

Discrimination of numerical proportions: A comparison of binomial and Gaussian models

Aire Raidvee^{1,2,3} · Jüri Lember⁴ · Jüri Allik^{1,5}

Published online: 26 September 2016
© The Psychonomic Society, Inc. 2016

Abstract Observers discriminated the numerical proportion of two sets of elements ($N = 9, 13, 33,$ and 65) that differed either by color or orientation. According to the standard Thurstonian approach, the accuracy of proportion discrimination is determined by irreducible noise in the nervous system that stochastically transforms the number of presented visual elements onto a continuum of psychological states representing numerosity. As an alternative to this customary approach, we propose a Thurstonian-binomial model, which assumes discrete perceptual states, each of which is associated with a certain visual element. It is shown that the probability β with which each visual element can be noticed and registered by the perceptual system can explain data of numerical proportion discrimination at least as well as the continuous Thurstonian–Gaussian model, and better, if the greater parsimony of the Thurstonian-binomial model is taken into account using AIC model selection. We conclude that Gaussian and binomial models represent two different fundamental principles—internal noise vs. using only a fraction of available information—which are both plausible descriptions of visual perception.

Keywords Divided attention and inattention · Theoretical and computational models · Math modeling

Numerous carefully performed experiments have shown that many animals—such as bees (Gross et al., 2009), fish (Gomez-Laplaza & Gerlai, 2013; Krusche, Uller, & Dicke, 2010; Petrazzini, Agrillo, Piffer, & Bisazza, 2014; Piffer, Agrillo, & Hyde, 2012), salamanders (Krusche et al., 2010), pigeons (Emmerton & Renner, 2006), dogs (Ward & Smuts, 2007), chimpanzees (Beran, Evans, & Harris, 2008), to say nothing about human infants (Brannon & Van de Walle, 2001) or people whose language lacks words for numbers beyond five (Pica, Lemer, Izard, & Dehaene, 2004)—are capable of discriminating relatively large quantities with remarkable precision, even when other cues are not available and there is no opportunity or ability for one-by-one counting. Considerable progress has been made in understanding where and how the brain represents quantities, including numerical proportions (Jacob, Vallentin, & Nieder, 2012). It was noted, for example, that the same cortical regions participate in the encoding of both absolute quantities and quantity ratios (proportions), and that single neurons in the prefrontal and parietal cortices of various species are tuned to preferred proportions (Nieder & Miller, 2004; Vallentin & Nieder, 2010). Based on these and other observations, it was proposed that the brain uses an analog, labeled-line code to represent the true numerical values of quantity ratios (Jacob et al., 2012; Moskaleva & Nieder, 2014; Vallentin, Jacob, & Nieder, 2012; Viswanathan & Nieder, 2013). Although discrimination of nonsymbolic proportions does not require verbal coding or propositional thinking, an intuitive approximate number sense, which is believed to form the basis of the perception of proportions, may serve as an antecedent of higher numerical abilities (Halberda, Mazzocco, & Feigenson, 2008). However, the last proposition is vigorously

✉ Aire Raidvee
aire.raidvee@ut.ee

¹ Department of Psychology, University of Tartu, Näituse 2, Tartu 50409, Estonia

² Laboratory of Psychophysics, Brain Mind Institute, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

³ New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

⁴ Institute of Mathematics and Statistics, University of Tartu, Tartu, Estonia

⁵ Estonian Academy of Sciences, Tallinn, Estonia

debated in the recent literature (Agrillo, Piffer, & Adriano, 2013; Göbel, Watson, Lervåg, & Hulme, 2014).

If the observer's task is to discriminate the relative proportion of two distinct sets of randomly distributed elements, these two sets can be separated in two principal ways. First, the sets occupying two distinct areas can differ by their spatial or temporal position (Allik & Tuulmets, 1991, 1993); second, they can be spatially intermixed but distinguished by a certain visual attribute, such as color, orientation (Honig & Matheson, 1995; Honig & Stewart, 1993; Tokita & Ishiguchi, 2009, 2010), or motion in two different directions (Raidvee, Averin, Kreegipuu, & Allik, 2011). As it turned out, the ability to discriminate numerical proportion depends heavily on the visual attributes by which the two sets can be distinguished. Rather surprisingly, human observers are extremely inaccurate in discriminating proportion between two spatially overlapping sets of randomly distributed elements moving in opposite directions (Raidvee et al., 2011). In a wide range of set sizes, decisions about motion direction are made as if only a very limited number of elements (in some cases, less than 0.5 %) are taken into account, even if the motion direction of each element in isolation can be determined with near absolute certainty (Raidvee, Averin, & Allik, 2012). In a similar task, where instead of motion direction the elements differed by color, observers were able to discriminate the relative number of red and green dots as if they had taken into account 69 elements from a total of 100 (Tokita & Ishiguchi, 2009) provided that psychometric response curves are interpreted in terms of dot counts used in the decision (Raidvee, Averin, et al., 2012; Raidvee et al., 2011). However, the same observers' ability to discriminate between the relative number of parallel and converging lines was much poorer, with a precision equal to discrimination decisions made on the basis of no more than two elements out of the 100 available (Tokita & Ishiguchi, 2009). This may mean that there is no single fixed brain circuitry for the discrimination of numerical proportions. It is more likely that differences in the discrimination of proportion based on varying stimulus attributes share the same late stages in the processing stream from the sensory code to the motor response. Nevertheless, the brain's representation of proportions must also include sensory codes which are specific to different visual attributes and which are probably located in different brain regions (cf. Gebuis & Reynvoet, 2012, 2013; Tokita & Ishiguchi, 2012).

When the observer's task is to discriminate between numerical proportions, empirical psychometric functions are typically approximated by a cumulative normal distribution, with a mean (μ) roughly corresponding to the perceived level of balance between the two judged quantities, and with a standard deviation (σ) characterizing the precision of numerosity discrimination. The slope of the psychometric function can be characterized by the inverse value of the standard deviation, $1/\sigma$. According to a customary approach, this imprecision σ is caused by a spontaneous, irreducible noise in the nervous system that stochastically transforms a fixed stimulus value into one of many

possible internal states. All these explanations, relating response errors to internal noise, are variants of the Thurstonian psychophysical analysis (Thurstone, 1927a, b), according to which, stimuli are projected onto a set of psychological states, thus forming subjective images with positions varying randomly among these states. The magnitude of this internal noise, the discriminational dispersion, in Thurstonian terminology, is believed to be quite straightforwardly related to the slope of the observed psychometric function.

Because the normal or Gaussian distribution usually fits empirical data rather well, it has been convenient to assume that the internal noise can indeed be modeled by a continuum of internal states with approximately normally distributed occurrence probabilities. Because numerous stochastic distributions, including discrete ones, closely approach the normal distribution, the assumption about the distribution of internal noise remains hypothetical in the least (cf. Neri, 2013). Furthermore, the assumption of normally distributed internal noise is not very convincing when the observer's task is to estimate other than the magnitude of continuous physical attributes such as weight, size, or luminance. For example, it is rather problematic to assume that, in the numerosity discrimination task, discrete visual elements would give rise to continuous subjective images. According to this conception, even small quantities are projected onto a continuum of internal states resembling the axis of real numbers. For instance, sometimes a set consisting of six visual elements may be perceived as more numerous than another set consisting of seven elements because the internal image of the first set has a higher value (e.g., 6.6 perceptual units) than the internal image of the second set (e.g., 6.5 perceptual units). This is not an unlikely outcome, provided that the standard deviation of the internal noise is equal to, for example, three or four elements. This assumption, however, runs into problems when trying to explain why there is no confusion when the number of elements is less than four (Jevons, 1871), to say nothing about finding a good explanation for what it really means to perceive one half or two thirds of a dot or of any other geometrical element. In addition, all Thurstonian–Gaussian models make an unrealistic assumption that the observer's decisions are based on all information. It is much more realistic to assume that the capacity of the visual system is strongly limited in the result of which only a fraction of all information is used.

At least hypothetically, we can entertain a version of the Thurstonian model in which internal states are represented not by a continuum, but by a finite set of discrete states. Let us suppose that a psychophysical function transforms each element presented on the screen into a corresponding mental state, which can take only two values: zero (“off”) when the presence of the element was not recorded, or one (“on”) when the element was noticed and recorded. Each element of a certain type is transformed into its “on” state with a probability β . This general definition leaves the probability β open to various psychological interpretations. For instance, the probability β can be interpreted

as perceptual salience of a certain type of elements, but it could also be considered as the level of attentiveness with which elements can be noticed and registered by the perceptual system. We can call this the Thurstonian-binomial approach to discrimination of numerical proportions.

