Five dichotomies in the psychophysics of ensemble perception

Joshua A. Solomon (J.A.Solomon@city.ac.uk)

Centre for Applied Vision Research, City, University of London EC1V 0HB

Abstract

1. Whereas psychophysicists may formulate hypotheses about appearance, they can only measure performance. Bias and imprecision in psychophysical data need not necessarily reflect bias and imprecision in perception.

2. Sensory systems may exaggerate the differences between each item and its neighbors in an ensemble. Alternatively, sensory systems may homogenize the ensemble; thereby removing any apparent differences between neighboring items.

3. Ensemble perception may be involuntary when observers attempt to report the identities of individual items. Conversely, when asked to make a (voluntary) decision about the ensemble as a whole, observers may find it very difficult to compute statistics that are based on more than a very small number of individual items.

4. Modeling decisions about prothetic continua like size and contrast can be tricky, because sensory signals may be distorted before and/or after voluntarily computing ensemble statistics. With metathetic continua, like spatial orientation, distortion is less problematic; physically vertical things necessarily appear close to vertical and physically horizontal things necessarily appear close to horizontal.

5. Decision processes are corrupted by noise that, like distortion, may be added to sensory signals prior to and/or after voluntarily computing ensemble statistics.

The ginkgo leaves fell like fine rain from the boughs and dotted the lawn with yellow….I said I would like to distinguish the sensation of each single ginkgo leaf from the sensation of all the others, but I was wondering if it would be possible.
–Italo Calvino, *If on a winter's night a traveler*, 1979

There are two types of people: Those that see dichotomies everywhere, and I'm one of them. It is my hope that some of what follows may serve as a tutorial to students of ensemble perception. Readers of this listicle should be forewarned: most of the examples provided below were produced in my own laboratory. Whitney and Yamanashi-Leib (2018) offer a more balanced review.

Appearance v Performance

The epigraph suggests that individual sensations may depend on whether they form part of an ensemble. The author of this passage is a character in Calvino's book. His suggestion is almost certainly correct, but I want to underscore the fact that psychophysicists are fundamentally incapable of directly measuring sensation. Physiologists can directly measure neural output, but all a psychophysicist can do is measure visual performance. The ensemble may bias our perception of any individual ginkgo leaf, but bias and imprecision in psychophysical data need not necessarily reflect bias and imprecision in perception.

As an example, consider Brady and Alvarez's (2011) investigation of the effect of ensemble membership on reproduction error. On each trial of their experiment, human observers re-sized a circle (with a computer mouse) to match one member of a previously seen ensemble of circles. Errors tended to be in the direction of the ensemble mean, but Brady and Alvarez did not unnecessarily infer any homogenization of perceived sizes within the ensemble. Although each individual circle's size may have appeared closer to that of the ensemble mean, Brady and

Alvarez concluded that reproductions were consistent with a type of rational decision-making, under uncertainty, given the limited capacity of visual working memory.

One of the most pernicious pitfalls in attempting to infer appearance from behavior is the subject-expectancy effect (e.g. Rosenthal & Rubin, 1978), in which experimenter expectations are transmitted to the participants in such a way as to alter the outcome of the experiment in a desired direction. There is, perhaps, no way to preclude non-perceptual biases from contaminating psychophysical estimates of appearance, but certain measures can be taken to minimize their likelihood and/or severity.

One strategy is to use a psychophysical paradigm in which observers' response categories are de-coupled from experimenter expectations. For example, Morgan, Melmoth, & Solomon (2013) expected that vertically aligned spots would appear to have a clockwise tilt, when they were displayed within a square frame having a counter-clockwise tilt. Rather than asking observers to report the apparent tilt of the spots, they asked observers to inspect two sets of spots and report which set was closest to vertical. Sometimes both sets were clockwise, sometimes both sets were counter-clockwise, and sometimes one set was clockwise of vertical while the other set was counter-clockwise. Variants of this "comparison-of-comparisons" strategy have become increasingly popular (e.g. Maloney & Yang, 2003; Jogan & Stocker, 2004; Morgan, Grant, Melmoth, & Solomon, 2015; Morgan, Schreiber, & Solomon, 2016; Yarrow, Martin, Di Costa, Solomon, & Arnold, 2016; Ismail, Solomon, Hansard, & Mareschal, 2016).

