Attention, Perception, & Psychophysics (2022) 84:981–991
https://doi.org/10.3758/s13414-022-02456-9

How are local orientation signals pooled?

Jüri Allik[1,2] · Mai Toom[1] · Richard Naar[1] · Aire Raidvee[1]

Accepted: 3 February 2022 / Published online: 2 March 2022

Abstract

Visual perception is capable of pooling multiple local orientation signals into a single more accurate summary orientation. However, there is still a lack of systematic inquiry into which summary statistics are implemented in that process. Here, the task was to recognize in which direction, clockwise or counter-clockwise, the mean orientation of a set of randomly distributed Gabor patches ($N = 1$, 2, 4, and 8) was rotated from the implicit vertical. The mean orientation discrimination accuracy did not improve with the increase of the number N of elements in proportion to the square-root-N, as could be expected if noisy internal representations were arithmetically averaged. The *Proportion of Informative Elements* (*PIE*), defined as the percentage of elements having an orientation different from the vertical, also affected the discrimination precision, violating the arithmetic averaging rules. The decrease in the orientation discrimination precision with the increase of the *PIE* would suggest that the orientation pooling could be more adequately described by a quadratic or higher power mean. Thus, we parameterized the averaging process for the power parameter of the generalized mean formula. The results indicate that different pooling rules in different trials may apply, for example, the arithmetic mean in some and the maximal deviation rule in others. It is concluded that pooling of orientation information is a relatively inaccurate process for which different perceptual cues and their combination rules can be used.

Keywords Pooling orientation · Ensemble perception · Statistical averaging · Representational noise · Proportion of informative elements · Generalized mean

Introduction

One of the most fundamental discoveries about the human mind was the magical number seven (Miller, 1956), later reduced to four (Cowan, 2015), characterizing the amount of information that can be received, processed, and remembered. As expected, the discovery of a mechanism that can somehow bypass this limit was enthusiastically welcomed (Ariely, 2001; Chong & Treisman, 2003, 2005a, 2005b). A large number of studies that followed these attempts to identify mechanisms capable of computing the mean values of various perceptual attributes across a large number of elements were united into a vigorously growing research field usually called ensemble perception (Alvarez, 2011; Baek & Chong, 2020; Bauer, 2015; Bayne & McClelland, 2019; Whitney et al., 2021; Whitney & Leib, 2018). One of the main incentives for the study of ensemble perception was to discover smart perceptual mechanisms that can cope with limited information processing capacity (cf. Baek & Chong, 2020).

A suggestion that these smart mechanisms exist were primarily based on an observation that the discrimination of the mean value of some perceptual attribute, usually size or orientation, was roughly constant and changed little with the total number N of elements. For example, it was observed that the accuracy of the mean size discrimination remained approximately constant with the increase of the set size from 4 to 16 (Ariely, 2001, Fig. 4). Subsequent studies confirmed the observation that the accuracy of the mean size discrimination is typically independent of the number of elements in the set (Allik et al., 2013; Chong & Treisman, 2005b). This independence was considered as evidence that the number of elements does not affect the accuracy with which the mean value of a perceptual attribute can be determined (Alvarez, 2011; Ariely, 2001; Chong & Treisman, 2005b). However, this was a mistake because simple statistical considerations could lead to a conclusion that if the information from N elements is pooled together arithmetically, the accuracy is expected to increase proportionally with the square root of N (Allik et al., 2013;

✉ Jüri Allik
juri.allik@ut.ee

[1] Institute of Psychology, University of Tartu, Näituse 2, 50409 Tartu, Estonia

[2] Estonian Academy of Sciences, Tallinn, Estonia

Fouriezos et al., 2008; Sorkin et al., 1991). Thus, a nearly constant discrimination precision demonstrates, in fact, a significant drop in the accuracy with which each additional element is processed. Strictly speaking, this also means that the pooling of information from multiple elements did not follow the rules of simple arithmetic addition.

However, there is one important subtlety. The square-root-*N* rule holds only if all *N* elements are noticed and processed. It was demonstrated that deviations from the square-root-*N* rule could also be explained by inattentional feature blindness (Myczek & Simons, 2008). The observed discrimination precision of the mean value can be obtained assuming that only two to three elements out of *N* are attended, and information recorded from them is pooled together for finding their mean size (Myczek & Simons, 2008). However, inattentional feature blindness cannot be separated from the representational noise based on the discrimination precision alone (Allik et al., 2013). Because these two factors have a trade-off effect on the discrimination precision, the slope of the psychometric discrimination function or any of its characteristics is an ambiguous indicator (Allik et al., 2013). To resolve this conundrum, a novel method for separating the effects of inattentional feature blindness from perceptual noise was recently proposed (Raidvee et al., 2021). This method is based on presenting only one single informative element among "dummies" that carry no useful information for solving a given perceptual task. If the single informative element goes unprocessed, the correct answer can only be given by a random guess. This random guessing rate can be modelled using a lapse rate parameter λ, which changes the shape of the psychometric discrimination function, not only its slope or standard deviation (Raidvee et al., 2021). By evaluating this deformation of the psychometric discrimination function it is possible to determine how often the critical element is missed independently from the accuracy with which a perceptual attribute can be perceptually represented. This study provided a strong endorsement that inattentional feature blindness played no significant role in the perception of the mean orientation when stimuli were identical to the study we will present here. Even when the maximal number of objects was $N = 8$, the observed lapse rate was not noticeably different from zero (Raidvee et al., 2021). This implies we can ignore inattentional feature blindness (or subsampling) as a potential mechanism involved in the pooling of local information signals, at least for the experimental settings that were used in the current study.

Evidently, in addition to varying the total number *N* of elements, other methods are needed for testing arithmetic averaging. Typically, simple averaging models are fitted to experimental data, which leads to inconclusive results because common indicators of the observer's performance are unable to tell different factors apart (Raidvee et al., 2021). Because simple averaging did not explain data accurately enough, additional factors such as under-sampling or inattentional blindness were proposed (e.g., Allik et al., 2013; Dakin, 2001; Solomon et al., 2011). Even if the two simplest rules – arithmetic averaging and maximal value – were found to provide a poor fit to data, usually no alternative explanations were proposed (e.g., Solomon, 2010). Thus, deviations from arithmetic averaging have been acknowledged, but no feasible alternatives have been proposed so far.

