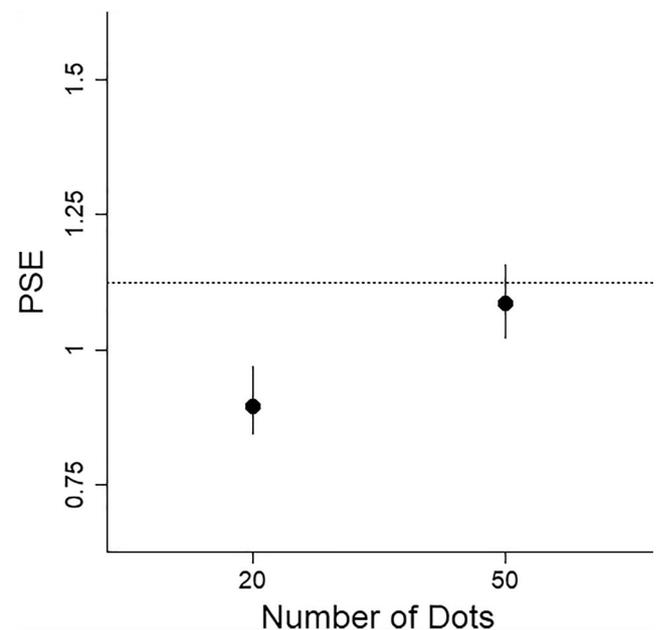
## Experiment 2

In Experiment 1, we observed a bias to overestimate the variability of the spread of a cluster of dots presented simultaneously and noted that the bias was only observed with the smaller of our two dot counts (20 dots). With 50 dots, participants were better able to estimate the true variability in the display, and we observed no bias.

This finding is reminiscent of the role of *support ratio* in the perception of visual illusions (Erlikhman et al., 2018;

Erlikhman et al., 2019; Shipley & Kellman, 1992; Zosky & Dodd, 2021). Support ratio is the idea that the strength of an illusory percept is influenced by the ratio of visible to illusory details in a display. As the amount of visible information increases, so too does the ability of the observer to overcome bias. For example, Zosky and Dodd (2021) recently demonstrated that the perceived direction of motion of an orb—which varied in particle count and could be plausibly perceived as rotating clockwise or counterclockwise—is strongly influenced by the presence of a bounding object: If a particle orb rotating to the left appears within a cube that is rotating to the right, participants tend to see the orb as moving in the same direction as the cube. This effect of the cube, however, disappears at higher particle counts for the orb. The increase in visible information (more particles) removes ambiguity from the display and allows the observer to better determine the true direction of motion of the orb

This is consistent with the results of Experiment 1 as more dots afforded a less biased determination of variability, but the variability overestimation effect was observed with fewer dots. The purpose of Experiment 2 was to replicate Experiment 1 and test the effect with even fewer dots. This result would have a real-world implication: When radiologists diagnose breast cancer based on the distribution of calcifications, they may have to make their judgments with only a few calcifications, making it important to determine the strength of the bias when less visual information is present.



**Fig. 4** PSEs for Experiment 1. The horizontal, dashed line is at the POE. Lower PSEs are indicative of overestimating the spread of the dots. Error bars are 95% CIs

### Method

**Participants** The participants were 40 volunteers recruited from Amazon’s Mechanical Turk. They participated in exchange for payment.

**Stimuli and procedure** The stimuli were created in the same way and shared the same parameters as in Experiment 1 except that displays had 10 or 20 dots rather than 20 or 50 dots. The experiment was administered via a Qualtrics survey. The experiment started with four training trials. Participants were shown one example for each number of dots at the lowest level of spread and at the highest. Each display was labeled as being “less” or “more” spread out. Viewing time was unlimited, and order was randomized.