Also popular are "double-blind" strategies, where steps are taken to remove the bias-inducing stimuli from the observer's awareness. For example, it is possible to observe a tilt aftereffect after adapting to gratings with unresolvably high spatial

frequency (He & MacLeod, 2001). The tilt aftereffect also survives removal of the

adapting stimulus from awareness by crowding (defined below; He, Cavanagh, &

Intrilligator, 1996) and critical-flash suppression (Bahrami, Carmel, Walsh, Rees, &

Lavie, 2008).

One final strategy is to make performance depend upon appearance.

Inferences regarding the latter can be made from objective measurements of the

former. For example, Morgan and Solomon (2019) asked observers to select the odd-

man-out in a variety of visual search tasks, where performance is determined by the

target's relative salience, i.e. how different it appears from the distractors. Adaptation

was used to manipulate target salience, and aftereffect strength was inferred from the

degree to which performance was facilitated or impaired by adaptation.

Repulsion v Assimilation

Although it may be relatively trendy to consider how ensemble membership affects

the appearance of its individual members, studies relating an item's appearance to its

general spatial and temporal context are nothing new. The motion aftereffect, for

example, was described by Addams in 1834, who reported that the nearby rocks

appeared to move upwards, after adapting to the downward motion of the Falls of

Foyers. In 1855 (p. 8) Chevreul described the exaggeration of simultaneous

luminance contrast between neighboring stripes with different graylevels. The general

term for these sensory phenomena is "repulsion," and it occurs in every feature

domain I have examined.

Figure 1 (top panel) shows an example of the visual system exaggerating the

simultaneous orientation contrast between neighboring Gabor patterns. The truly

vertical central elements on either side of the fixation cross should appear to be tilted

in opposite directions when you stare directly between them. This can be described as an exaggeration of the difference between the orientation of each central element and the orientation of its flanking distractors. This exaggeration is usually pretty large for briefly presented stimuli (i.e., short flashes; Wenderoth & van der Zwan, 1989), but it also depends on the retinal eccentricity of the central Gabor patterns.
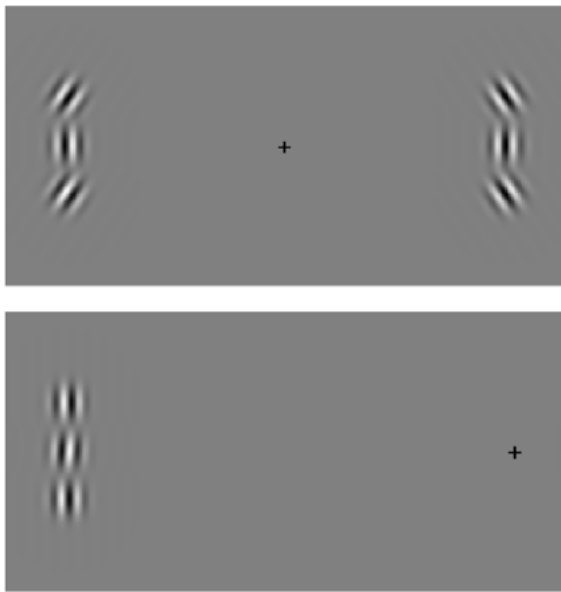

Fig. 1. Demonstrating repulsion and assimilation. The visual system exaggerates the orientation contrast between the (45°) tilted flankers and the vertical elements in the top panel. Small orientation differences, on the other hand, are effectively squelched. Thus, vertical flankers in the bottom panel can cause the small (4°) tilt to go unnoticed.