Very few studies have assessed the linear additivity in perceptual pooling despite the simplicity of the testing rules. For instance, it is not complicated to test the symmetry axiom which states that the order of addends does not change their sum (cf. Pascucci et al., 2021). Another axiom – associative – states that the form of distribution of a constant sum among addends does not change the result. For example, if the observer's task was to discriminate the mean size of four test circles from the size of the reference circle then it was irrelevant whether all four units of increment or decrement relative to the reference size were added to only one of the test circles (4-0-0-0, where "0" indicates a "dummy" element identical to the reference) or equally distributed among all of them (1-1-1-1, where four units of increment/decrement are distributed equally among all four elements) (Allik et al., 2014). Thus, if the basic rules of arithmetic are obeyed, the precision of discrimination is indifferent to the distribution of increments/decrements among elements. This is exactly what happened in the task of discriminating the mean size of four circles (Allik et al., 2014). However, there is no guarantee that pooling orientation information, or any other attribute, obeys the same simple rules of addition.

Let us not forget that the arithmetic mean is only one of many alternative averaging methods such as the geometric, harmonic, and quadratic means, to say nothing about taking the minimum and maximum values, which are also variants of the summary statistics. If data do not fit arithmetic averaging then it is possible that some other averaging scheme provides a more adequate explanation. All these variants of averaging can be embraced by a single formula known as the *generalized mean* (*Hölder* or *power mean*) in which aggregated values are raised to power *p* and from their sum, the p^{th} root is taken. Let us suppose that we are interested in the pooling of orientation information from *N* elements, the i^{th} of which has orientation $\Delta\phi_i$ relative to some reference orientation. The generalized mean orientation $\Delta\phi^*$ across *N* elements can be found with this formula:

$$\Delta\phi^* = \left(\frac{1}{N}\sum_{i=1}^{N}\Delta\phi_i{}^p\right)^{\frac{1}{p}}$$

where *p* is the power-parameter of the generalized mean. For example, if $p = 1$, then the result corresponds to the arithmetic mean; if $p \to 0$, then this corresponds to the geometric mean, if $p = -1$, then we have the harmonic mean, and if $p = 2$, it is what is called the quadratic mean. One particularly interesting case is when $p \to \infty$, in which case the power mean corresponds to the maximal value or, in our particular case, the maximal tilt from the reference.

There have been multiple psychological studies in which the generalized mean was used for parameterizing information collected from multiple stimulus elements (e.g., Bimler et al., 2013; Shepard, 1987; To et al., 2011). For example, before applying multidimensional scaling (cf. Borg & Groenen, 2005), it is routine to test for something other than the Euclidean metric that could underlie analysed judgements. In ensemble perception, however, other alternatives to the arithmetic averaging are seldom considered (e.g., Whitney & Leib, 2018).

To compare different averaging algorithms, it is necessary to have a stimulus attribute that could differentiate between these various averaging schemes. A promising variant for testing averaging, as mentioned above, is the distribution of the relevant stimulus information among displayed elements. It is possible to render some elements as "dummies" by assigning them attribute values identical to the reference, which make them carry no useful information based on which orientation or any other attribute can be discriminated. The *Proportion of Informative Elements* (*PIE*) can be defined as a ratio of the number of elements that contain usable information to the total number N of displayed elements. Using elements identical to the reference is not a novel approach, and can be found in many previous studies (e.g., Allard & Cavanagh, 2012; Hess et al., 2003; Solomon, 2010). However, we are not aware of studies in which this stimulus attribute was systematically used as a tool for the identification of the averaging algorithm.

Let us assume that the task was to determine the average tilt of four elements that were all rotated by one unit angle θ clockwise from the reference orientation (symbolically, θ - θ - θ - θ). For this combination, the *PIE*=4/4 (numerator shows the number of informative elements and denominator corresponds to the total number N of elements) signifying that all four elements are informative. The same average tilt, however, can be achieved when two elements are rotated by two units of angle and two have zero orientation carrying no useful information (2θ – 2θ – 0 – 0; *PIE*=2/4). Finally, we can assign the whole amount of the usable information to only one element (4θ – 0 – 0 – 0; *PIE*=1/4). Because the total amount of tilt was held constant, the arithmetic mean value also remained unchanged. However, this is not so for other forms of the generalized mean. Let us consider the quadratic mean ($p = 2$) as an example. The quadratic mean for these three hypothetical distributions will be different: 1, 1.41, and 2, respectively, for distributions θ – θ – θ – θ, 2θ – 2θ – 0 – 0, and 4θ – 0 – 0 – 0. Thus, the quadratic mean increases with the decrease of the *PIE* value. This implies that if the total amount of tilts is distributed evenly among elements, the quadratic mean would be smaller than for any uneven distributions. Consequently, the *PIE* is a stimulus attribute that is helpful in testing whether the pooling information from multiple elements is performed in accordance with the rules of arithmetic addition/subtraction or not.

The main goal of this study was to investigate the effects of the number N of elements and the *PIE* on the accuracy with which the mean orientation of a set of objects could be determined. If the orientation of elements is arithmetically averaged, it is expected that the accuracy of discrimination increases proportionally to the square-root-N while the *PIE* has no effect on the mean orientation discrimination accuracy. If these two predictions are violated, then local orientation signals are pooled together using an algorithm that is different from arithmetic addition and subtraction.

Methods

Stimuli and procedure

Stimuli were presented on a flat LCD monitor (60 Hz) at a distance ensuring that one pixel of the screen would subtend about 1 min of arc. Stimuli were generated using MATLAB (The MathWorks, Inc., Natick, MA, USA), the Psychophysics Toolbox (Brainard, 1997) and CircStat Toolbox (Berens, 2009).

Each trial started with the presentation of a fixation cross for 1 s, after which a set of sinusoidal gratings embedded in a Gaussian envelope – Gabor's patches or Gabors – were presented for 250 ms against a homogenous grey background (see Fig. 1). The position of each Gabor was chosen randomly within a circular test area with a diameter of 10.6° of visual angle. An inhibitory radius surrounded each Gabor to avoid spatial overlap between the elements, and guaranteed that two Gabors were never closer than 90 pixels. Because orientation discrimination precision decreases as a function of eccentricity, the main purpose of random positioning was to equalize stimulus elements by their average positions at least.

Participants were asked to indicate, with the corresponding button press, whether the presented set had an average tilt to the left (clockwise) or right (counter-clockwise) compared to the implicit vertical orientation. Previous studies have also used the vertical orientation as an implicit reference (cf. Dakin, 2001). This implicit reference was used because its efficiency was shown to be no different from using explicit standards (Morgan et al., 2000; Nachmias, 2006).

As feedback, one of two simple tones, lower (400 Hz) or higher (600 Hz), was played after each response to indicate its correctness (lower tone for correct responses). The main purpose of feedback was to reduce response biases and to achieve the highest possible response accuracy.