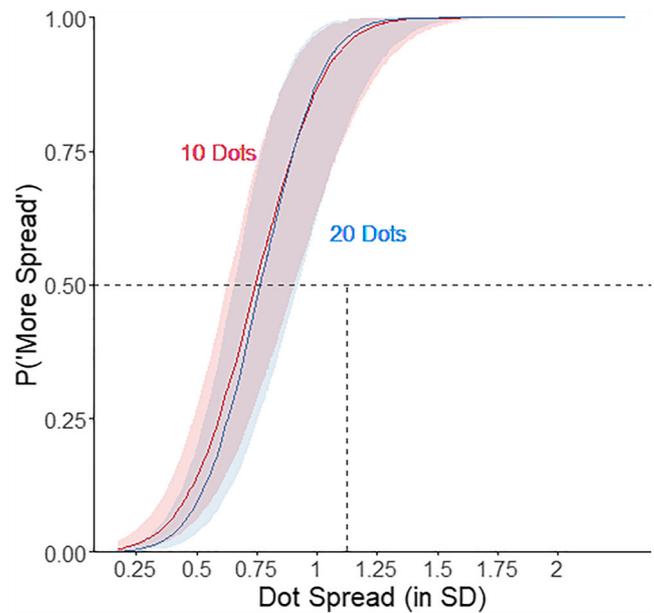
On the experimental trials, participants viewed a display of dots and indicated whether the spread of the dots was “less” or “more” by clicking on the corresponding box. Viewing time was unlimited, and feedback was not given. Participants completed 4 blocks of experimental trials, with each block consisting of a single presentation of each of the 144 images. Presentation order was randomized within each block.

### Results

Data were first explored for outliers using a simple GLMM with response as the dependent measure, dot spread as the fixed effect, and participant as the random effect. Dot spread was calculated at the *SD* of the visible dots. Due to random sampling, some dot locations were not visible in the display, leaving only nine, 18, or 19 dots (number of images in these categories were three, two, and 12, respectively, out of 144 images). The random effects coefficients were used to calculate PSEs and examined to determine outliers (see <https://osf.io/xgsgjc/> for curves for each participant). Four participants had PSEs that were beyond 1.5 times the IQR. These participants were excluded.

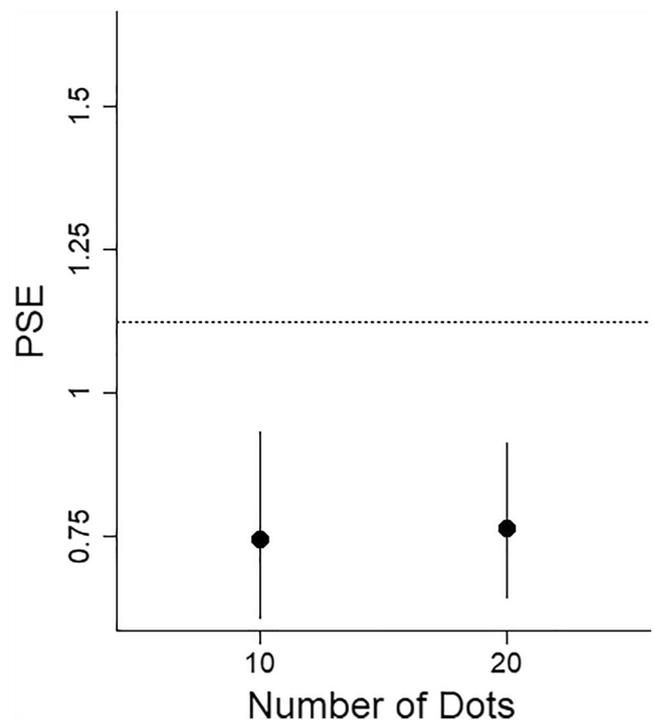
Data were analyzed by calculating the PSEs from a GLMM with the same model specification as in Experiment 1. Model results are shown in Fig. 5. PSEs and PSE difference score between 10 and 20 dots are shown in Fig. 6. As can be seen from the figures, participants overestimated the variability of the dot positions for both 10 and 20 dots. The magnitude of the bias was 34% and 32% for 10 and 20 dots, respectively. Thus, the bias was similar regardless of whether there were 10 or 20 dots. The 95% CI for the difference between the two number of dot conditions overlapped zero [−0.03, 0.06].