If we accept these simple premises, then the derivation of computational formulae for the perception of numerical proportions is a matter of simple calculations. Let N_A and N_B be the numbers of elements in two distinct sets of elements from Type A and B (e.g., red vs. green), respectively. It is easy to predict how many K_A and K_B elements from the total number of presented elements N_A and N_B are on average taken into account if there is the constant probability β_i with which each element of a certain type is counted (where $i = 1, \dots, n$, and n is the number of different types of elements in the display; in this case, $n = 2$). These two numbers, K_A and K_B , are obviously not constants but two random variables, with the binomial distributions $B(N_A, \beta_1)$ and $B(N_B, \beta_2)$, respectively, whereas the two binomial distributions are assumed mutually independent. The expected values of these two binomially distributed variables are

$E(K_A) = N_A\beta_1$ and $E(K_B) = N_B\beta_2$. Observers' decisions about relative numerosity are assumed to be based on a comparison of these two binomial values in a given trial: if $K_A > K_B$, the observer answers that "Type A is more numerous than Type B." Otherwise, Type B is judged as more numerous than Type A. In case of equality of $K_A = K_B$, an unbiased observer answers randomly, with each answer being equally probable. To know how often K_A is expected to be larger than K_B , the distribution of the difference between these two binomially distributed variables is required.

The exact probability of responding "Type A is more numerous than Type B" (denoted as "Answer = A") is equal to the sum of the probability of K_A being larger than K_B plus 0.5 times the probability of K_A being equal to K_B , given certain probabilities β_1 and β_2 of detecting Type A and Type B elements, as described by the following formulae for the situations where Type A is either less numerous than Type B (Eq. 1), or more numerous than Type B (Eq. 2; for derivation, consult Appendix 1):

$$\begin{aligned}
 P(\text{Answer} = A | N_A < N_B) &= P(K_A > K_B | N_A < N_B) + 0.5P(K_A = K_B | N_A < N_B) \\
 &= \sum_{k_1=0}^{N_A} \sum_{k_2=0}^{k_1-1} \binom{N_A}{k_1} \binom{N_B}{k_2} \beta_1^{k_1} (1-\beta_1)^{N_A-k_1} \times \beta_2^{k_2} (1-\beta_2)^{N_B-k_2} \\
 &\quad + 0.5 \sum_{k=0}^{N_A} \binom{N_A}{k} \binom{N_B}{k} \beta_1^k (1-\beta_1)^{N_A-k} \times \beta_2^k (1-\beta_2)^{N_B-k}
 \end{aligned} \tag{1}$$

$$\begin{aligned}
 P(\text{Answer} = A | N_A > N_B) &= P(K_A > K_B | N_A > N_B) + 0.5P(K_A = K_B | N_A > N_B) \\
 &= 1 - \sum_{k_1=0}^{N_B} \sum_{k_2=0}^{k_1-1} \binom{N_B}{k_1} \binom{N_A}{k_2} \beta_2^{k_1} (1-\beta_2)^{N_B-k_1} \times \beta_1^{k_2} (1-\beta_1)^{N_A-k_2} \\
 &\quad - 0.5 \sum_{k=0}^{N_B} \binom{N_A}{k} \binom{N_B}{k} \beta_1^k (1-\beta_1)^{N_A-k} \times \beta_2^k (1-\beta_2)^{N_B-k}
 \end{aligned} \tag{2}$$

where N_A and N_B are the total numbers of elements in the Categories A and B, respectively (e.g., red vs. green); K_A and K_B are the numbers of elements taken into account from the total number of presented elements N_A and N_B , given constant binomial probabilities β_1 and β_2 with which each element of a certain type is counted.

If the expectations of K_A and K_B are sufficiently large, they are approximately normally distributed (given that with large N , binomial distribution approximates the Gaussian distribution) and we can use normal approximation. For this we need to express the difference between $N_A\beta_1 - N_B\beta_2$ in z scores (units of the standardized normal distribution):

$$z = \frac{N_A\beta_1 - N_B\beta_2}{\sqrt{N_A\beta_1(1-\beta_1) + N_B\beta_2(1-\beta_2)}} \tag{3}$$

If these two detecting probabilities are identical, $\beta_1 = \beta_2$, then the psychometric function becomes symmetrical, and its shape can be approximated by a normal cumulative distribution. In fact, there is a simple equation relating the standard deviation σ of a cumulative normal distribution to the counting probability β (see Appendix 2 for derivation):

$$\sigma = \sqrt{\frac{1-\beta}{4N\beta}} \tag{4}$$

By a direct analogy, we can also construct the classical Thurstonian model assuming that there is a continuum of internal states and the distribution of the internal noise is sufficiently close to normal. However, it is important to notice that mechanical approximation of empirical data with a single Gaussian function is not necessarily an implementation of the Thurstonian model. As with the previously described binomial model, we need to start from a proposition that each stimulus element has a stochastic internal representation, and all these elementary representations are pooled to make a decision about relative numerosity of elements. A formula describing the Thurstonian normal or Gaussian model is as follows:

$$P(N_A \geq N_B) = \Phi\left(\frac{\mu_1 N_A - \mu_2 N_B}{\sqrt{N_A \sigma_1^2 + N_B \sigma_2^2}}\right), \quad (5)$$

where $P(N_A \geq N_B)$ is the probability with which elements from Category A are judged to be more numerous than elements from the Category B (e.g., red vs. green); N_A and N_B are the total numbers of elements in the Categories A and B, respectively; μ_1 and μ_2 are the mean values, and σ_1 and σ_2 standard deviations of stochastic internal representation of individual visual elements of Type A and B, respectively; and Φ stands for the standard normal integral. Please notice that this formula does not coincide with a simple cumulative normal approximation that is widely used for the description of empirical psychometric functions. For its derivation, it is useful to remember that the sum of two independent, normally distributed random variables is normal, with its mean being the sum of the two means, and its variance being the sum of the two variances. Only when $\mu_1 = \mu_2$ and $\sigma_1 = \sigma_2$, Eq. 5 reduces to the most simple case, Case V, in Thurstone's (1927a) own terminology.

The aim of this article is to compare the classical Thurstonian–Gaussian model (Eq. 5) with the Thurstonian–binomial model (Eqs. 1 and 2). We expect that, in terms of model residuals, both models perform sufficiently well, especially because with large N , binomial distribution approximates the Gaussian distribution. When the greater parsimony of Thurstonian–binomial model is taken into account (using, for example, an AIC-based approach), it clearly surpasses the Thurstonian–Gaussian model.

We expected the differences in the predictions from the two models to be most pronounced in cases where the number of trials, or in this case, number of sampled elements, is small. This could happen, for example, when discriminating between relatively small numbers of visual elements, or when a visual attribute by which the two sets of elements can be distinguished is not salient enough, and thus the expected values of K_A and K_B are small. This expectation derives directly from the fact that binomial distribution approaches the normal distribution when the number of trials—in this case, the number of elements that are taken into account in the discrimination process—increases.

For this reason we selected two visual attributes—color and orientation—which are known to have considerably different capacities to distinguish between two sets of visual elements (Tokita & Ishiguchi, 2009). We are also going to discuss which of these two models is intuitively more plausible and provides a psychologically more reasonable interpretation.

Method

Participants Four 20-year-old female observers (Agne, Dea, Kristin, and Moonika), with normal or corrected-to-normal vision, were asked to decide which of the two distinctive sets of objects was more numerous.