Stimulus display consisted of N = 1, 2, 4, or 8 Gabor patches (with phase of 0°, 2.3 cycles per degree, the value of

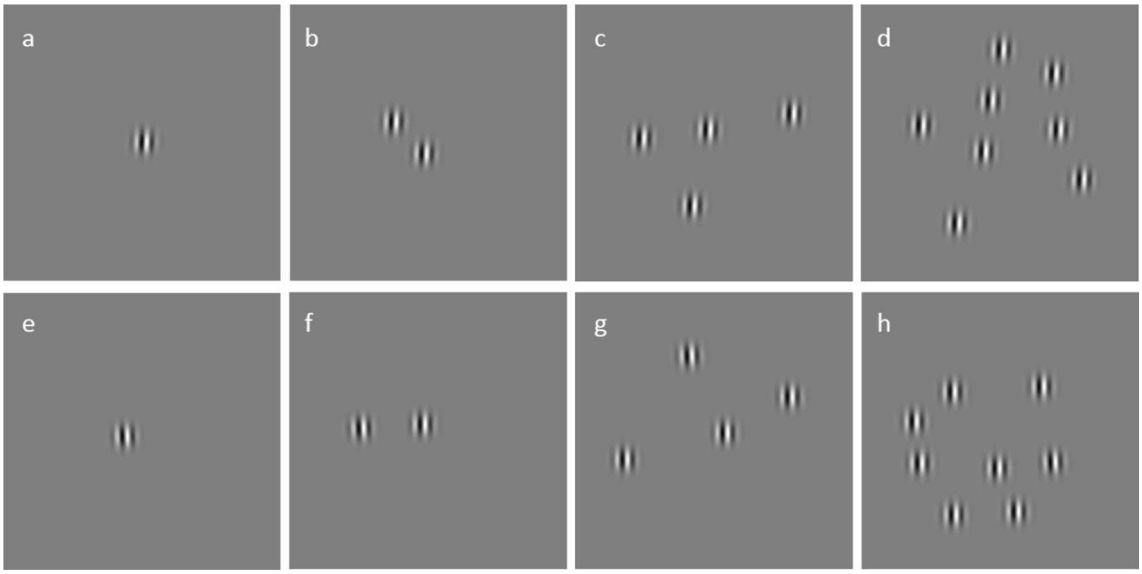

Fig. 1 Examples of stimuli used in this study. The number of Gabor patches $N = 1, 2, 4$, and 8 increases from left to right. In the upper row (panels **a–d**), all Gabors are tilted $\Delta\phi = 2°$ clockwise from the vertical. In the bottom row (**e–h**), only one of N presented Gabors is tilted counter-clockwise from the vertical

the spatial constant of the Gaussian envelope function was 11, Michelson contrast of 75%, 1° of visual angle per grating). All Gabors' parameters except orientation were fixed because this was the only relevant attribute to attend. Stimuli were presented in 16 experimental blocks with set size N held constant within a block. One experimental block consisted of 560 trials (in case of one Gabor patch) or 640 trials (in case of two, four, and eight Gabors) and lasted around 40 min (three blocks were administered in case of one and two Gabor patches, and five blocks in case of four and eight Gabor patches). Participants could take self-paced breaks between the blocks. For $N = 1$ there can only be one *PIE* value, *PIE* = 1/1. For the set of two Gabors two variants exist: *PIE*=1/2 and 2/2. For four Gabors four *PIE* values were used: *PIE* = 1/1, 2/4, 3/4, and 4/4. The *PIE* values used for eight Gabors were: *PIE* = 1/8, 2/8, 4/8 and 8/8. The *PIE* = 1/8 is particularly interesting because even if only one of the elements is missed due to inattentional blindness, it is possible that this element has all necessary information for a correct answer. When this happens the only possibility to give a correct answer is by random guessing, which is expected to change the shape of psychometric function due to the lapse rate λ (Raidvee et al., 2021). Because these preliminary experiments demonstrated a near-zero lapse rate for *PIE* = 1/8, we did not consider inattentional feature blindness as an important factor.

It is necessary to mention that observers were generally unaware of the *PIE* value that was used in each trial. Because all deviations $\Delta\phi$ from the reference were close to the detection threshold, the participants rarely had the impression that all displayed elements had the same tilt or that only one of them deviated significantly from the vertical axis. Thus, there is no strong support for explanations arguing that contrast between neighbor elements' orientations, or perceptual pop-out of some of them, played any role in the pooling of orientation information.

The range of the individual tilt values $\Delta\phi$ was chosen from the extreme anti-clockwise to the extreme clockwise to guarantee that the psychometric functions would approach zero in one of its ends and one in the other end. To cover the whole range of tilts, eight tilt values with an equal step were selected. The mean (or summary) tilt value of all N Gabors was always different from the vertical reference. For all test conditions, a particular combination of tilt values was randomly selected from a list of possible combinations until each condition was replicated at least 100 times. To make conditions comparable, all orientations were expressed in terms of the mean tilt (i.e., the summary tilt from the reference of all presented Gabors divided by the number N of elements). Within each block, the total number of Gabors presented was kept constant while the *PIE* values were randomized.

Observers

Five adult participants – S1, S2, S3, S3, and S5 – took part in the experiment. Two of them were very experienced observers and authors of this study. Three participants were not aware of the design and purposes of this study. All participants had normal or corrected-to-normal vision.

The experimental procedures were approved by the Research Ethics Committee of the University of Tartu.

Results

Each set of discrimination data (11 combinations of the number N of elements and the *PIE*) were approximated by a normal cumulative distribution function (the best fit in terms of non-linear least squares) to find the optimal values of the mean μ and standard deviation . For saving space, Fig. 2 illustrates empirical psychometric functions for three different values of *PIE* = 1/1, 1/8, and 8/8, for all five observers. In the first column of panels, the discrimination functions for a single element (*PIE*=1/1) are demonstrated. This is a special case because the standard deviation ( ) of the best fitting cumulative distribution function characterizes the precision (the inverse of the representational noise) of identifying the orientation of a single Gabor. The values of representational noise ranged from 0.58 (S4) to 1.1 (S1).