Some research has suggested that when judging variability, people use the range of the stimuli as a proxy for variability (Lau & Brady, 2018). To test whether our participants also used the range of the dots, rather than their variability, we modeled the data using range as an independent variable. Range was calculated for both the horizontal and vertical



**Fig. 5** The estimated proportion of “more spread” responses as a function of the spread of the dots and the number of dots from the GLMM in Experiment 2. Shading represents 95% CIs. The vertical dashed line represents the POE

dimensions, and the mean range was used in the model. We used the Bayesian information criterion (BIC) as a measure of model fit. A difference in BIC scores of 2 is considered evidence for better model fit. The BIC was substantially lower (better) when we used the spread of the dots than when we



**Fig. 6** PSEs for Experiment 2. The horizontal, dashed lines is at the POE. Lower PSEs are indicative of overestimating the spread of the dots. Error bars are 95% CIs

used range as the independent measure (BICs = 11,274, 11,486, respectively). This suggests participants used the spread (*SD*) of the dots rather than using range as a proxy for judging spread. However, a model that included both spread and range had a slightly better fit (BIC = 11,272) than the model with only spread. This outcome seems most consistent with Lau and Brady (2018) based on our understanding of their results. Although they claimed range serves as a proxy, their results do not support the idea that participants used range instead of variability. Rather, their results suggest that variability was the main driver of participants' responses, but the range of the stimuli affected responses too. Our results also support this idea. There may be other information that participants used as a proxy for spread when making their judgments such as whether the dots are touching each other. Information like contact between dots would change with the number of dots even if the spread itself was held constant. Future studies should explore whether these other sources of information impact variability judgments.

## General discussion

Ensemble perception refers to the visual system's ability to quickly and effectively summarize properties of a group of items such as their mean size or the variability in their orientation (Ariely, 2001; Morgan et al., 2008). Although the visual system is quite adept at perceiving these statistical properties, recent research has uncovered biases in the perception of ensembles. One example, the variability overestimation effect, reveals a tendency for observers to overestimate the variability of a group of stimuli. This bias has been found for the variability of line orientations (Witt, 2019) and with circle color value and size (Warden et al., 2022; Witt et al., 2019). Here, we tested whether the bias would also be found regarding the spread of a cluster of dots. The current studies tested the generalizability of the variability overestimation effect in several ways. The previous studies used sequentially presented stimuli, which raises the issue of whether the bias is due to memory-related processes given the stimuli were not in view when the judgment was made. In addition, the previous studies manipulated object features such as orientation, color value, and size, whereas the current studies manipulated spatial position of the items.

In the experiments reported here, participants judged the spread of a cluster of dots. When there were 10 or 20 dots, we observed a variability overestimation effect: Participants overestimated the spread of the dots. However, when there were 50 dots, no bias was found. The current research is the first to document the variability overestimation effect in static displays with simultaneously presented stimuli. The research is also the first to document this bias with variability defined by the spatial properties between items in the display relative

to variability defined by the feature properties of objects. The visual system is effective at perceiving the mean location of a cluster of dots (Sun et al., 2016, 2018), yet we still observed a bias to overestimate the spread of the dots.

Despite several differences between the current studies and previous ones—such as (1) animated, sequential displays versus static, simultaneous displays, (2) object features versus spatial properties, and (3) the use of a 2AFC that used a reference image versus verbal labels—we still found the variability overestimation effect. This speaks to the generalizability of the effect. However, most experiments on the bias used a bisection task for which participants had to form an implicit midpoint and for which we assumed a linear midpoint. Future research will be needed to determine whether the perceived magnitude of spread is linear. In addition, biases in the measured PSE cannot specify the underlying process, which could be a bias in visual processes, memory-related processes, or even a response bias to select “more spread” more frequently. As with the signal detection measure of the criterion, additional studies are needed to determine the source of the bias (Witt et al., 2015, 2016). For example, although we have no a priori reason for thinking there would be a response bias towards selecting “more spread,” additional studies should be run to determine whether response bias plays a role in the variability overestimation effect.