Stimuli In two separate series, these two sets of objects were distinguished either by color or by orientation. A schematic view of the two types of stimulus configurations is depicted in Figure 1. In the first series, a randomly distributed collection of red and green circles was presented. The red and green circles had equal luminance of about 23.5 cd/m². To diminish the impact of total red versus green area on responses, the size of the circles was randomly varied in the range of 11 to 22 minutes of arc (the summed area being proportional to the number of elements). In the second series of experiments, a collection of short, black line segments of luminance 0.3 cd/m² and a tilt of 20°, either to the left or to the right of the vertical direction, was presented. The width and length of a line subtended 2' and 19', respectively (and the height of its vertical projection was 16'). Both types of stimuli were presented within an elliptical gray background, with a luminance of 54 cd/m² and with lengths of horizontal and vertical axes of 8.86° and 8.70°, respectively. This elliptical background was in the center of a rectangular area of luminance 64 cd/m², filling the rest of the screen. To avoid overlap between elements, each element was positioned within an invisible inhibitory area that prevented other elements from being closer than 22'. Each stimulus element had contrast sufficient to guarantee its 100 % identification had it been presented in isolation. The total number of objects, N , presented on the display was kept constant through each experimental session and was equal to $N = 9, 13, 33$, or 65 elements. During experimental sessions, the relative proportion of Type A and Type B elements was varied in random order. The total number of elements was constant throughout each session. For the total number of elements $N = 9$ and $N = 13$, the relative proportions of A (red or tilted to the left) and B (green or tilted to the right) element categories were varied with a change of one element to the numerosity of both sets: from 1:8 to 8:1 and from 1:12 to 12:1, respectively. For the total number of elements $N = 33$, the proportions 13:20, 14:19, 15:18, 16:17, and the reverse, were used; for $N = 65$, the proportions 23:42, 26:39, 32:33,

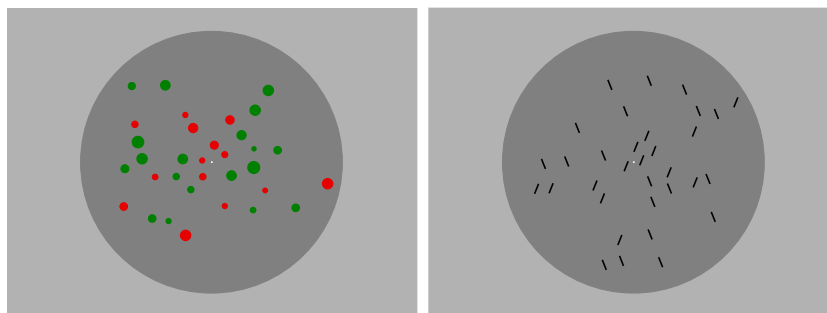


Fig. 1 Schematic view of two types of stimuli used for the study of discrimination of numeric proportions

and the reverse, were used. The given proportions were chosen in an attempt to yield psychometric functions covering most of the range of response probabilities. The stimuli were presented for 200 milliseconds, with 3 seconds for responding. In the case of nonresponse, the trial was repeated at a later, randomly selected position in the same series of trials.

Every stimulus condition was replicated 100 times. The choice probability of red circles was plotted as a function of the proportion of red elements N_A of the total number of elements on the display $N = N_A + N_B$, where N_B refers to the number of green elements. Similarly, in the orientation experiment, the probability of the choice of leftward-tilted elements was measured as a function of the proportion of leftward-tilted elements N_A in the total number of elements on the display N .

Part of the data presented here has been published previously (Raidvee, Pölder, & Allik, 2012) in a different modeling approach.

Apparatus All stimuli were presented at a viewing distance of 170 cm on the screen of a Mitsubishi Diamond Pro 2070SB 22-in. color monitor (the frame rate was 140 Hz, with a resolution of 1024×768 pixels, subtending 12.9° horizontally and 9.8° vertically) with the help of a ViSaGe (Cambridge Research Systems Ltd.) stimulus generator.

Statistical analysis The models described by Eqs. 1 and 2, and 5, were fitted to the data by custom software written in MATLAB. Nonlinear least squares optimization was used to find the best fit parameter estimates. The 95 % parameter confidence intervals were obtained by bootstrap resampling of the model residuals (1,000 repetitions). The explained variance was estimated by the degrees of freedom adjusted R^2 of the nonlinear regression fit. Corrected Akaike information criterion (AICc; Hurwitsch & Tsai, 1989) was computed as $AIC = n \cdot \ln(SSE/n) + 2K$, and $AICc = AIC + 2K \cdot (K + 1)/(n - K - 1)$, where n is the number of data points (in this experiment, varies from six to 12); K is the number of model parameters + 1; and SSE is sum of squared errors of the model.

Results

Figure 2 demonstrates empirical data (red circles) fitted by theoretical psychometric functions: Thurstonian–Gaussian model (black dotted curves) described by Eq. 5, and Thurstonian-binomial model (black plus signs), described jointly by Eqs. 1 and 2. The probability of saying that the number of elements from Type A (*red or tilted to the right*) is larger than Type B (*green or tilted to the left*) is plotted against the proportion of N_A elements: $N_A / (N_A + N_B)$. Each panel corresponds to the total number of elements ($N = 9, 13, 33,$ and 65) separately for color and orientation as two distinguishing attributes, pooled across the four participants.

Because the majority of binomial predictions lay on the continuous Gaussian psychometric functions it is expected that both models, Gaussian and binomial, provide similarly close fit to empirical data. Under the information theoretic approach, the more parsimonious binomial model has an advantage.

Table 1 provides an overview of the best-fitting values for both Thurstonian–Gaussian and Thurstonian-binomial models. The most optimal parameter values were determined by fitting the functions to empirical data by nonlinear least squares. There are several points of principal interest in the presented data, as follows.

Which of the two models provides a better fit? The differences between the two models in terms of R^2 and AIC values are outlined in Table 2. Judging by the values of adjusted R^2 s, both models, the binomial and Gaussian, provided approximately equally good fits to the collected response data. Averaging across all conditions and observes the binomial function explained on average 98.87 % of the observed variance. The Gaussian model explained 98.62 % of the variance in exactly the same set of response data. Thus, both models left on average only about 1.26 % of variance unexplained. Both models explained color data about 1.3 % . . . 1.6 % better than orientation data. The average AIC values were -46.65 and -48.22 for Thurstonian-binomial and Thurstonian–Gaussian models, respectively. On average, the AICs were smaller by 3.4 (for the binomial model) and by 2.86 (for the Gaussian model) for color data compared to