The middle column of panels corresponds to *PIE*=1/8, where the orientation of a single element deviated from the reference (vertical) orientation. The standard deviations of the best fitting normal cumulative distribution ranged from 0.36 (S4) to 0.73 (S5). This standard deviation , however, characterizes representational noise of a summary distribution of all eight elements, not of a single element. To distinguish the standard deviation   of the psychometric discrimination function from the imprecision (representational noise) with which orientation of each element can be registered, we use a symbol ς (final-sigma) to characterize representational noise of each single element (cf.; Allik et al., 2013). Only if N = 1, the psychometric   corresponds directly to the representational noise ς. Otherwise, assuming that recorded orientations of elements are arithmetically aggregated, theoretically expected representational noise ς must be $\sqrt{N}$-times larger than the standard deviation   of the best fitting normal cumulative function:

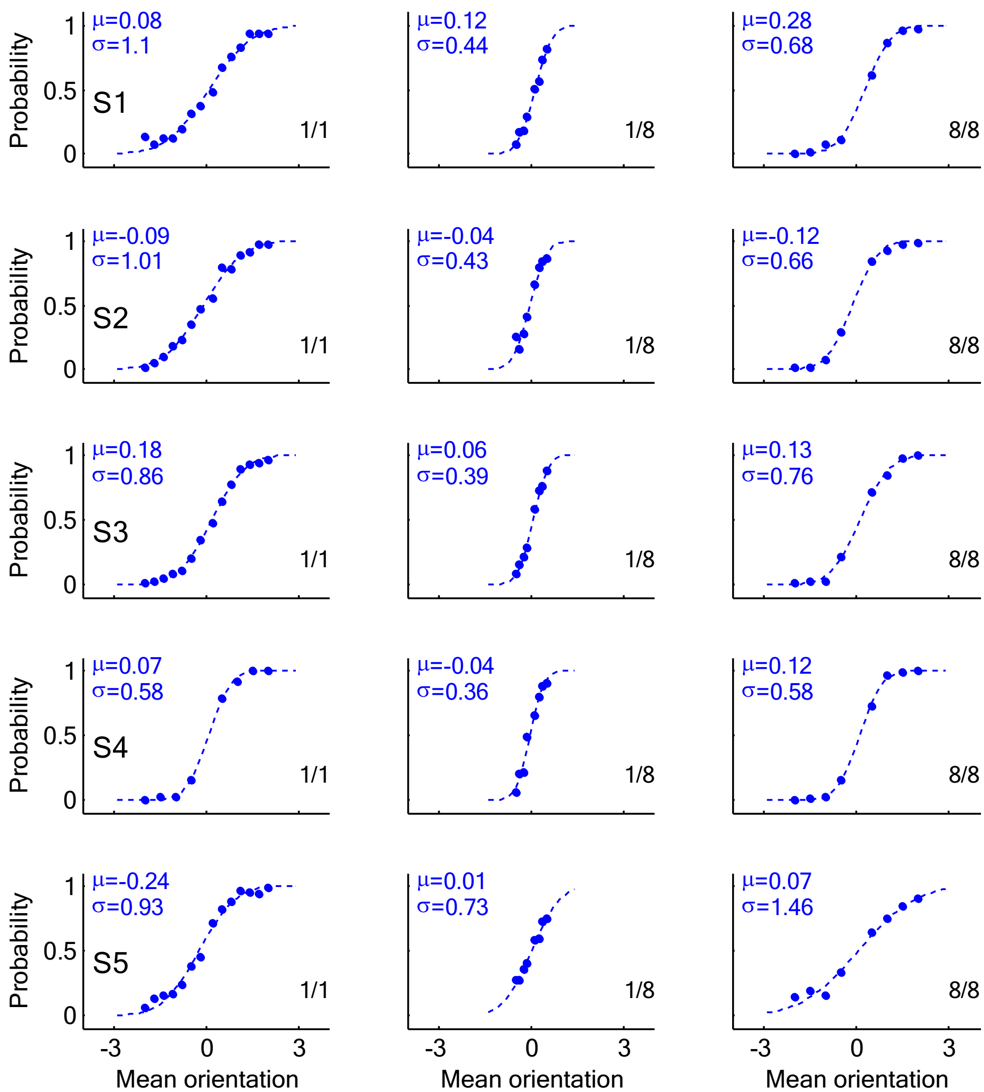

Fig. 2 Examples of empirical discrimination functions for all five observers (in rows) for three most critical stimulus conditions *PIE* = 1/1, 1/8, and 8/8 in the first, second, and third column, respectively

ς = √*N*. In other words, the standard deviation of the summary distribution is √*N*-times smaller than the standard deviation of each individual component. Please notice that the best fitting psychometric functions for *PIE*=1/8 have usual forms in which the lower and upper tails are asymptotically approaching zero and one, respectively, indicating that the lapsing rate λ is close to zero. This is a confirmation of our previous observation that no stimulus elements were missed due to inattentional feature blindness (cf. Raidvee et al., 2021).

This implicates that the representational noise ς of each individual element out of eight is expected to be about √8 = 2.83 times higher than the standard deviation of the psychometric discrimination function. For all observers, the calculated representational noise values ς (1.24, 1.22, 1.10, 1.02, and 2.06) were significantly higher than the respective values for *PIE* = 1/1. This means that the square-root-*N* rule was not obeyed. If the representational noise ς is independent of the number *N* of elements then it is a sign that the recorded orientations of elements are not arithmetically summed or averaged.

The last column demonstrates the condition in which eight Gabors had identical tilt Δϕ (*PIE* = 8/8). Although the arithmetic averaging predicts no difference between *PIE*=1/8 and *PIE* = 8/8, in empirical data the slope of the best fitting psychometric function was always flatter (the standard deviation was larger) in the latter case (i.e., in case of eight elements having the same tilt).

This means that the orientation discrimination was less precise when all eight elements were informative rather than when only one of them carried all relevant information. Thus, the *PIE* determines the mean orientation discrimination accuracy which clearly violates the rules of arithmetic averaging.

Figure 3 demonstrates all calculated representational noise values for a single element calculated from the standard deviations of the best fitting psychometric discrimination function by multiplying it with √*N* for all 11 combinations of *N* and *PIE,* and for all five observers. As in every perceptual task, there were conspicuous individual differences. However, the results of all five observers were similar (at least from the perspective of testing for linear additivity in perceptual pooling). The calculated representational noise values ς were always higher than the respective values for *N* = 1 (black symbols), which indicated a significant drop in the accuracy with each additional element. Higher values of ς mean that summation precision in any number of Gabors cannot inductively, by arithmetic summation be predicted from data where orientation of a single Gabor was judged.

Another invariant property characterizing all five observers is the increase of representational noise ς accompanied by the increase of *PIE* value. For clarity, a linear regression was computed for each set of representational noise estimates for different *PIE*s with the same number *N* of elements. Expectedly, this relationship was obvious for eight Gabors (*N* = 8), where the regression line had a positive slope. The mentioned regularity was less clear for *N* = 4 and especially for *N* = 2, where only two *PIE* values exist. Nevertheless, a condition with eight Gabors (*N* = 8) clearly demonstrates that the precision of orientation discrimination is considerably higher (and thus the value of ς smaller) when only one of eight elements carries relevant information (*PIE* = 1/8) compared to a condition where all eight elements are tilted away from the reference orientation. In other words, the total amount of tilt away from the reference that was necessary for the discrimination of the mean orientation was considerably larger when all eight elements were tilted than when only one element was different from the reference.