That the variability overestimation effect is present with multiple visual features (orientation, color value, size), spatial features (spread) and with different types of presentation (sequential, simultaneous) is consistent with the idea that the bias being an inherent part of the visual system and its ensemble processes. The idea that there is an inherent bias to overestimate variability of sets of items raises the question of the underlying mechanism. Here, we consider two possible mechanisms: sampling strategies and the boundary effect.

## Bias via the subsampling strategies

Subsampling refers to the concept that rather than take into account all stimuli within the ensemble, perceivers may process only a subset of the items available. It is unclear what determines the number of items to be selected. With regard to perceiving the mean, it has been suggested that the number of items selected is the square-root of the number of total items (Whitney & Yamanashi Leib, 2018) or perhaps as few as a single item (Marchant et al., 2013). With respect to perceiving mean variability of dot distributions, it has been suggested that performance is equivalent to using 69% of the dots (Sun et al., 2018). This means that given 26 dots in the display, performance attained by human observers could be achieved by calculating the mean position of just 18 of the dots. This subsampling notion is also consistent with Franconeri and colleagues' (Boger et al., 2021; Franconeri et al., 2012) suggestion that all stimuli in static displays may not be processed in

parallel but rather follow a series of rapidly sequential steps in which information is extracted by considering subsets of visual information over time. Indeed, the potentially longer presentation times of static displays may facilitate processing in this manner.

Given that subsampling may be involved in estimation of the mean of a group of items, subsampling could also be involved in estimating the variability. However, it is unclear how a subsampling strategy could account for the variability overestimation effect. When subsampling occurs, the *SD* of the subset may deviate from the *SD* of the full set, but subset *SD* is just as likely to be greater than the full set *SD* as it is to be lesser than the full set *SD*. In other words, the subset *SD* does not deviate from the full set *SD* in systematic ways. As a result, if subsampling occurs, it will lead to attenuated sensitivity to the variability in the ensemble but would not lead to a systematic bias to overestimate variability.

To observe a bias like the one reported here, the sampling would have to be systematic or biased itself. For example, if observers extracted the variability of only the most deviant dots, the variability of the dots at the extreme edges would be greater than the variability of the full set. Thus, a systematically biased subsampling strategy could create a bias to overestimate variability. It is worth noting that one type of strategic subsampling has been documented in the literature for which features closer to the mean are weighted more heavily when estimating the mean (De Gardelle & Summerfield, 2011). This strategy has been referred to as *robust averaging* and has thus far been documented with respect to the mean but not yet to perception of variability. It is unclear how robust averaging could account for the variability overestimation effect given that items closer to the mean are more heavily weighted, which seems it would lead to a bias to underestimate variability rather than overestimate variability as we have found.

### Bias via the boundary effect

Another possible account for the variability overestimation effect is the boundary effect. The boundary effect is the idea that an object's category can bias perception of the object, but that the bias only occurs for stimuli that are located near the boundary of two categories (Huttenlocher et al., 1991; Jazayeri & Movshon, 2007). For example, in one experiment, participants had to judge the direction of a rotating set of dots. There is a natural boundary between clockwise and counterclockwise rotation. When the rotation of the dots was closer to this boundary, there was a bias to judge the dots as rotating further from the boundary. Specifically, when the dots were rotating in a slightly clockwise direction, they were judged as rotating even more clockwise, and when the dots were rotating slightly in a slightly counterclockwise direction, they were

judged as rotating even more counterclockwise (Jazayeri & Movshon, 2007).

The boundary effect has been suggested as a mechanism for why people overestimate variability (Witt, 2019). In a previous study, participants judged the variability in orientation of a set of lines. When the set had low variability, there was a greater bias to overestimate their variability whereas when the set had high variability, there was no bias to overestimate variability. Given the category of *same* versus the category of *different*, a set of lines with low variability would be closer to the boundary between these two categories. A boundary effect would create a bias away from the boundary. When the line orientations were more similar but not identical, the boundary effect would lead to judgments that the set was more variable, pushing it away from the *same* category and towards the *different* category (Witt, 2019). Conversely, when the lines are identical, small perturbations in their perceived orientation due to noise in the visual system would result in them being perceived as having different orientations. The boundary effect was proposed as a way to bias the lines to be perceived as the same, thereby nullifying the effect of noise when perceiving the individual lines on perceiving the set of lines (Morgan et al., 2008).