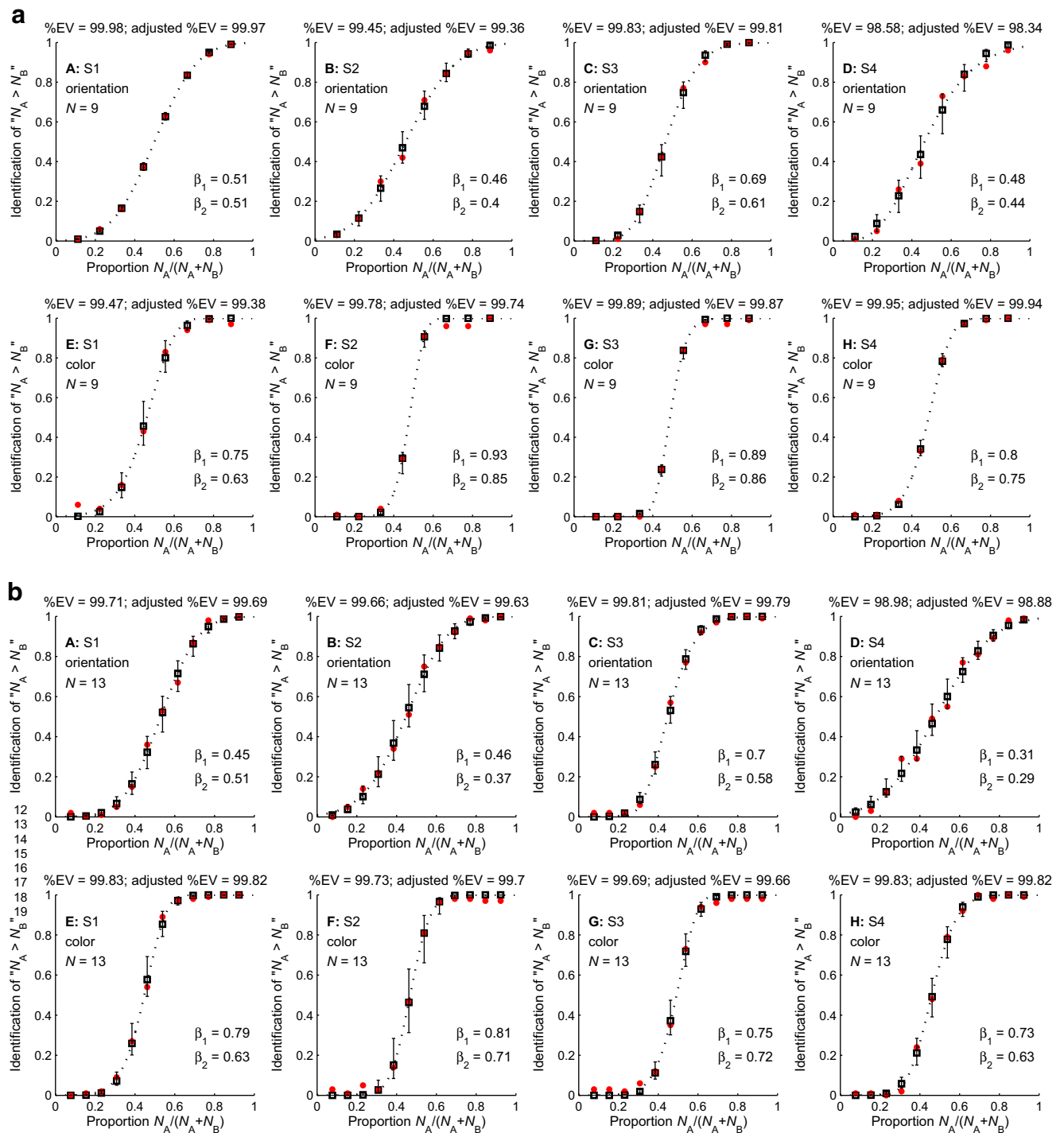


Fig. 2 Empirical choice probabilities (red circles) and predictions of the Thurstonian–Gaussian (continuous curves) and Thurstonian-binomial (black hollow squares) models separately for each subject: a. $N = 9$, b. $N = 13$, c. $N = 33$, d. $N = 65$, and e. pooled across the four subjects.

Vertical lines denote the 95 % bootstrap confidence intervals for the Thurstonian-binomial estimates of the response probabilities. (Color figure online). %EV denotes the percentage of explained variance in terms of (adjusted) R^2

the orientation data. The differences in terms of the adjusted R^2 (values differ by 0.24 % in favor of the binomial model), and the AIC statistics (difference of 1.95 in favor of the Gaussian model) are small and imply that both models should receive full consideration in the explanation of proportion discrimination data.

Because the number of fitted data points in this study was small, we applied the corrected AIC, denoted as AICc (Hurvich & Tsai, 1989) as it provides an unbiased estimate of relative loss of information. It is shown that in practice, AICc should always be preferred over AIC, for

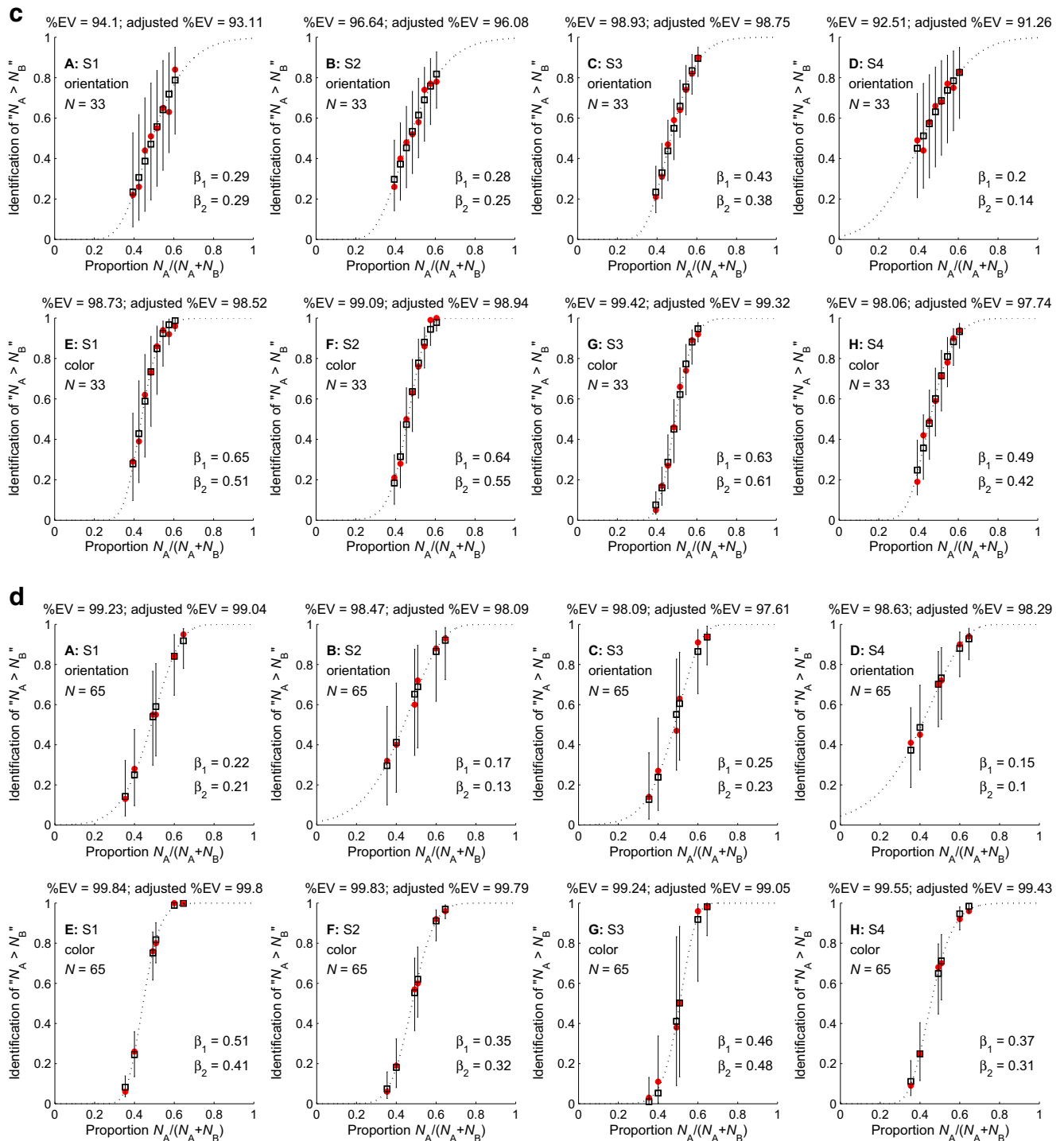


Fig. 2 (continued)

AICc will asymptotically reach AIC as n gets large (Burnham & Anderson, 2004). The average AICc values were -44.35 and -35.22 for Thurstonian-binomial and Thurstonian–Gaussian models, respectively. Thus, on average, the AICc statistics favored the binomial model by $-35.22 - (-44.35) = 9.13$, which means that the relative likelihood of the binomial model, compared to the

Gaussian model, is $\exp(9.13/2) = 96.1$ (Burnham & Anderson, 2002).

Because there is a simple formula (Eq. 4) on how to convert binomial β into standard deviation of the normal distribution, it was expected that in terms of the percentage of explained variance, both models provided almost equally good fits to the observed data. One obvious difference between these two