Next, we looked for the generalized mean parameter *p* value, which could account for the departure from the flat horizontal function corresponding to the arithmetic mean (*p* = 1). The best fit was obtained with *p* = 2.2, 4.0, 2.2, 1.6, and 1.5 for the observers S1, …, S5, respectively. Thus, values closer to the quadratic rather than arithmetic mean are in better agreement with the obtained data.

Discussion

No evidence of arithmetic averaging

Although extracting statistics from multiple elements is a central theme in ensemble perception (e.g., Alvarez, 2011; Baek & Chong, 2019; Bauer, 2015; Whitney & Leib, 2018), the arithmetic averaging (or linear additivity) was not often tested or demonstrated. In this study, we examined two predictions of the assumption that local orientation signals are pooled together using arithmetic averaging. First, if stochastic representations of the individual elements' attributes are summed together then the discrimination performance is expected to improve according to the square-root-*N*. Second, because of the association rule of arithmetic, the discrimination performance is expected to be independent of the proportion of informative elements or *PIE*. Neither of these two predictions was confirmed. The orientation discrimination performance improved considerably less than by the predicted factor of √*N*. Also, the orientation discrimination performance was significantly better when only one element deviated from the reference compared with a condition in which all elements were tilted away from the reference orientation. Thus, there was no evidence that the orientation information from multiple elements was pooled together by following closely the rules of arithmetic averaging.

However, it is important to remember that voluntary orientation averaging was examined in this study. The observers were instructed to report the perceived mean orientation of all displayed elements. Involuntary averaging occurs when the

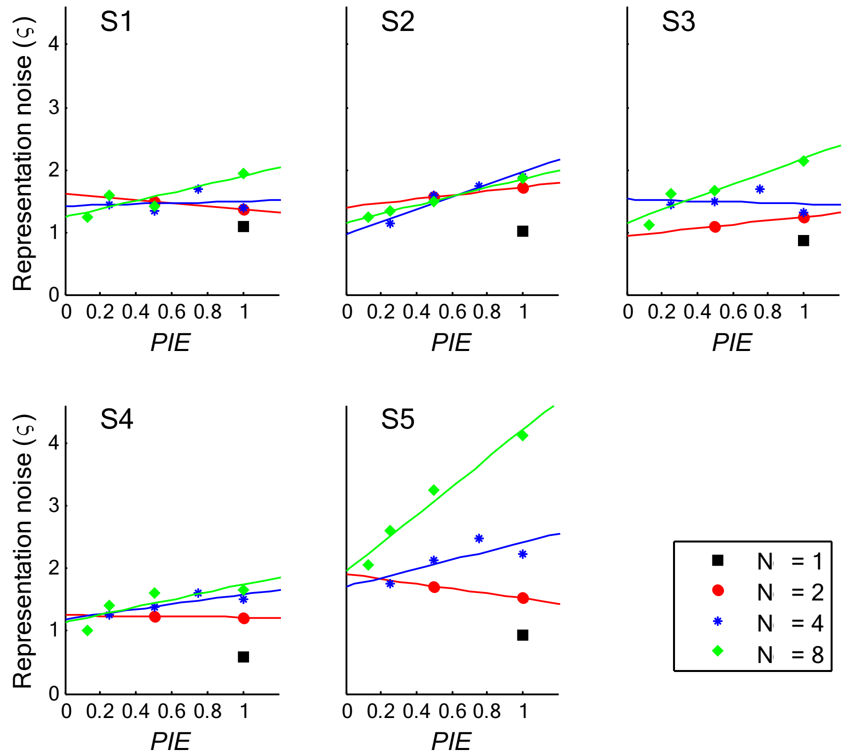

Fig. 3 Calculated representation noise ς as a function of *PIE* and the number of Gabors *N* for the five observers S1, …, S5

samples are unavoidably pooled, for example, under crowding conditions when the samples are near one another in the periphery and are difficult to report independently (Allard & Cavanagh, 2012). It was reported that it was impossible to determine the orientation of the central target without taking into account the orientation of parafoveal objects that were explicitly instructed to be ignored (Parkes et al., 2001). No convincing proof seems to exist, however, that the involuntary compulsory pooling of orientation information "knows" the rules of arithmetic better than the voluntary pooling that was studied in this paper.

An idea that a perceptual system is capable of computing effortlessly statistical properties of an image (Ariely, 2001) incited a hope that it could be a general ability generalizable across many perceptual attributes (cf. Cha et al., 2021; Whitney & Leib, 2018). For a reminder, ensemble perception was defined as a perceptual system's ability to extract summary statistical information from groups of visual objects even if they were presented for a very short period (Whitney & Leib, 2018). Although a confirmation exists that the mean perceived size is indeed computed according to the rules of arithmetic (Allik et al., 2014), the results of the present and other studies demonstrate that these rules are not always applied. For example, there is robust evidence that human visual system is essentially incapable of estimating the mean position of even two different objects (Hess et al., 2003). These findings seem to suggest that the perceptual system may be less proficient of an "intuitive statistician" than previously thought. This may also be a signal for an adjustment of the previous research agenda. Instead of one general perceptual ability, it seems more productive to talk about multiple specific mechanisms that are used for pooling information about various visual attributes across multiple elements.

One mechanism that could potentially explain anomalies in arithmetic averaging is differential weighting of the elements to be added (Iakovlev & Utochkin, 2020; Kanaya et al., 2018; Pascucci et al., 2021). For instance, it is well known that the precision of orientation discrimination decreases towards the periphery of the visual field (Mäkelä et al., 1993). In the same way, if there was a single informative element among eight elements (*PIE*=1/8) then we clearly observed the decrease in the orientation discrimination precision with the distance from the centre of the display. Unfortunately, in most other conditions there were multiple informative elements, and the relationship between discrimination accuracy and spatial position was not determined. As already mentioned, the main purpose of random positions was equalizing the elements' average distance from the centre of the visual field. Even if elements have different weights depending on their spatial positions these weights are expected to regress towards a common average value. However, according to the proposed explanation, no weights are needed to take into account the variable distance from the centre of the visual field. The only required modification is to make the representational noise ς with which the orientation of an element can be perceived variable depending on its distance from the centre. Unfortunately, multiple informative elements make this dependence ambiguous, preventing its exact specification. Thus, we know that the representational noise ς increases towards the periphery, but we had no constraints allowing to specify this increase.