Exaggerations of simultaneous contrast are routinely ascribed to lateral inhibition somewhere within the visual system. The cartoon at the top of Figure 2 represents 15 receptive fields somewhere in visual cortex, with lateral inhibition spreading around the cell whose receptive field is red. If lateral inhibition is responsible for the exaggeration of simultaneous contrast, then we have psychophysical data suggesting that lateral inhibition spreads very far indeed. Consider the red symbols in the lower half of Figure 2. They represent trials in which

flankers were tilted either 22° degrees clockwise or 22° degrees counter-clockwise of horizontal. When these flankers were present, they produced a negative or "repulsive" bias of roughly 10° when observers attempted to report the orientations of near-horizontal targets. This was true even when the distance between flanker and target was almost half the latter's eccentricity.
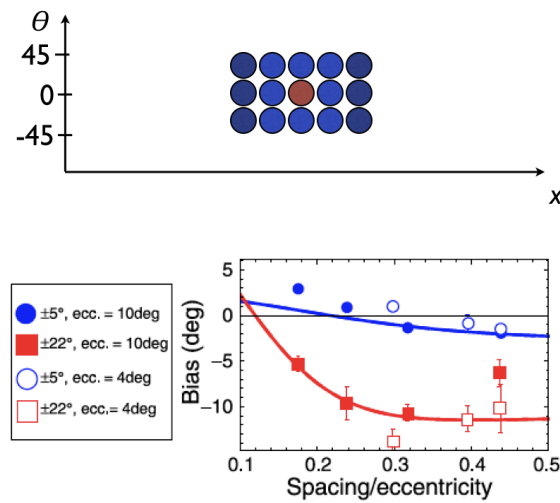

Fig. 2. Repulsion and assimilation: theory and experiment. Top panel contains a cartoon of 15 "classical" receptive fields somewhere in visual cortex. Axes are one dimension of visual space (e.g. horizontal position x) and orientation preference θ. Horizontal stimuli positioned so as to maximally stimulate the central (red) cell indirectly and strongly inhibit nearby (light blue) cells and weakly inhibit the somewhat further (dark blue) cells, via lateral connections. Reciprocality in those connections means that the latter cells can also inhibit the central one. Thus, the amalgamation of these 15 receptive fields can be considered the "extra-classical" receptive field of the central cell. The lower panel shows average biases (symbols; ± 1 standard error) and model fits (solid and dashed curves) to Experiment 3 from Mareschal, Morgan, & Solomon (2010), in which observers judged the orientation (with respect to a horizontal reference) of a flanked Gabor pattern at 4 and 10 degrees of retinal eccentricity. All flankers were identically rotated either –22°, –5°, +5°, or +22° with respect to the horizontal.

The blue symbols, on the other hand, represent trials in which the flankers were tilted just 5° off horizontal. When those flankers were close to the target, they seemed to produce a positive bias. This assimilation means that the visual system effectively makes it hard to notice subtle changes in orientation (and any other feature dimension you care to name), particularly when you don't look directly at the

stimulus. Orientation assimilation is demonstrated in Figure 1B. You might need to look slightly to the left or right of the black cross, but you should be able to find a fixation for which you can still see all three Gabors without being able to see the differences in their orientations.

Involuntary Averaging v Voluntary Averaging

Whatever is responsible for small-angle assimilation, it too extends over a vast swathe of visual cortex. The only reason small-angle assimilation disappears when flankers and target are widely separated is because lateral inhibition is stronger. To get around the problem of lateral inhibition, some researchers have adopted tasks that should be immune to it. That is, they use visual tasks that do not require making fine discriminations between similar orientations or luminances or whatever. One task they particularly like is letter identification.

The idea behind Figure 3 is that the visual system is compelled to combine information from discrete stimuli, and consequently cannot disentangle the identity of one individual element from the overall statistics of information coming from that region of the visual field. In general, the name given to this phenomenon is "crowding," and experiments with letter identification suggest a compulsory combination of visual information from individual letters separated by anything up to half their average eccentricity (Bouma, 1970).

Fig. 3. Crowding. When fixating the dot on the top left, the N to the right is easily identified. When fixating the dot just below, the N seems to meld into a jumbled texture. Adapted from Cavanagh (2001).

Parkes, Angelucci, Lund, Solomon, & Morgan (2001) described crowding as "textural processing when we don't want it to happen." In other words, crowding is a form of involuntary ensemble representation. It is evident only when the visual task requires a decision about individual visual objects. I would like to distinguish such tasks from those that require a decision about the ensemble as a whole. In such cases, the computation of ensemble statistics may very well be voluntary.