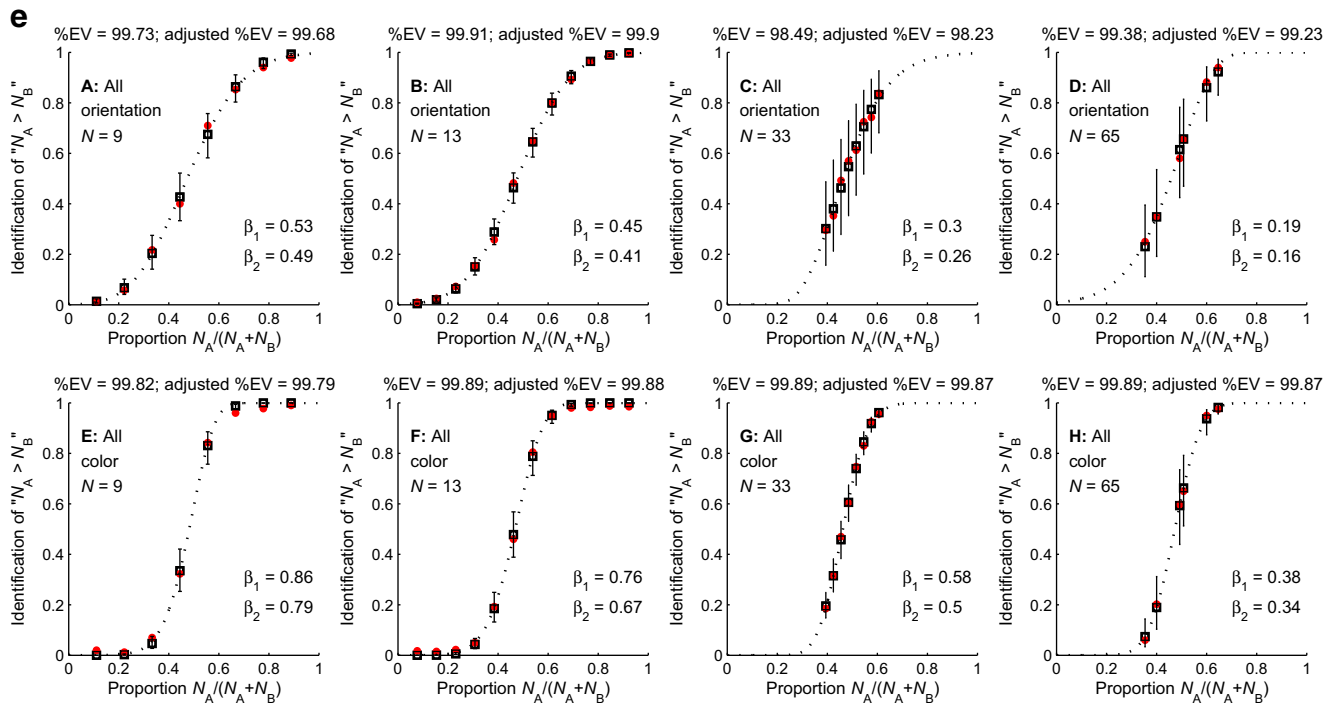


Fig. 2 (continued)

models is the number of the model's free parameters. The binomial model has only two free parameters, contrary to the Gaussian model, which has four. For the sake of clarity we have to notice that Eq. 5 cannot be reduced, unlike customary practice, to two parameters, because the difference between two means $\mu_1 - \mu_2$ and the sum of two standard deviations $\sigma_1 + \sigma_2$ cannot be partialled out. Thus, the greater parsimony of the binomial model compared to the Gaussian model is irreducible.

What is the impact of the total number of elements N ? The Thurstonian-binomial model is particularly transparent with regard to the total number of elements. With the increase of N from nine to 65, the best fitting value of β_i drops approximately two times, irrespective of the distinguishing attribute, be it color or orientation. Because the number of elements increased more than 7 times and probability of counting β_i only 2 times, this means that the relative number of elements taken into account does not remain constant. In the Gaussian model, the relationship between the model's free parameters μ_i and σ_i and the total number of elements is not so evident. It seems that for color, the mean values μ_i slightly decrease whereas the standard deviations σ_i increase. For the orientation, there seems to be no simple pattern of changes.

However, it is obvious that the decrease of β_i with the increase on N is not a linear function. If we approximate the observer's responses simply with a simple cumulative normal function, not with the function given in Eq. 5, then the

standard deviation of this function characterizes the minimal number of elements (just noticeable difference, or JND) that was required to tell the difference between numerical proportions in 84.1 % of all trials. When we tested how JND is related to the total number of elements N , it turned out that the relationship is fairly linear. Many previous studies have shown that the precision of discrimination of numerical proportions decreases in about proportion to the total number of elements (Allik & Tuulmets, 1991; Burgess & Barlow, 1983; Emmerton & Renner, 2006; Halberda & Feigenson, 2008). This means that Weber's law $JND/N \approx \text{Constant}$ held, approximately at least, for this and many previous studies. Usually, confirmation of Weber's law is interpreted as evidence for a universal measurement unit (Dehaene, Dehaene-Lambertz, & Cohen, 1998; Merten & Nieder, 2009). Besides single cell recordings (Nieder & Miller, 2004; Vallentin & Nieder, 2010), Weber's law is also perceived as evidence for the existence of neurons tuned for numbers (Piazza, Izard, Pinel, Le Bihan, & Dehaene, 2004; Ross, 2003). Because N was not related, transparently at least, to the either model's free parameters, β_i in the binomial model and σ_i in the Gaussian model, we are slightly cautious of attributing any critical significance to the fact that the slope of the empirical psychometric function is very often a linear function of the total number of elements N .

Which attribute—color or orientation—is perceptually more salient? Irrespective of the total number of elements

Table 1 The best-fitting parameter values and goodness of fit indices (percentages of explained variance as reflected by adjusted R^2 -s, and AIC-s) for the Thurstonian-binomial and Thurstonian-Gaussian models

Thurstonian-Binomial model							
<i>N</i>	Subject	β_1	β_1 95 % CI	β_2	β_2 95 % CI	adjusted R^2	AICc
color							
9	S1	0.749	0.682...0.827	0.632	0.567...0.690	99.376	-44.589
	S2	0.934	0.900...0.950	0.850	0.827...0.876	99.740	-49.823
	S3	0.894	0.872...0.902	0.863	0.852...0.883	99.871	-55.258
	S4	0.804	0.786...0.825	0.750	0.731...0.766	99.944	-62.711
	All	0.858	0.823...0.886	0.787	0.755...0.817	99.787	-51.946
13	S1	0.792	0.752...0.833	0.633	0.600...0.664	99.817	-87.937
	S2	0.809	0.761...0.860	0.714	0.668...0.753	99.704	-82.323
	S3	0.755	0.709...0.803	0.723	0.676...0.766	99.662	-80.881
	S4	0.730	0.691...0.766	0.632	0.599...0.670	99.816	-87.843
	All	0.763	0.732...0.793	0.668	0.641...0.694	99.881	-93.302
33	S1	0.650	0.581...0.706	0.507	0.452...0.557	98.522	-45.639
	S2	0.641	0.588...0.696	0.546	0.497...0.592	98.936	-45.687
	S3	0.626	0.583...0.666	0.610	0.570...0.651	99.324	-47.888
	S4	0.487	0.421...0.556	0.416	0.351...0.475	97.736	-42.442
	All	0.583	0.565...0.601	0.504	0.488...0.520	99.866	-63.305
65 ^a	S1	0.509	0.477...0.547	0.411	0.385...0.444	99.797	-36.793
	S2	0.345	0.316...0.375	0.319	0.293...0.348	99.785	-37.28
	S3	0.462	0.384...0.578	0.476	0.391...0.589	99.051	-27.157
	S4	0.367	0.321...0.407	0.312	0.274...0.348	99.432	-31.878
	All	0.381	0.356...0.406	0.342	0.320...0.365	99.867	-39.865

Thurstonian-Binomial model							
<i>N</i>	Subject	β_1	β_1 95 % CI	β_2	β_2 95 % CI	adjusted R^2	AICc
orientation							
9	S1	0.514	0.501...0.526	0.514	0.501...0.526	99.971	-73.692
	S2	0.462	0.412...0.516	0.396	0.354...0.448	99.361	-50.055
	S3	0.686	0.646...0.719	0.608	0.578...0.645	99.807	-57.166
	S4	0.480	0.399...0.575	0.437	0.368...0.525	98.341	-42.032
	All	0.532	0.491...0.574	0.485	0.449...0.526	99.680	-54.582
13	S1	0.454	0.417...0.495	0.512	0.471...0.557	99.685	-84.798
	S2	0.461	0.422...0.510	0.368	0.333...0.405	99.625	-84.154
	S3	0.702	0.660...0.745	0.583	0.546...0.618	99.792	-88.610
	S4	0.309	0.266...0.361	0.287	0.246...0.335	98.878	-72.694
	All	0.454	0.433...0.476	0.415	0.396...0.435	99.896	-99.116
33	S1	0.291	0.214...0.381	0.285	0.211...0.376	93.115	-40.616
	S2	0.281	0.224...0.339	0.252	0.199...0.307	96.079	-46.516
	S3	0.429	0.384...0.472	0.383	0.340...0.422	98.752	-51.653
	S4	0.196	0.139...0.260	0.141	0.095...0.198	91.264	-45.146
	All	0.296	0.255...0.337	0.261	0.222...0.299	98.234	-52.676
65	S1	0.224	0.193...0.263	0.207	0.178...0.244	99.040	-35.201
	S2	0.168	0.134...0.207	0.129	0.101...0.161	98.090	-33.872
	S3	0.254	0.196...0.327	0.233	0.182...0.298	97.613	-29.192
	S4	0.153	0.129...0.188	0.105	0.086...0.130	98.289	-35.973
	All	0.193	0.170...0.222	0.160	0.140...0.185	99.228	-38.024

Table 1 (continued)