In the present study, we also only observed an overestimation of variability at low levels, but not an underestimation of variability at high levels. This pattern aligns with predictions of a boundary effect mechanism. The reason the boundary effect only predicts a bias at low levels of variability is because there is a natural boundary between “same” and “different”. At high levels of variability, it is unclear to us what two categories would exist such that there would be a natural boundary between them. For example, “somewhat spread” and “very spread” reflect a difference of magnitude rather than distinct categories per se, so there would not be a boundary between two categories to cause a boundary effect.

In one account of the boundary effect, an object's category serves as an additional source of information. Huttenlocher et al. (1991) suggested that objects are encoded at two levels of representation: the fine-grained details and the object's category. In their experiments, participants judged the location of a single dot located in a circle. Participants spontaneously divided the circle into four quadrants (upper-left, upper-right, lower-left, lower-right) as if a vertical and a horizontal line intersected in the middle of the circle. Judgments of dot locations within the circle were biased away from the boundaries between the quadrants. In other words, the object's categorized quadrant influenced judgments of its position. If the category of the object serves as an additional source of information, that raises the question of how this source of information is weighted by the visual system. It could be that the strength of the source of information about the object's category on the final judgment is a function of the reliability of each source of information.

There are many cases in vision for which reliability dictates how much a given source of information influences the resulting percept (Ernst & Banks, 2002; Ernst & Bühlhoff, 2004; Knill, 2007; Stocker & Simoncelli, 2006; Weiss et al., 2002). The current results are consistent with this explanation. When there were more dots, there was more information about their variability. When there were fewer dots, there was less information about their variability. In other words, the information about variability was less reliable when there were fewer dots than when there were more dots. This model would predict a greater influence of category—a greater boundary effect—when there are fewer dots because that is when the information about variability is less reliable.

Rather than reliability, there could be other reasons why the magnitude of the variability overestimation effect is stronger when there are fewer dots. We previously noted the relationship between the current bias and the concept of support ratio. In many visual illusions, the strength of an illusory percept is directly related to the amount of visible versus illusory visual information. An increase in visible information leads to a less biased percept overall. This was the case in the present experiments as the variability overestimation effect was not observed with high dot counts but was apparent with only 10 to 20 dots in the display. It is possible that at higher dot counts, participants have more information with which to judge variability which in turn makes them more accurate in their appraisal. In the absence of sufficient visible information, perceivers are more susceptible to bias.

Future research is needed to test the mechanism driving the bias to overestimate variability. If the mechanism is due to a boundary effect, it will be important to test whether this is a function of the reliability of the information, a function of the support ratio, or related to something else. For example, a reliability explanation would predict a greater bias to overestimate variability when there is more noise or less clear information in the display (e.g., if the dot contrast was lower or the display was only briefly visible), whereas a support ratio account would predict that the magnitude of bias might be directly linked to the strength and amount of visual information in the display. It is important to note, however, that while we observe the bias at low dot counts, there did not appear to be a difference in the strength of the bias when there was 10 versus 20 dots in the display. It is unclear whether there is a specific threshold of visual information that needs to be passed in order to accurately judge variability, or whether there is a more direct relationship between the degree of overestimation and the amount of information in a display (it was initially posited that we may observe a greater likelihood to overestimate with even fewer dots but that is not consistent with the results of Experiment 2).