When investigating ensemble representations, most of my work has been on voluntary averaging. For example, Solomon and Morgan (2018) recently assessed how well observers could estimate the average number of dots in a dot patch. The metric we used was efficiency. The paradigm was two (temporal)-alternative forced-choice: observers were asked whether the dot patches in the first ensemble had greater or fewer dots on average than patches in a second ensemble. Obviously, if both ensembles had the same number of patches, then the question would be identical to asking which stimulus had more dots. However, in some blocks of trials, the first ensemble had eight patches while the second had just four patches. We also had blocks of four followed by eight. A Binomial distribution was sampled to determine the number of dots in each patch. To compute efficiency, you need to calculate the

ideal performance with various fractions of the available information. For example, if human performance were similar to what an ideal observer could do with half the available information, then human efficiency would be 50%.

Figure 4A shows a flow-diagram of the Inefficient Observer model for voluntary averaging. There were N dot clusters, observer JAS selected M of these at random, measured their numerosities, and reported the mean of those M measurements. M is known as the effective sample size and the ratio of $M:N$ is known as either the sampling efficiency or the calculation efficiency or simply "the efficiency." For JAS M was about 3.
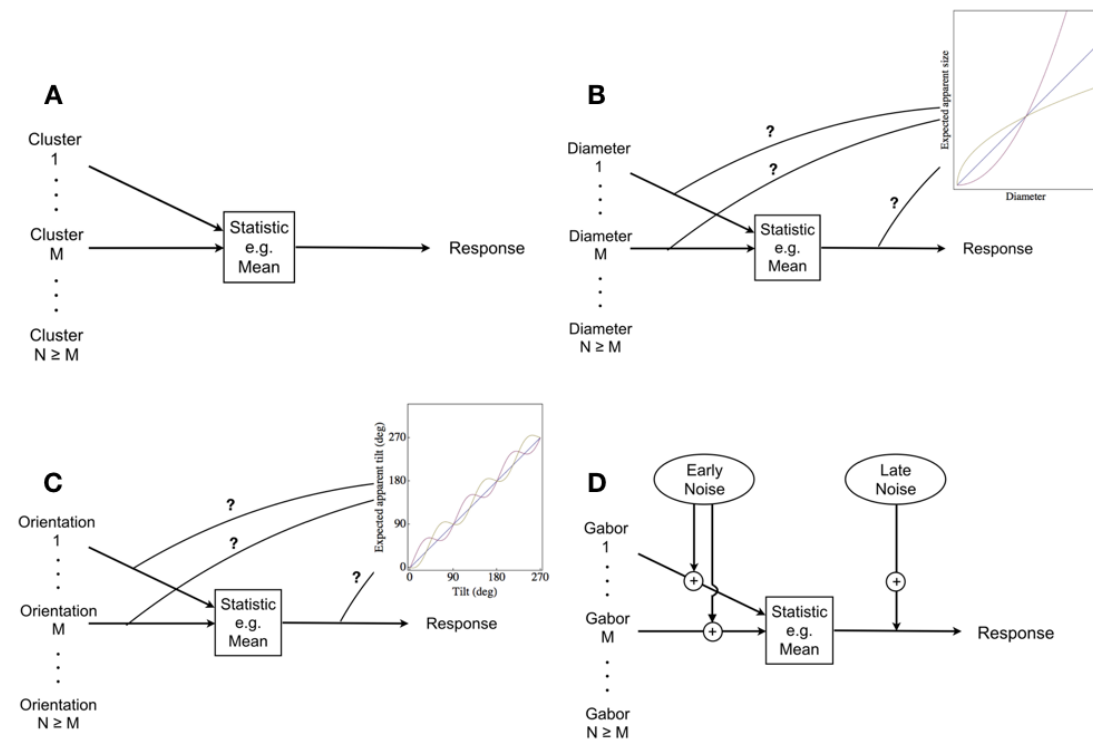
Fig. 4. The Inefficient Observer model applied to voluntary averaging of (A) numerosities, (B) sizes, and (C) orientations. Psychophysical transformations (from physical dimensions into visual signals) are shown in the insets of panels B and C. Note these transformations (expansive, compressive, or quasi-linear) may occur prior to or after statistical summarization. Panel (D) shows the Noisy Inefficient Observer model with both (pre-summarization) early and (post-summarization) late sources of noise.