What rule of pooling?

Because the exact arithmetic averaging provided a poor fit to the collected data, a question arises what rule of pooling was used. Obviously, arithmetic averaging is not an imperative of how information from multiple elements is pooled together. Because of the direct relationship between the orientation discrimination precision and the *PIE,* we were able to identify that a quadratic or higher power mean was computed when information from multiple elements was pooled together. Thus, instead of the arithmetic mean, the pooling rule was closer to squaring or even higher powers of individual measures of orientation.

One problem, however, is the reality of these non-linear operations. Although squaring or multiplying is not an unknown operation for the visual system (e.g., Morrone & Burr, 1988; Rashbass, 1970), it is not easy to see the benefits of these non-linear transformations for the pooling orientation or size information. One possible answer is to notice that the generalized mean characterizes average performance and not the process of decision making in every single trial. It is possible, for instance, that different rules are applied in different trials. Notice that if $p \to \infty$ then the power mean is simply corresponding to the maximal tilt value. In other words, alternating the mean ($p = 1$) and maximum ($p \to \infty$) decision strategies in different trials it is possible to achieve a different average p value between these two extremes. In other words, the fitted value of power $p > 1$ may indicate that different strategies were used in different trials. In some of the trials, the answer was given according to the perceived tilt of the element with the largest absolute deviation from the reference; in other trials, however, the answer was given based on the average tilt value. This is, of course, a speculation that can be confirmed by developing experimental procedures that manipulate response strategies.

It is important to note that the generalized mean with $p > 1$ is not the only explanation to the reduction in discrimination accuracy with the increase of proportion of informative elements – *PIE*. For example, it was observed that when it is surrounded by lines of a differing orientation, a test line changes its apparent orientation in a direction away from that of the surrounding lines (Solomon & Morgan, 2006;Virsu & Taskinen, 1975 ; Wenderoth & Johnstone, 1988 ; Westheimer, 1990). Although there is no evidence that this phenomenon, dubbed the *tilt illusion*, can occur between a real test stimulus and an implicit reference, it could be a potential mechanism extending apparent tilt away from the reference. An assumption that a physical angle is transformed into an enlarged perceived angle is technically enough to explain why the tilt of a single element is perceived more accurately than the same summary tilt split into smaller portions and distributed equally between all elements. Mathematically speaking, consequences of the tilt illusion would give results very similar to the generalized mean with $p > 1$. However, there are multiple problems with this explanation. There are no data showing that the perceived angle away from the implicit vertical deviates significantly from its physical value. The tilt illusion itself seems to have properties not favoring this type of explanation. For example, the largest orientation contrast effect occurs when the angle between a target and reference is about 15° (Virsu & Taskinen, 1975; Wenderoth & Johnstone, 1988; Westheimer, 1990). There is no evidence that small tilt angles used in this study are perceptually exaggerated enough for the discrimination accuracy to decrease as the *PIE* value increases. Thus, an explanation relying on the generalized mean has an obvious advantage over the tilt illusion explanation.

Relation to previous studies

Most previous studies on the orientation pooling – voluntary or involuntary – used a wide range of orientations (Anderson et al., 2007; Dakin, 2001; Dakin & Watt, 1997; Gheri & Baldassi, 2008; Husk et al., 2012; Põder, 2012; Solomon et al., 2016; Webb et al., 2010). A distinctive characteristic of this study is that all orientations were in a narrow range, extremely close to the just noticeable tilt away from the reference value. This is one of the main reasons why explanations assuming that orientations of neighbouring stimuli can influence the perceived orientation of a given element are not realistic. Nevertheless, a justified question arises how generalizable are these results from a near threshold conditions to the whole 180° range of orientations? For example, although not likely, the perceptual system may be inept at pooling tiny tilts while in a wide range of orientations, the voluntary pooling can be better described by arithmetic averaging. Indeed, it was found in a voluntary averaging paradigm that if a set of stimuli had a wide range of orientations, then averaging their orientations was possible. However, if orientations were similar to one another then observers seemed to lose their ability to average individual orientation signals (Allard & Cavanagh, 2012). However, without additional experiments it would be premature to conclude that the ability to average orientations vanishes when stimulus elements have remarkably similar orientations.

Many previous studies on combining local orientation signals were inspired by perception of textures using a large number of elements. For example, the maximal number of elements used by Dakin (2001) was 1,024. Baldwin et al. (2014) displayed 841 Gabors to their participants in each trial. It is not realistic to expect that all displayed elements can be processed if their number becomes sufficiently large. It is possible that in addition to inattention, rules of combining local orientation information also change with the number of elements. As an alternative to arithmetic averaging, it was proposed that the perceived orientation of multielement textures is decided based on the centroid of the elements' orientation distribution (Dakin & Watt, 1997), which corresponds to the generalized mean parameter $p \rightarrow 0$. Following studies indicated, however, that the mean orientation is computed based on the arithmetic mean of a fixed fraction of information in elements regardless of their spatial arrangement or density (Dakin, 2001). Although the computation of the arithmetic mean was questioned in some later studies (e.g., Bauer, 2015; Solomon, 2010), no alternative pooling algorithm was proposed. Understandably, additional experiments are needed to demonstrate that arithmetic averaging is indeed a pooling mechanism for textures composed of many repetitive elements.

Another distinction of this study was the avoidance of the concept of *external noise*. Initially, the noise was defined as the response of the perceptual system in the detection tasks which occurs without any input signal (Barlow, 1957). This means that a spontaneous perceptual response can be elicited in the absence of an external signal. If we consider a tilt away from the vertical as a signal, then elements with perfectly vertical orientation could be used to study spontaneous activity in the absence of the visual signal that is a deviation from the reference orientation. Indeed, even if the orientations of all displayed elements were perfectly vertical, their perceived orientation was incoherently scattered around the vertical, demonstrating the presence of internal noise. Later, however, the concept of noise was modified by an assumption that the variance (or the standard deviation) of the feature values of the presented stimuli is a factor which in combination with internal noise, limits the observer's performance (e.g., Allard & Cavanagh, 2012; Dakin, 2001; Florey et al., 2016; Florey et al., 2017; Husk & Hess, 2013; Im & Halberda, 2013). When the observer's task is to detect the presence of a weak signal, the performance is limited by a quantum fluctuation of this signal which creates uncontrollable external noise (Rose, 1957). However, an analogy with the quantal noise or a noisy communication channel is misleading because stimuli used for the study of the perceived size or orientation can be determined perfectly without anything that obscures them. The size and orientation of the displayed stimuli is always certain, corresponding to these parameters' values that were specified in the program running the experiment. Thus, there is no external noise because these parameters can be deterministically specified.