That the variability overestimation effect has now been observed with both sequential and simultaneous displays, and as it relates both featural and spatial properties would

seem to suggest a general perceptual bias that may be inherent to the visual system. It is important to acknowledge, however, that we cannot entirely rule out the possibility that memory processes contribute to the bias. Though simultaneous displays do not require the individual to hold array items in memory, working memory is required to remember which button corresponds to which response option and the more/less spread decision is likely influenced by memory of the spread of dots on previous trials. Moreover, Franconeri and colleagues (Boger et al., 2021; Franconeri et al., 2012) have provided evidence that static displays may not be processed strictly in parallel inasmuch as via a series of rapid sequential steps in which information is extracted over time and presumably integrated in working memory. We also cannot rule out a potential contribution of response bias given a difficulty dissociating the perceptual experience from the motor response, though we do not have any reason to believe that response bias would only lead to an overestimation of variability at low levels of variability or with 20 or fewer dots. Future research will be necessary to determine the boundary conditions for the variability overestimation effect and whether other processes beyond vision play any role in the effect.

## Implications

The variability overestimation effect has real-world consequences. With the increasing use of Big Data and visualizations of said data, this could lead to biases in graph comprehension when individuals attempt to estimate the variability in a dataset being shown in a data visualization such as a scatterplot.

Another example relates to diagnostic tests that depend on a radiologist judging the distribution of clusters in a display. In mammography, radiologists must classify the distribution of calcifications in breast tissue to identify cancer (see Fig. 1a). At its heart, this is a visual task and therefore susceptible to biases within the visual system, and a bias to overestimate the variability of object spread could severely impact accurate diagnosis. Specifically, a bias to overestimate the variability of calcifications in breast tissue would lead to more benign diagnoses, which would mean the cancer would be misdiagnosed. This would lead to untreated cancer and any eventual treatments would be delayed.

## Summary

The current study demonstrated a bias to overestimate the variability in positions for a group of dots presented simultaneously. Critically, the variability overestimation effect was present when fewer dots were presented, and biased estimation decreased to the point of being unbiased when 50 dots were presented. The bias to overestimate variability has previously been demonstrated with line orientation, color value,

and size (Warden et al., 2022; Witt, 2019; Witt et al., 2019). The current results are the first demonstration that variability of spread is also prone to a similar bias to overestimate variability. It is also the first demonstration of the variability overestimation effect using static, simultaneous displays. The generalizability of the bias across different kinds of visual features and across static and dynamic displays suggests a bias inherent in the visual system. The current results also have implications for any task for which judging variability is relevant for decision making. For example, the current results suggest the possibility of increased false negative diagnoses for breast cancer due to a visual bias (see Fig. 1). As another example, data visualizations such as scatterplots for which the spread of the points are relevant are likely to be misinterpreted by up to 30%, particularly when there are fewer data points. Any task for which perceiving the variability of a set of items is relevant may be at risk of misinterpretations due to the bias to overestimate variability.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.3758/s13414-022-02528-w>.

**Author note** Jessica K. Witt, Department of Psychology, Colorado State University. Mengzhu Fu and Michael D. Dodd, Department of Psychology, University of Nebraska–Lincoln

Jessica Witt <https://orcid.org/0000-0003-1139-1599>

Stimuli, data, and analysis scripts can be found at <https://osf.io/xgsgjc/>.

This work was supported by grants from the National Science Foundation (BCS-1632222 and SES-2030059) to J.K.W. and NSF/EPSCoR Grant #1632849 to M.D.D.

Correspondence concerning this manuscript should be addressed to Jessica K. Witt, Department of Psychology, Colorado State University, Fort Collins, CO 80523. [Jessica.Witt@colostate.edu](mailto:Jessica.Witt@colostate.edu)