Prothetic v Metathetic Continua

The inefficient observer model can be traced back to one (Dakin, 2001) of the two papers that massively popularized the study of ensemble statistics. Dakin studied voluntary averaging of orientation and Ariely (2001) studied voluntary averaging of size. Subsequently, I studied both, but orientation is easier, because it is a metathetic continuum.

The distinction was made by Stevens (1957). Simply put, prothetic continua are those like size and contrast, where one value is either greater than, equal to, or less than another. Metathetic continua like orientation and direction don't work like that.[1] Prothetic continua like size can be tricky to work with, because apparent size may not scale linearly with physical size. Indeed, without sufficiently specific instructions (cf. Raidvee, Toom, Averin, & Allik, *in press*) it isn't necessarily clear whether average size reflects average diameter, average area, or something in between.

Moreover, the transformation between physical size and apparent size could theoretically occur before and/or after voluntarily computing whatever statistic we wanted to study. The inset in Figure 4B shows three psychophysical (transducer) functions. The blue one is linear, the violet one is expansive, and the yellow one is compressive.

Solomon & Tyler (2018) studied voluntary averaging in the contrast domain. When asked to adjust the contrast of a static, black-and-white texture to match the average contrast of another black-and-white texture that was flickering between high and low contrast, all observers gave the static texture greater average contrast than the

---

[1] Regarding this dichotomy, here is the exact quote: "Continua having to do with *how much* belong to what we have called Class I, or prothetic; continua having to do with *what kind* and *where* (position) belong to Class II, or metathetic" (italics as in Stevens, 1957, p. 154).

flickering texture. This result is consistent with an accelerating or expansive transformation of physical contrast, prior to voluntary averaging.

With metathetic continua, like orientation, transformations like this don't really matter, because we can be pretty confident that physically vertical things appear close to vertical and physically horizontal things appear close to horizontal. Thus the psychophysical function for orientation may stray a little from the blue line of veridicality in Figure 4C, but it can't stray too far.