Thurstonian-Gaussian model											
<i>N</i>	Subject	μ_1	μ_1 95 % CI	μ_2	μ_2 95 % CI	σ_1	σ_1 95 % CI	σ_2	σ_2 95 % CI	adjusted R^2	AICc
color											
9	S1	0.895	0.830...1.000	0.767	0.682...0.878	0.054	0.000...0.523	0.594	0.165...0.926	99.141	-21.276
	S2	0.995	0.939...1.000	0.917	0.849...0.941	0.000	0.000...0.264	0.250	0.000...0.304	99.669	-25.340
	S3	1.000	0.935...1.000	0.953	0.896...0.968	0.296	0.193...0.327	0.000	0.000...0.086	99.875	-32.959
	S4	0.932	0.897...1.000	0.879	0.835...0.955	0.075	0.000...0.227	0.444	0.248...0.614	99.925	-39.590
	All	0.953	0.905...1.000	0.885	0.829...0.943	0.000	0.000...0.340	0.399	0.044...0.506	99.735	-27.668
13	S1	0.978	0.882...1.000	0.790	0.707...0.819	0.000	0.000...0.239	0.610	0.310...0.696	99.879	-87.132
	S2	0.893	0.866...1.000	0.786	0.753...0.885	0.267	0.000...0.635	0.212	0.000...0.555	99.626	-75.176
	S3	0.922	0.860...1.000	0.894	0.820...0.977	0.000	0.000...0.391	0.625	0.190...0.808	99.666	-75.302
	S4	0.846	0.827...0.998	0.722	0.705...0.845	0.486	0.250...0.936	0.160	0.000...0.418	99.804	-82.769
	All	0.904	0.857...1.000	0.796	0.746...0.892	0.123	0.000...0.338	0.475	0.219...0.733	99.862	-87.196
33	S1	0.648	0.592...0.804	0.502	0.464...0.620	0.488	0.000...0.876	0.022	0.000...0.420	98.005	-22.481
	S2	0.768	0.583...0.849	0.658	0.491...0.728	0.000	0.000...0.464	0.709	0.060...0.861	98.913	-22.977
	S3	0.873	0.772...0.989	0.843	0.745...0.951	0.953	0.309...1.198	0.000	0.000...0.487	99.411	-26.450
	S4	0.605	0.568...0.803	0.510	0.477...0.678	0.720	0.000...1.349	0.053	0.000...0.918	96.745	-18.782
	All	0.632	0.586...0.703	0.543	0.504...0.600	0.427	0.254...0.706	0.171	0.030...0.287	99.844	-41.304
65	S1	0.723	0.663...0.847	0.575	0.532...0.670	1.080	0.222...1.462	0.035	0.000...0.692	99.692	13.568
	S2	0.575	0.564...0.699	0.530	0.520...0.643	0.739	0.355...1.669	0.513	0.131...1.065	99.573	14.682
	S3	0.802	0.663...0.960	0.830	0.689...0.982	0.000	0.000...0.665	1.651	0.467...2.177	99.399	20.373
	S4	0.851	0.610...0.849	0.711	0.515...0.713	2.681	0.897...2.715	0.000	0.000...0.410	99.793	12.339
	All	0.590	0.578...0.686	0.529	0.517...0.613	0.559	0.266...1.289	0.555	0.263...1.075	99.721	12.416
Thurstonian-Gaussian model											
<i>N</i>	Subject	μ_1	μ_1 95 % CI	μ_2	μ_2 95 % CI	σ_1	σ_1 95 % CI	σ_2	σ_2 95 % CI	adjusted R^2	AIC
orientation											
9	S1	0.725	0.719...0.736	0.725	0.718...0.736	0.550	0.490...0.630	0.550	0.490...0.633	99.971	-65.936
	S2	0.690	0.655...0.776	0.589	0.543...0.699	0.621	0.362...1.045	0.576	0.304...1.155	99.059	-39.265
	S3	0.842	0.813...0.965	0.732	0.703...0.823	0.571	0.348...0.984	0.209	0.017...0.392	99.807	-49.469
	S4	0.778	0.680...0.997	0.676	0.589...0.859	1.212	0.546...2.578	0.254	0.000...0.673	98.227	-33.814
	All	0.744	0.721...0.838	0.669	0.644...0.736	0.677	0.460...1.142	0.410	0.180...0.667	99.604	-45.191
13	S1	0.642	0.612...0.732	0.724	0.705...0.813	0.532	0.251...1.042	0.531	0.276...0.972	99.578	-79.573
	S2	0.724	0.667...0.848	0.585	0.524...0.701	0.448	0.216...0.633	0.785	0.451...1.406	99.638	-82.855
	S3	0.938	0.824...1.000	0.760	0.670...0.806	0.887	0.450...1.062	0.033	0.000...0.221	99.872	-92.751
	S4	0.586	0.552...0.703	0.542	0.490...0.662	0.836	0.444...1.504	0.748	0.402...1.692	98.479	-67.341
	All	0.695	0.682...0.721	0.633	0.617...0.664	0.645	0.521...0.826	0.582	0.442...0.786	99.886	-96.318
33	S1	0.662	0.530...0.846	0.638	0.513...0.822	2.178	0.000...3.449	0.000	0.000...2.819	92.078	-30.019
	S2	0.639	0.539...0.799	0.563	0.480...0.706	2.032	0.000...3.484	0.000	0.000...2.430	95.870	-36.626
	S3	0.682	0.566...0.771	0.601	0.503...0.681	1.234	0.348...1.491	0.000	0.000...0.585	98.947	-43.539
	S4	0.626	0.580...0.844	0.450	0.404...0.609	1.648	0.000...5.168	1.361	0.000...4.815	86.950	-34.246
	All	0.595	0.542...0.763	0.517	0.475...0.665	1.556	0.486...2.995	0.071	0.000...1.224	97.725	-42.962
65	S1	0.573	0.516...0.688	0.538	0.481...0.652	0.447	0.000...1.560	1.833	0.569...3.325	98.373	-19.197
	S2	0.599	0.562...0.704	0.466	0.434...0.545	0.396	0.000...2.446	3.046	1.005...5.132	97.095	-18.514
	S3	0.704	0.528...0.806	0.658	0.484...0.745	0.000	0.000...1.703	2.855	0.378...3.692	97.672	-14.069
	S4	0.658	0.595...0.722	0.454	0.403...0.508	0.045	0.000...1.964	4.645	1.931...5.631	98.186	-22.780
	All	0.613	0.553...0.677	0.519	0.465...0.576	0.051	0.000...0.620	3.053	1.808...3.773	99.622	-29.463

^a In computing the AICc for both models in case of $N = 65$, the number of data points was taken to be equal to seven instead of six to avoid division by zero

N , color is always more salient attribute than orientation. For all N elements distinguished by their color had higher probability to be counted β_1 or larger weight μ_1 in an internal representation. This is not surprising because it was repeatedly shown that color is a more salient attribute than form or orientation for the discrimination of numerical proportions at least (Tokita & Ishiguchi, 2009, 2010). Of course, not only numerosity discrimination has shown

that color is perceptually a stronger or more salient feature than orientation for segmenting visual scenes and searching for elements in it (Anderson, Heinke, & Humphreys, 2010, 2011; Nothdurft, 1993; Zhuang & Papatomas, 2011). Thus, it was not news that the observers were more accurate in the discrimination of proportions when elements were discriminated by color rather than by orientation. However, this indicates that exactly

Table 2 Difference between the two models: the values of adjusted R^2 and AIC of the Thurstonian–Gaussian model are subtracted from the respective values for Thurstonian–binomial model.