However, besides conceptual issues, this interpretation of the concept of external noise is confronted with some empirical difficulties. Although range or variance of the stimulus values makes the finding of their mean value less precise (Fouriezos et al., 2008), Allard and Cavanagh (2012) noticed that the orientation of four identically oriented Gabors was more difficult to perceive compared to conditions where they had different orientations. Thus, they found that not only increase but also decrease of variance in the stimulus values deteriorates the precision with which the average tilt can be discriminated. In this study, we also saw that when all elements had identical orientation and consequently zero variance the discrimination performance was considerably worse than in situations where elements had a variance in their orientation values. It seems confusing to call the variance of stimulus values *external noise* because instead of expected deterioration of subject's performance the variance among stimulus values seems to facilitate the orientation discrimination precision.

Conclusions

One of the main premises of the ensemble perception is the visual system's ability to compute arithmetic averages across multiple objects for many perceptual attributes such as size, orientation, and even facial expressions. Although the mean size seems to be computed by following the rules of arithmetic addition and subtraction (Allik et al., 2014), no evidence was found in support for arithmetic averaging of orientations across a moderate number of elements, at least not in every single trial. Applying the concept of the generalized or power mean it was found that orientation pooling can be more adequately described by computing quadratic or higher power means, which could be an indication that different pooling rules are used on different trials. It was concluded that the pooling of orientation information is a relatively inaccurate process for which different perceptual cues and their combination rules can be used.

Acknowledgements We thank Endel Põder for his comments at the early stages of this project. The authors gratefully acknowledge critical comments by Joshua Solomon concerning an earlier version of the manuscript. We are grateful to Tiit Mogom for the technical support and assistance. We thank Michael Herzog for helpful comments on an earlier version of the manuscript. We thank Mallory S. E. Roberts for useful discussions.

This research received support from the Estonian Research Council grant PRG1151. Aire Raidvee was supported by the Mobilitas Pluss Returning Researcher Grant (MOBTP91) by Estonian Research Council.

References

Allard, R., & Cavanagh, P. (2012). Different processing strategies underlie voluntary averaging in low and high noise. *Journal of Vision, 12*(11). https://doi.org/10.1167/12.11.6

Allik, J., Toom, M., Raidvee, A., Averin, K., & Kreegipuu, K. (2013). An almost general theory of mean size perception. *Vision Research, 83*, 25-39. https://doi.org/10.1016/j.visres.2013.02.018

Allik, J., Toom, M., Raidvee, A., Averin, K., & Kreegipuu, K. (2014). Obligatory averaging in mean size perception. *Vision Research, 101*, 34-40. https://doi.org/10.1016/j.visres.2014.05.003

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

Anderson, N. D., Murphy, K. M., & Jones, D. G. (2007). Temporal aspects of orientation pooling using visual noise stimuli. *Journal of Vision, 7*(1). https://doi.org/10.1167/7.1.9

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science, 12*(2), 157-162. https://doi.org/10.1111/1467-9280.00327

Baek, J., & Chong, S. C. (2019). Distributed attention model of perceptual averaging. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-019-01827-z

Baek, J., & Chong, S. C. (2020). Ensemble perception and focused attention: Two different modes of visual processing to cope with limited capacity. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-020-01718-7

Baldwin, A. S., Husk, J. S., Meese, T. S., & Hess, R. F. (2014). A two-stage model of orientation integration for Battenberg-modulated micropatterns. *Journal of Vision, 14*(1), 30-30. https://doi.org/10.1167/14.1.30

Barlow, H. B. (1957). Increment thresholds at low intensities considered as signal/noise discriminations. *J Physiol, 136*(3), 469-488. https://doi.org/10.1113/jphysiol.1957.sp005774

Bauer, B. (2015). A selective summary of visual averaging research and issues up to 2000. *Journal of Vision, 15*(4), 14-14. https://doi.org/10.1167/15.4.14

Bayne, T., & McClelland, T. (2019). Ensemble representation and the contents of visual experience. *Philosophical Studies, 176*(3), 733-753. https://doi.org/10.1007/s11098-018-1037-3

Berens, P. (2009). CircStat: A MATLAB Toolbox for Circular Statistics. *Journal of Statistical Software, 31*(10), 1-21. https://doi.org/10.18637/jss.v031.i10

Bimler, D., Izmailov, C., & Paramei, G. (2013). Processing bimodal stimuli: integrality/separability of color and orientation. *Frontiers in Psychology, 4*(759). https://doi.org/10.3389/fpsyg.2013.00759

Borg, I., & Groenen, P. J. F. (2005). Modern multidimensional scaling: Theory and applications (2nd ed.). : Springer.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 437-442. https://doi.org/10.1163/156856897X00357

Cha, O., Blake, R., & Gauthier, I. (2021). Contribution of a common ability in average and variability judgments. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-021-01982-1

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research, 43*(4), 393-404. https://doi.org/10.1016/S0042-6989(02)00596-5

Chong, S. C., & Treisman, A. (2005a). Attentional spread in the statistical processing of visual displays. *Perception & Psychophysics, 67*(1), 1-13. https://doi.org/10.3758/BF03195009

Chong, S. C., & Treisman, A. (2005b). Statistical processing: computing the average size in perceptual groups. *Vision Research, 45*(7), 891-900. https://doi.org/10.1016/j.visres.2004.10.004

Cowan, N. (2015). George Miller's Magical Number of Immediate Memory in Retrospect: Observations on the Faltering Progression of Science. *Psychological Review, 122*(3), 536-541. https://doi.org/10.1037/a0039035

Dakin, S. C. (2001). Information limit on the spatial integration of local orientation signals. *Journal of the Optical Society of America a-Optics Image Science and Vision, 18*(5), 1016-1026. https://doi.org/10.1364/JOSAA.20.000974

Dakin, S. C., & Watt, R. J. (1997). The computation of orientation statistics from visual texture. *Vision Research, 37*(22), 3181-3192. https://doi.org/10.1016/s0042-6989(97)00133-8

Florey, J., Clifford, C. W. G., Dakin, S., & Mareschal, I. (2016). Spatial limitations in averaging social cues. *Scientific Reports, 6*. https://doi.org/10.1038/srep32210

Florey, J., Dakin, S. C., & Mareschal, I. (2017). Comparing averaging limits for social cues over space and time. *Journal of Vision, 17*(9). https://doi.org/10.1167/17.9.17