Early v Late Noise

Perhaps the most direct way to test how well an observer can voluntarily compute an average is to ask the observer to reproduce that average. Solomon & Tyler (2015) showed observers an ensemble comprised of eight low-frequency Gabor patterns in a ring around fixation (see Figure 5A). Their orientations were drawn from a Wrapped Normal distribution. The mean of that distribution was totally random, but its standard deviation was either zero or 16°. The ensemble was visible for 0.42 s. A high-frequency, "probe" Gabor was presented at fixation after the ensemble had disappeared. Observers rotated the probe until its orientation matched their memory of the ensemble's average orientation.
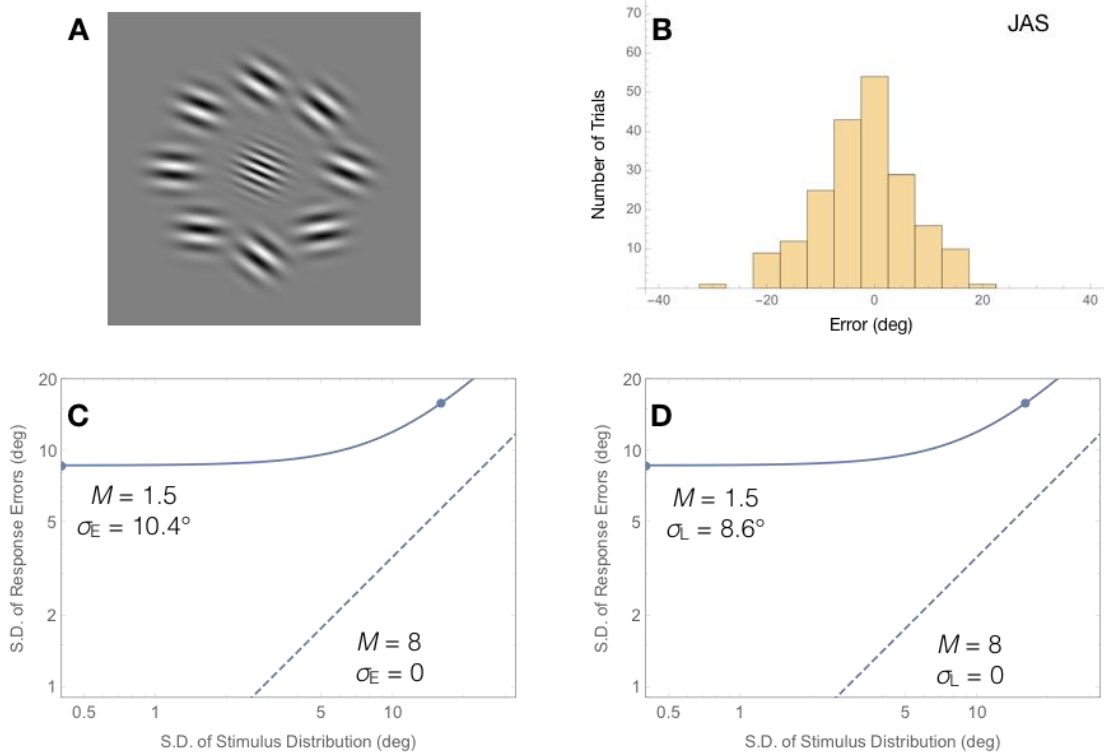
Fig. 5. Voluntary averaging of orientation. (A) Stimulus from Solomon & Tyler (2015). Observers adjusted the central probe to match their memory of the surrounding ensemble's average orientation. (B) Histogram of a representative observer's adjustment errors when the SD of low-frequency Gabor orientations was zero. (C, D) Threshold v Noise (in this case, the SD of adjustment errors v the SD of the distribution of ensemble orientations) plots, fit with the Noisy, Inefficient Observer model assuming zero late noise (i.e. $\sigma_L = 0$; C) and zero early noise (i.e. $\sigma_E = 0$; D). Smooth curves are maximum-likelihood fits to thresholds (dots) from JAS. Dashed lines illustrate the ideal observer.

Although the Inefficient Observer model (Figure 4A) is useful in some cases; in general, it is not a good model of human performance, because humans cannot measure feature values like orientation with infinite precision. If they could, then response errors would have dropped to zero when all eight of the low-frequency Gabors were parallel. Errors do not drop to zero, as can be seen in Figure 5B. Thus the Inefficient Observer model needs modification. Perhaps the most straightforward way to modify the model is to assume that each measurement made by the observer gets perturbed by an independent, identically distributed sample of noise, as illustrated in Figure 5C.

12

This Noisy, Inefficient Observer model for voluntary averaging (and other summary statistics) has two free parameters: the effective set-size M, and the standard deviation of noise σ_E:

$$\text{error S.D.}^2 = \frac{\sigma_E^2 + \text{stimulus S.D.}^2}{M}. \tag{1}$$

Within this simple formula, the standard deviation of response errors increases as a function of the standard deviation of orientations in the ensemble. A fit to one observer's data is shown in Figure 5C's threshold-v-noise plot. The axes are logarithmic, so data from the condition with parallel Gabors appears at the leftmost edge of the plot. When the model is best-fit to these two data points, its noise has a standard deviation of 10.4 degrees. The effective set size is 1.5.

Non-integer values such of effective set size can be interpreted in two ways. One possibility is that they reflect a mixture of effective set sizes. For example, on some trials I might have ignored or forgotten all but one randomly selected element, whereas on other trials I computed the average of two or more. In this case, M would reflect the root-mean-square of this number. Alternatively, non-integer values of M might reflect an unequal weighting of two or more of the texture elements.[2]

The smooth curve in Figure 5D illustrates a slightly different Noisy, Inefficient Observer:

$$\text{error S.D.}^2 = S_L^2 + \frac{\text{stimulus S.D.}^2}{M}. \tag{1}$$

In this case, it's as though the entire ensemble gets tilted by a random amount, rather than each individual element. The standard deviation of that random amount σ_L is