Difference between two models					
N	Subject	Orientation		Color	
		$\Delta(\text{adjusted } R^2)$	$\Delta(\text{AICc})$	$\Delta(\text{adjusted } R^2)$	$\Delta(\text{AICc})$
9	S1	0	-7.756	0.235	-23.313
	S2	0.302	-10.790	0.071	-24.483
	S3	0	-7.697	-0.004	-22.299
	S4	0.114	-8.218	0.019	-23.121
	All	0.076	-9.391	0.052	-24.278
13	S1	0.107	-5.225	-0.062	-0.805
	S2	-0.013	-1.299	0.078	-7.147
	S3	-0.08	4.141	-0.004	-5.579
	S4	0.399	-5.353	0.012	-5.074
	All	0.01	-2.798	0.019	-6.106
33	S1	1.037	-10.597	0.517	-23.158
	S2	0.209	-9.890	0.023	-22.710
	S3	-0.195	-8.114	-0.087	-21.438
	S4	4.314	-10.900	0.991	-23.660
	All	0.509	-9.714	0.022	-22.001
65 ^a	S1	0.667	-16.004	0.105	-50.361
	S2	0.995	-15.358	0.212	-51.962
	S3	-0.059	-15.123	-0.348	-47.530
	S4	0.103	-13.193	-0.361	-44.217
	All	-0.394	-8.561	0.146	-52.281

^a In computing the AICc for both models in case of $N = 65$, the number of data points was taken to be equal to seven instead of six to avoid division by zero.

the same proportion between the two sets of elements may have quite different psychological significance (discriminability), dependent on the distinctive visual attribute. Therefore, the recent discovery of neurons tuned to a fixed proportion (Jacob et al., 2012; Nieder & Miller, 2004; Vallentin & Nieder, 2010) remains at least psychologically ambiguous because there is no information whether this corresponds to all available elements N , or to a smaller sample of elements, actually represented in the perceptual system. Thus, results of this study seem to support those who do not believe in one universal code for the discrimination of proportions irrespective of visual attributes which determines segregation visual elements into two categories (Gebuis & Reynvoet, 2012; Tokita & Ishiguchi, 2012).

How symmetric are empirical response functions? If empirical response functions are symmetrical, then there will be no need for two sets of parameters characterizing

perception, each value of the distinguishing attributes, green versus red and left tilt versus right tilt ($\beta_1 = \beta_2$, $\mu_1 = \mu_2$, and $\sigma_1 = \sigma_2$). In reality, most of the empirical responses were nearly symmetric, and the optimal pairs of values are not very far from each other. Only in a minority of experimental series, and only for some of the participants, the parameter values for the Type A elements needed to be substantially different from the values for the Type B elements. The average difference between β_1 and β_2 was about 7 % for color and 4 % for orientation. Beside these probably erratic results, both models, the binomial and Gaussian, also described a systematic asymmetry in a similar way. Table 1 shows that in 30 out of 32 cases, $\beta_1 \geq \beta_2$. This means, in terms of the binomial model, that red dots and bars tilted to the right were noticed and counted with a slightly higher probability than were green dots and bars tilted to the left. In the same proportions, $\mu_1 \geq \mu_2$, in the Gaussian model, red color and rightward orientation had slightly more weight than green color and leftward orientation. Thus, a minor asymmetry in empirical response functions originates from the fact Type A that elements were perceived slightly differently from the Type B elements.

Discussion

It has been noted that the idea behind Thurstonian models is so irresistible that nobody has ever really tried to escape it (Luce, 1977). Indeed, the Thurstonian–Gaussian model has been the main tool for the description of empirically collected psychometric functions. However, its frequent use has led to numerous abuses. It has become perhaps too easy to approximate the observer's choice probabilities with a cumulative normal distribution identifying its mean and standard deviation that provide the best fit to data. Very often, the obtained standard deviation is automatically identified with a hypothetical sensory noise, forgetting that to get an estimate of internal noise it is, in even the simplest, Thurstonian Case V, necessary to divide the standard deviation by a factor of $\sqrt{2}$ (Thurstone, 1927a). Perhaps the most typical mistake is to equate the determined standard deviation with an unidentified general noise. There is no abstract general noise in perception. The noise is always associated with particular operations and originates from specific processes of how these elements are transformed and represented in a corresponding set of perceptual states. For this very reason, both proposed models, the binomial and Gaussian, start from the description of what every single and identifiable visual element could contribute. General noise can be nothing but an aggregate of individually identifiable noises that are pooled together by certain rules.

An almost default assumption that each visual event can be represented by a continuum of ordered sensory states, occurrences of which are sufficiently well approximated by a Gaussian function, is a result of deeply rooted habits and computational convenience. The success of Gaussian approximation is primarily due to lack of viable alternatives. This study presented, we hope, a feasible alternative, which describes internal representation, not by a continuous and infinite Gaussian function, but with a discrete binomial function. This replacement also means that, instead of an infinite number of sensory states, it is sufficient to have a single perceptual mechanism that has two possible outcomes. Although single neurons can represent highly sophisticated information (Barlow, 1972, 2009), there is no pressure to identify these hypothetical processing elements with any known neurophysiological mechanisms. Indeed, if there are neurons sensitive to human faces and other complex attributes, it is also realistic to suppose the existence of neurons, single or joined into an assembly of neurons, which are able to record relatively simple perceptual elements such as colored dots. For the construction of the model, it was enough to assume that these hypothetical perceptual mechanisms or automatons are endowed with a single ability to register the presence of displayed elements and remember to which of the two types, A or B, they belonged. Because nothing is perfect, these perceptual automatons make mistakes. Each displayed element can be recorded and used in subsequent processing with a more or less fixed probability β . Although the internal representation of each displayed element has only two values, *on* and *off*, the resulting psychometric curve of discrimination has the shape that is practically indistinguishable from that of cumulative normal distribution.

Like Gaussian models, the binomial models also have a long and respectable history. This tradition is first associated with attempts to determine efficiency of visual perception or, in other words, the proportion out of all available information that an observer is able to take into account in decision making. In many situations, it is obvious that the observer's decisions are not based on all available information. Only a fraction of all available information appears to be exploited in many perceptual judgments. For example, from all light quanta that fall on the cornea, only every 10th or even every 100th is actually used for visual processing (Jones, 1959). At a higher level of processing, it is also typical that only a small fraction of all available information is effectively used for making perceptual decisions. Although, in some exceptional cases, the estimated efficiency of the perceptual system reaches the 50 % level (Barlow & Lal, 1980; Burgess, Wagner, Jennings, & Barlow, 1981; Morgan, Raphael, Tibber, & Dakin, 2015), it is not uncommon that only a few percent from all available information are effectively used

(Raidvee et al., 2011; Simpson, Falkenberg, & Manahilov, 2003; Swensson & Judy, 1996). This study provided further evidence that, under the most favorable conditions, efficiency hardly surpasses the 70 % level for color and 40 % level for orientation, and, in many cases, the observer is behaving as if she can take into account no more than one fifth or even one tenth of all elements. One relevant conclusion of this study is that efficiency with which visual elements are counted is not a fixed value. Different visual attributes, even in their most distinctive forms, lead to quite different counting probabilities. It is also obvious that processing efficiency drops with the number of processed elements N , irrespective of the visual attributes that are used to discriminate between the two sets. Overall, the proposed binomial model can be conceptualized as a counting model that gives the following simple meaning to the parameter β —a probability with which any displayed element is recorded and processed to make perceptual judgments about numerical proportions.

Based on the strict formal criteria alone, the Thurstonian-binomial model wins by greater parsimony over the Thurstonian–Gaussian model. As we have already stressed, the formal equality of the goodness of fits is not surprising, given a relatively simple link (see Eq. 4) between the parameters of these two models. It is well known that binomial function becomes practically inseparable from the normal distribution with the increase of the number of binomial trials N . If only the total number of elements is sufficiently small, it becomes possible to distinguish between different distributions, even as close as binomial and hypergeometric (Raidvee, Pölder, et al., 2012). Thus, we expected to see some meaningful differences between binomial and Gaussian models, mainly with the smallest number of used elements, $N = 9$. This means that it was necessary to discriminate, for example, five red dots from four green dots. Nevertheless, even in the range close to absolute accuracy, which by the way was established by one of the fathers of neoclassical economics (Jevons, 1871), we did not observe that the predictions of the binomial model were much more accurate than those of the Gaussian model.

If all preconceptions are left behind, a binomial model, which among other things does not presume an infinite continuum of perceptual states, looks as a rather promising explanation for how numerical proportions are judged. It does not make unrealistic pledges and is intuitively transparent. One obvious advantage of the binomial model before the Gaussian is to question whether the already-processed elements are separated from the to-be-processed elements, or is it possible that some of the elements are processed more than once (Raidvee, Pölder, et al., 2012). Perhaps the most salient benefit of the binomial model over the Gaussian is that it is not committed to an unrealistic, and only occasionally tested, assumption that all

available information is used for making perceptual decisions.