## References

- Albers, D., Correll, M., & Gleicher, M. (2014). Task-driven evaluation of aggregation in time series visualization. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 551–560.
- Ariely, D. (2001). *Seeing Sets: Representation by Statistical Properties*, 12(2), 157–162. <https://doi.org/10.1111/1467-9280.00327>
- Bates, D., Machler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Boger, T., Most, S. B., & Franconeri, S. L. (2021). Jurassic mark: Inattentive blindness for a Datasaurus reveals that visualizations are explored, not seen. *IEEE Visualization Conference (VIS)*, 2021, 71–75.
- Dakin, S. C., & Watt, R. J. (1997). The computation of orientation statistics from visual texture. *Vision Research*, 37(22), 3181–3192.
- De Gardelle, V., & Summerfield, C. (2011). Robust averaging during perceptual judgment. *Proceedings of the National Academy of Sciences*, 108(32), 13341–13346.
- Drew, S. A., Chubb, C. F., & Sperling, G. (2010). Precise attention filters for weber contrast derived from centroid estimations. *Journal of Vision*, 10(10), 20–20.
- Durgin, F. H. (1995). Texture density adaptation and the perceived numerosity and distribution of texture. *Journal of Experimental Psychology: Human Perception and Performance*, 21(1), 149.
- Erlikhman, G., Caplovitz, G. P., Gurariy, G., Medina, J., & Snow, J. C. (2018). Towards a unified perspective of object shape and motion processing in human dorsal cortex. *Consciousness and Cognition*, 64, 106–120.
- Erlikhman, G., Fu, M., Dodd, M. D., & Caplovitz, G. P. (2019). The motion-induced contour revisited: Observations on 3-D structure and illusory contour formation in moving stimuli. *Journal of Vision*, 19(1), 7–7.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415, 429–433.
- Ernst, M. O., & Bühlhoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, 8(4), 162–169.
- Franconeri, S. L., Scimeca, J. M., Roth, J. C., Helseth, S. A., & Kahn, L. E. (2012). Flexible visual processing of spatial relationships. *Cognition*, 122(2), 210–227.
- Friedenberg, J., & Liby, B. (2002). Perception of two-body center of mass. *Perception & Psychophysics*, 64(4), 531–539.
- Haberman, J., Lee, P., & Whitney, D. (2015). Mixed emotions: Sensitivity to facial variance in a crowd of faces. *Journal of Vision*, 15(4), 16.
- Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review*, 98(3), 352.
- Jazayeri, M., & Movshon, J. A. (2007). A new perceptual illusion reveals mechanisms of sensory decoding. *Nature*, 446, 912. <https://doi.org/10.1038/nature05739> <https://www.nature.com/articles/nature05739#supplementary-information>.
- Kanaya, S., Hayashi, M. J., & Whitney, D. (2018). Exaggerated groups: Amplification in ensemble coding of temporal and spatial features. *Proceedings of the Royal Society B: Biological Sciences*, 285(1879), 20172770.
- Kareev, Y., Aron, S., & Horwitz-Zeliger, R. (2002). On the misperception of variability. *Journal of Experimental Psychology: General*, 131(2), 287–297. <https://doi.org/10.1037//0096-3445.131.2.287>
- Knill, D. C. (2007). Learning Bayesian priors for depth perception. *Journal of Vision*, 7(8), 13. <https://doi.org/10.1167/7.8.13>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13).
- Lathrop, R. G. (1967). Perceived variability. *Journal of Experimental Psychology*, 73(4), 498–502.
- Lau, J. S.-H., & Brady, T. F. (2018). Ensemble statistics accessed through proxies: Range heuristic and dependence on low-level properties in variability discrimination. *Journal of Vision*, 18(9), 3–3.
- Lu, V. T., Wright, C. E., Chubb, C., & Sperling, G. (2019). Variation in target and distractor heterogeneity impacts performance in the centroid task. *Journal of Vision*, 19(4), 21–21.
- Marchant, A. P., Simons, D. J., & de Fockert, J. W. (2013). Ensemble representations: Effects of set size and item heterogeneity on average size perception. *Acta Psychologica*, 142(2), 245–250. <https://doi.org/10.1016/j.actpsy.2012.11.002>
- McGowan, J. W., Kowler, E., Sharma, A., & Chubb, C. (1998). Saccadic localization of random dot targets. *Vision Research*, 38(6), 895–909.
- Morgan, M. J., Chubb, C., & Solomon, J. A. (2008). A “dipper” function for texture discrimination based on orientation variance. *Journal of Vision*, 8(11), 9.
- Moscattelli, A., Mezzetti, M., & Lacquaniti, F. (2012). Modeling psychophysical data at the population-level: The generalized linear mixed model. *Journal of Vision*, 12(11), 1–17. <https://doi.org/10.1167/12.11.26>
- Norman, L. J., Heywood, C. A., & Kentridge, R. W. (2015). Direct encoding of orientation variance in the visual system. *Journal of Vision*, 15, 3.
- Penney, T. B., Gibbon, J., & Meck, W. H. (2008). Categorical scaling of duration bisection in pigeons (*Columba livia*), mice (*Mus musculus*),