---

[2] The Inefficient Observer model can be considered an extreme form of unequal weighting, in which $N - M$ elements get zero weight. Whereas any form of unequal weighting is sub-ideal and thus inefficient, equal weighting does not imply 100% efficiency. An observer whose responses on each trial are based on a single, randomly selected element will have an efficiency of $1/M$, but – on average – weight will have been distributed equally over all elements in the sample.

called "late noise" in this formulation.[3] The first thing to notice is that the two versions of the model make exactly the same predictions. The second thing to notice is that the best-fitting value of *M* does not depend on whether individual samples of equivalent noise are correlated. Thus, if all you care about is the effective set size, you can use either formula, it doesn't matter.

I attempted to ascertain whether the noise that limited the precision of orientation statistics was early or late (Solomon, 2010). This would have been impossible if I had merely asked obserers to judge the average orientation of ensembles having a fixed size. Consequently, I varied the size of the ensemble, taking care not to change the precision with which the orientation of each individual element could be assessed, and I also varied the task: in some trials, observers were asked to assess the variance of orientations in the ensemble, rather than the mean. The results contained compelling evidence for both early and late noise. When members of an ensemble are all roughly identical, any effect of sample size *N* on the precision of voluntary averaging must be attributed to early noise. On the other hand, late noise is implicated when the effect of the sample's size increases with the sample's variance.

In conclusion, I would like to underscore two features/caveats with regard to the efficiency metric. First, it is a purely descriptive measure of performance in relation to the available information. Given any sample size *N*, efficiency is the ratio of *M* to *N*, where *M* is the sample size that the ideal observer would need in order to estimate a

---

[3] Note that segregating the cause of imprecision into an early and late stage is somewhat arbitrary. If, instead of late noise, we allow the random perturbation added to the orientation of any two elements to have correlation ρ, then we can reparameterize Equations 1 and 2 such that:

$$\text{error S.D.}^2 = \frac{\rho(M-1)}{M}\sigma^2 + \frac{\sigma^2 + \text{stimulus S.D.}^2}{M}. \tag{3}$$

Note that when $\rho=0$, Equation 3 reduces to Equation 1 with $\sigma = \sigma_E$; when $\rho=1$, Equation 3 reduces to Equation 2 with $\sigma = \sigma_L$.

statistic with the same precision as a human observer. Since the ideal observer is, by definition, at least as good as any other observer, human decisions are necessarily based on *M* or more elements. Second, and most important, estimates of efficiency tell us nothing about ensemble perception *per se*. With strategic decision-making strategies, reasonably high efficiencies can be achieved from independent representations of individual stimuli. Authors discussing various strategies include: Myczek & Simons (2008), Solomon (2010), Dayan & Solomon (2010), Gorea, Belkoura, & Solomon (2014), and Li, Herce Castañon, Solomon, Vandormael, & Summerfield (2017).

References

Addams, R. (1834). An account of a peculiar optical phenomenon seen after having looked at a moving body etc. London and Edinburgh Philosophical Magazine and Journal of Science, 5, 373–374 (3rd series).

Ariely, D. (2001). Seeing sets: representation by statistical properties. Psychological Science, 157–162.

Bahrami, Carmel, Walsh, Rees, & Lavie (2008). Spatial attention can modulate unconscious orientation processing. Perception, 37, 1520–1528.

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. Nature, 226, 177–178.

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical Encoding in Visual Working Memory: Ensemble Statistics Bias Memory for Individual Items. Psychological Science, 22, 384–392.

Cavanagh, P. (2001). Seeing the forest but not the trees. Nature Neuroscience, 4, 673–674.

Chevreul, M. E. (1855). Harmony and contrast of colours, and their application to the arts. Translated by C. Martel. London: Longman, Brown, Green, and Longmans.

Dakin, S. C. (2001). Information limit on the spatial integration of local orientation signals. Journal of the Optical Society of America A. 18, 1016–1026.

Dayan, P., & Solomon, J. A. (2010). Selective Bayes: Attentional load and crowding. Vision Research. 50, 2248–2260.