Fouriezos, G., Rubenfeld, S., & Capstick, G. (2008). Visual statistical decisions. *Perception & Psychophysics, 70*(3), 456-464. https://doi.org/10.3738/pp.70.3.456

Gheri, C., & Baldassi, S. (2008). Non-linear integration of crowded orientation signals. *Vision Research, 48*(22), 2352-2358. https://doi.org/10.1016/j.visres.2008.07.022

Hess, R. F., Barnes, G., Dumoulin, S. O., & Dakin, S. C. (2003). How many positions can we perceptually encode, one or many? *Vision Research, 43*(14), 1575-1587. https://doi.org/10.1016/s0042-6989(03)00122-6

Husk, J. S., & Hess, R. F. (2013). Global processing of orientation in amblyopia. *Vision Research, 82*, 22-30. https://doi.org/10.1016/j.visres.2013.02.005

Husk, J. S., Huang, P. C., & Hess, R. F. (2012). Orientation coherence sensitivity. *Journal of Vision, 12*(6). https://doi.org/10.1167/12.6.18

Iakovlev, A. U., & Utochkin, I. S. (2020). Roles of saliency and set size in ensemble averaging. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-020-02089-w

Im, H. Y., & Halberda, J. (2013). The effects of sampling and internal noise on the representation of ensemble average size. *Attention Perception & Psychophysics, 75*(2), 278-286. https://doi.org/10.3758/s13414-012-0399-4

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. https://doi.org/10.1098/rspb.2017.2770

Mäkelä, P., Whitaker, D., & Rovamo, J. (1993). Modelling of orientation discrimination across the visual field. *Vision Research, 33*(5), 723-730. https://doi.org/10.1016/0042-6989(93)90192-Y

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97. https://doi.org/10.1037/h0043158

Morgan, M. J., Watamaniuk, S. N. J., & McKee, S. P. (2000). The use of an implicit standard for measuring discrimination thresholds. *Vision Research, 40*(17), 2341-2349. https://doi.org/10.1016/s0042-6989(00)00093-6

Morrone, M. C., & Burr, D. C. (1988). Feature detection in human vision: a phase-dependent energy model. *Proceedings of the Royal Society of London. Series B. Biological Sciences, 235*(1280), 221-245. https://doi.org/10.1098/rspb.1988.0073

Myczek, K., & Simons, D. J. (2008). Better than average: Alternatives to statistical summary representations for rapid judgments of average size. *Perception & Psychophysics, 70*(5), 772-788. https://doi.org/10.3758/pp.70.5.772

Nachmias, J. (2006). The role of virtual standards in visual discrimination. *Vision Research, 46*(15), 2456-2464. https://doi.org/10.1016/j.visres.2006.01.029

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience, 4*(7), 739-744. https://doi.org/10.1038/89532

Pascucci, D., Ruethemann, N., & Plomp, G. (2021). The anisotropic field of ensemble coding. *Scientific Reports, 11*(1), 8212. https://doi.org/10.1038/s41598-021-87620-1

Põder, E. (2012). On the rules of integration of crowded orientation signals. *I-Perception, 3*(7), 440-454. https://doi.org/10.1068/i0412

Raidvee, A., Toom, M., & Allik, J. (2021). A method for detection of inattentional feature blindness. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-020-02234-5

Rashbass, C. (1970). The visibility of transient changes of luminance. *Journal of Physiology, 210*(1), 165-186. https://doi.org/10.1113/jphysiol.1970.sp009202

Rose, A. (1957). Quantum Effects in Human Vision. In J. H. Lawrence & C. A. Tobias (Eds.), Advances in Biological and Medical Physics (Vol. 5, pp. 211-242): Elsevier.https://doi.org/10.1016/B978-1-4832-3111-2.50009-2

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237*(4820), 1317-1323. https://doi.org/10.1126/science.3629243

Solomon, J. A. (2010). Visual discrimination of orientation statistics in crowded and uncrowded arrays. *Journal of Vision, 10*(14). 19 https://doi.org/10.1167/10.14.19

Solomon, J. A., & Morgan, M. J. (2006). Stochastic re-calibration: contextual effects on perceived tilt. *Proceedings of the Royal Society B-Biological Sciences, 273*(1601), 2681-2686. https://doi.org/10.1098/rspb.2006.3634

Solomon, J. A., Morgan, M., & Chubb, C. (2011). Efficiencies for the statistics of size discrimination. *Journal of Vision, 11*(12), 13, 11-11. https://doi.org/10.1167/11.12.13.

Solomon, J. A., May, K. A., & Tyler, C. W. (2016). Inefficiency of orientation averaging: Evidence for hybrid serial/parallel temporal integration. *Journal of Vision, 16*(1). https://doi.org/10.1167/16.1.13

Sorkin, R. D., Mabry, T. R., Weldon, M. S., & Elvers, G. (1991). Integration of information from multiple element displays. *Organizational Behavior and Human Decision Processes, 49*(2), 167-187. https://doi.org/10.1016/0749-5978(91)90047-W

To, M. P. S., Baddeley, R. J., Troscianko, T., & Tolhurst, D. J. (2011). A general rule for sensory cue summation: evidence from photographic, musical, phonetic and cross-modal stimuli. *Proceedings of the Royal Society B-Biological Sciences, 278*(1710), 1365-1372. https://doi.org/10.1098/rspb.2010.1888

Virsu, V., & Taskinen, H. (1975). Central inhibitory interactions in human vision. *Experimental Brain Research, 23*(1), 65-74. https://doi.org/10.1007/bf00238729

Webb, B. S., Ledgeway, T., & McGraw, P. V. (2010). Relating spatial and temporal orientation pooling to population decoding solutions in human vision. *Vision Research, 50*(22), 2274-2283. https://doi.org/10.1016/j.visres.2010.04.019

Wenderoth, P., & Johnstone, S. (1988). The different mechanisms of the direct and indirect tilt illusions. *Vision Research, 28*(2), 301-312. https://doi.org/10.1016/0042-6989(88)90158-7

Westheimer, G. (1990). Simultaneous orientation contrast for lines in the human fovea. *Vision Research, 30*(11), 1913-1921. https://doi.org/10.1016/0042-6989(90)90167-J

Whitney, D., & Leib, A. Y. (2018). Ensemble Perception. *Annual Review of Psychology, 69*, 105-129. https://doi.org/10.1146/annurev-psych-010416-044232

Whitney, D., Solomon, J. A., & Hochstein, S. (2021). Introduction to the special issue on ensemble perception. *Attention Perception & Psychophysics, 83*(3), 899-903. https://doi.org/10.3758/s13414-021-02275-4