- and humans (*Homo sapiens*). *Psychological Science*, 19(11), 1103–1109.
- R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for statistical computing. <https://www.R-project.org/>
- Rashid, J. A., & Chubb, C. (2021). The density effect in centroid estimation is blind to contrast polarity. *Vision Research*, 186, 41–51. <https://doi.org/10.1016/j.visres.2021.04.005>
- Raslear, T. G. (1985). Perceptual bias and response bias in temporal bisection. *Perception & Psychophysics*, 38(3), 261–268.
- Semizer, Y., & Boduroglu, A. (2021). Variability leads to overestimation of mean summaries. *Attention, Perception, & Psychophysics*, 83(3), 1129–1140.
- Shipley, T. F., & Kellman, P. J. (1992). Strength of visual interpolation depends on the ratio of physically specified to total edge length. *Perception & Psychophysics*, 52(1), 97–106.
- Solomon, J. A., Morgan, M. J., & Chubb, C. (2011). Efficiencies for the statistics of size discrimination. *Journal of Vision*, 11(12), 13.
- Stocker, A. A., & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*, 9, 578–585.
- Sun, P., Chubb, C., Wright, C. E., & Sperling, G. (2016). Human attention filters for single colors. *Proceedings of the National Academy of Sciences*, 113(43), E6712–E6720.
- Sun, P., Chubb, C., Wright, C. E., & Sperling, G. (2018). High-capacity preconscious processing in concurrent groupings of colored dots. *Proceedings of the National Academy of Sciences*, 115(52), E12153–E12162.
- Szafir, D. A., Haroz, S., Gleicher, M., & Franconeri, S. (2016). Four types of ensemble coding in data visualizations. *Journal of Vision*, 16(5), 11. <https://doi.org/10.1167/16.5.11>
- Warden, A. C., Witt, J. K., Fu, M., & Dodd, M. D. (2022). *Overestimation of variability in ensembles of line orientation, size and color* Manuscript under review.
- Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, 5(6), 598–604.
- Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual Review of Psychology*, 69, 105–129. <https://doi.org/10.1146/annurev-psych-010416-044232>
- Witt, J. K. (2019). The perceptual experience of variability in line orientation is greatly exaggerated. *Journal of Experimental Psychology: Human Perception and Performance*, 45(8), 1083–1103. <https://doi.org/10.1037/xhp0000648>
- Witt, J. K., Dodd, M. D., & Edney, E. (2019). The perceptual experience of orientation variability. *Journal of Vision*, 19(10), 193a–193a.
- Witt, J. K., & Sugovic, M. (2010). Performance and ease influence perceived speed. *Perception*, 39(10), 1341–1353. <https://doi.org/10.1068/P6699>
- Witt, J. K., Taylor, J. E. T., Sugovic, M., & Wixted, J. T. (2015). Signal detection measures cannot distinguish perceptual biases from response biases. *Perception*, 44(3), 289–300.
- Witt, J. K., Taylor, J. E. T., Sugovic, M., & Wixted, J. T. (2016). Further clarifying signal detection theoretic interpretations of the Müller-Lyer and sound-induced flash illusions. *Journal of Vision*, 16(11), 19. <https://doi.org/10.1167/16.11.19>
- Witt, J. K., Warden, A. C., Dodd, M. D., & Edney, E. C. (2022). Visual bias could impede diagnostic accuracy of breast cancer calcifications. *Journal of Medical Imaging*, 9(3), 035503. <https://doi.org/10.1117/1.JMI.9.3.035503>
- Zosky, J. E., & Dodd, M. D. (2021). The Z-box illusion: Dominance of motion perception among multiple 3D objects. *Psychological Research*, 1–15.

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