Gorea, A., Belkoura, S., & Solomon, J. A. (2014). Summary statistics for size over space and time. Journal of Vision. 14(9):22, 1–14.

He, S., Cavanagh, P., & Intrilligator, J. (1996). Attentional resolution and the locus of visual awareness. Nature. 383, 334–337.

He, S., & MacLeod, D. (2001). Orientation-selective adaptation and tilt after-effect from invisible patterns. Nature, 411, 473–476.

Ismail, A. M. H., Solomon, J. A., Hansard, M., & Mareschal, I. (2016). A tilt after-effect for images of buildings: Evidence of selectivity for the orientation of everyday scenes. Royal Society Open Science, 3: 16–551.

Jogan, M., & Stocker, A. A. (2014). A new two-alternative forced choice method for the unbiased characterization of perceptual bias and discriminability. Journal of Vision, 14(3):20, 1–18.

Li, V. Herce Castañon, S., Solomon, J. A., Vandormael, H., & Summerfield, C. (2017). Robust averaging protects decisions from noise in neural computations. PLoS Computational Biology, 13(8) e1005723.

Maloney, L. T., & Yang, J. N. (2003). Maximum likelihood difference scaling. Journal of Vision, 3, 573–585.

Mareschal, I., Morgan, M. J. & Solomon, J. A. (2010). Cortical distance determines whether flankers cause crowding or the tilt illusion. Journal of Vision, 10(8):13, 1–14.

Morgan, M., Grant, S., Melmoth, D., & Solomon, J. A. (2015). Tilted frames of reference have similar effects on the perception of gravitational vertical and the planning of vertical saccadic eye movements. Experimental Brain Research, 233, 2115–2125.

Morgan, M. J., Melmoth, D., & Solomon, J. A. (2013). Linking hypotheses underlying Class A and Class B methods. Visual Neuroscience, 30, 197–206.

Morgan, M. J., Schreiber, K., & Solomon, J. A. (2016). Low-level mediation of directionally specific motion after-effects: motion perception is not necessary. Attention, Perception, & Psychophysics, 78, 2621–2632.

Morgan, M. J., & Solomon, J. A. (2019). Attention and the motion aftereffect. Quarterly Journal of Experimental Psychology, 72, 2848–2864.

Myczek, K., & Simons, D. J. (2008). Better than average: Alternatives to statistical summary representations for rapid judgments of average size. Perception & Psychophysics, 70, 772–788.

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A. & Morgan, M. J. (2001). Compulsory averaging of crowded orientation signals in human vision. Nature Neuroscience, 4, 739–744.

Raidvee, A., Toom, M., Averin, K., & Allik, J. (*in press*). Perception of means, sums, and areas. Attention, Perception, & Psychophysics.

Rosenthal, R., & Rubin, D. B. (1978). Interpersonal expectancy effects: The first 345 studies. The Behavioral and Brain Sciences, 1, 377–386.

Stevens, S. S. (1957). On the psychophysical law. The Psychological Review, 64, 153–181.

Solomon, J. A. (2010). Visual discrimination of orientation statistics in crowded and uncrowded arrays. Journal of Vision, 10(14):19, 1–16.

Solomon, J. A., & Morgan, M. J. (2018). Calculation efficiencies for mean numerosity. Psychological Science, 29, 1824–1831.

Solomon, J. A. & Tyler, C. W. (2015) Weight allocation in summary statistics. Perception, 44 Supplement, 248–249.

Solomon, J. A. & Tyler, C. W. (2018). A Brücke-Bartley effect for contrast. Royal Society Open Science, 5: 180171.

Wenderoth, P., & van der Zwan, R. (1989). The effects of exposure duration and surrounding frames on direct and indirect tilt aftereffects and illusions. Perception & Psychophysics, 46, 338–344.

Whitney, D., & Yamanashi-Leib, A. (2018). Ensemble perception. Annual Review of Psychology, 69, 12.1–12.25.

Yarrow, K., Martin, S. E., Di Costa, S., Solomon, J. A., & Arnold, D. H. (2016). A roving dual-presentation simultaneity judgment task to estimate the point of subjective simultaneity. Frontiers in Psychology, 7:416, 1–19.