Finally, there may be a good reason why it was so difficult to distinguish predictions of the Gaussian model from the predictions of the binomial model. Although their mathematical predictions are similar, and in limiting conditions even identical, they are still expressing two fundamentally different principles. The Gaussian model is based on the idea that every measurement executed by the perceptual system is susceptible to an unavoidable internal noise. In contrast, the binomial model is implementing an equally basic idea that, unlike an ideal observer, the real perceptual devices have capacity limitations as a result of which perceptual decisions are always made on the basis of a fraction of available information. These two fundamental principles—internal noise and using only a fraction of all information—are complementary, and possibly both necessary, for the description of how perception operates (Allik, Toom, Raidvee, Averin, & Kreegipuu, 2013). In this perspective, the most demanding task is not to decide which of the two models describes data best, but to devise new experimental protocols that will be able to separate consequences of both principles operating concurrently.

Acknowledgements We thank Kristin Kurjama and Agne Põlder for collecting data and assisting when we wrote this manuscript. We are indebted to Ehtibar Dzhafarov for very thorough feedback and suggesting a right direction of where to move, and grateful to Delaney Skerrett for helpful suggestions. We thank two anonymous reviewers for helpful discussion and comments. This research was supported by grants from the Estonian Ministry of Science and Education (IUT02-13) and Estonian Science Foundation (#ETF8231). Aire Raidvee was supported by a

Sciex-NMS.ch (Scientific Exchange Programme between Switzerland and the New Member States of the EU) Grant #12.196.

Appendix 1

Let us have N_A red dots and N_B green dots, with $N = N_A + N_B$ being the total number of dots. Let β_1 and β_2 be the probabilities of detecting a red dot and a green dot, respectively.

Then, the probability that more red than green dots would be detected is:

$$\begin{aligned}
 P(K_A > K_B) &= \sum_{k_1=0}^{N_A} \sum_{k_2=0}^{N_B} \mathbf{1}(k_1 > k_2) \binom{N_A}{k_1} \binom{N_B}{k_2} \beta_1^{k_1} (1-\beta_1)^{N_A-k_1} \times \beta_2^{k_2} (1-\beta_2)^{N_B-k_2} = \\
 &= \sum_{k_1=0}^{N_A} \sum_{k_2=0}^{(k_1-1) \wedge N_B} \binom{N_A}{k_1} \binom{N_B}{k_2} \beta_1^{k_1} (1-\beta_1)^{N_A-k_1} \times \beta_2^{k_2} (1-\beta_2)^{N_B-k_2},
 \end{aligned}
 \tag{6}$$

where $a \wedge b = \min \{a, b\}$. The probability that equal numbers of red and green dots get detected is the following:

$$P(K_A = K_B) = \sum_{k=0}^{N_A \wedge N_B} \binom{N_A}{k} \binom{N_B}{k} \beta_1^k (1-\beta_1)^{N_A-k} \times \beta_2^k (1-\beta_2)^{N_B-k}.
 \tag{7}$$

From Eqs. 6 and 7 it follows that the probability that an unbiased observer (i.e., an observer for whom the detection probability is equal for both types of dots) detects more red than green dots (given less red than green dots in the stimulus), is:

$$\begin{aligned}
 P(\text{Answer} = A | N_A < N_B) &= P(K_A > K_B | N_A < N_B) + 0.5 P(K_A = K_B | N_A < N_B) \\
 &= \sum_{k_1=0}^{N_A} \sum_{k_2=0}^{k_1-1} \binom{N_A}{k_1} \binom{N_B}{k_2} \beta_1^{k_1} (1-\beta_1)^{N_A-k_1} \times \beta_2^{k_2} (1-\beta_2)^{N_B-k_2} \\
 &\quad + 0.5 \sum_{k=0}^{N_A} \binom{N_A}{k} \binom{N_B}{k} \beta_1^k (1-\beta_1)^{N_A-k} \times \beta_2^k (1-\beta_2)^{N_B-k}.
 \end{aligned}$$

Similarly (just replace “A” with “B” and β_1 with β_2):

$$\begin{aligned}
 P(\text{Answer} = B | N_A > N_B) &= P(K_B > K_A | N_A > N_B) + 0.5 P(K_B = K_A | N_A > N_B) \\
 &= \sum_{k_1=0}^{N_B} \sum_{k_2=0}^{k_1-1} \binom{N_B}{k_1} \binom{N_A}{k_2} \beta_2^{k_1} (1-\beta_2)^{N_B-k_1} \times \beta_1^{k_2} (1-\beta_1)^{N_A-k_2} \\
 &\quad + 0.5 \sum_{k=0}^{N_B} \binom{N_A}{k} \binom{N_B}{k} \beta_1^k (1-\beta_1)^{N_A-k} \times \beta_2^k (1-\beta_2)^{N_B-k}.
 \end{aligned}$$

Therefore, given more red dots than green dots in the stimulus, the probability of more red than green dots being detected is:

$$\begin{aligned}
 P(\text{Answer} = A | N_A > N_B) &= 1 - P(\text{Answer} = B | N_A > N_B) \\
 &= 1 - \sum_{k_1=0}^{N_B} \sum_{k_2=0}^{k_1-1} \binom{N_B}{k_1} \binom{N_A}{k_2} \beta_2^{k_1} (1-\beta_2)^{N_B-k_1} \times \beta_1^{k_2} (1-\beta_1)^{N_A-k_2} \\
 &\quad - 0.5 \sum_{k=0}^{N_B} \binom{N_A}{k} \binom{N_B}{k} \beta_1^k (1-\beta_1)^{N_A-k} \times \beta_2^k (1-\beta_2)^{N_B-k}.
 \end{aligned}$$

Appendix 2

Let N_A, N_B be the numbers of red dots and green dots, respectively, $N = N_A + N_B$, and β be the probability of taking one dot (of any color) into account. Thus (given that N_A is sufficiently large), the number of red dots taken into account is approximately normally distributed, with a mean of βN_A and a variance of $\beta(1-\beta)N_A$. The same applies for green dots (N_A replaced by N_B). Given that the probabilities of taking each dot into account are mutually independent, then the normal distributions (of the numbers of red vs. green dots taken into account) are mutually independent. The difference between two independent normal distributions is normally distributed (with a variance equal to the sum of the component variances). Hence, the estimate of the difference between red and green dots (let this random variable be denoted as X) is approximately normally distributed, with mean $\beta(N_A - N_B)$ and variance $\beta(1-\beta)(N_A + N_B) = \beta(1-\beta)N$. Thus,

$$X \sim \mathcal{N}(\beta(N_A - N_B), \beta(1-\beta)N). \tag{8}$$

An observer judges the red dots to be more numerous if $X > 0$, so, the probability that an observer estimates the red dots to be more numerous compared to the green dots is approximately

$$\begin{aligned}
 P(X > 0) &= P\left(\frac{X - \beta(N_A - N_B)}{\sqrt{\beta(1-\beta)N}} > \frac{-\beta(N_A - N_B)}{\sqrt{\beta(1-\beta)N}}\right) \\
 &= P\left(-\frac{X - \beta(N_A - N_B)}{\sqrt{\beta(1-\beta)N}} < \frac{\beta(N_A - N_B)}{\sqrt{\beta(1-\beta)N}}\right) = \Phi\left(\frac{\beta(N_A - N_B)}{\sqrt{\beta(1-\beta)N}}\right), \tag{9}
 \end{aligned}$$

where Φ is the distribution function of standard normal distribution (i.e., with mean 0 and variance 1). Equation 9 is derived from the property that a normally distributed random variable minus its mean and divided by its standard deviation has standard normal distribution. Hence,

$$\frac{X - \beta(N_A - N_B)}{\sqrt{\beta(1-\beta)N}} \sim \mathcal{N}(0, 1), \quad -\frac{X - \beta(N_A - N_B)}{\sqrt{\beta(1-\beta)N}} \sim \mathcal{N}(0, 1). \tag{10}$$

Next, let

$$p_A := \frac{N_A}{N}, \quad p_B := \frac{N_B}{N}. \tag{11}$$

Hence, $N_A - N_B = N(2p_A - 1)$ and

$$\begin{aligned}
 \Phi\left(\frac{\beta(N_A - N_B)}{\sqrt{\beta(1-\beta)N}}\right) &= \Phi\left(\sqrt{\frac{\beta N}{1-\beta}}(2p_A - 1)\right) \\
 &= \Phi(c(2p_A - 1)), \tag{12}
 \end{aligned}$$

where

$$c := \sqrt{\frac{\beta N}{1-\beta}}. \tag{13}$$

Let $Y \sim \mathcal{N}(0, 1)$ be a standard normally distributed random variable. Thus,

$$\begin{aligned}
 P(X > 0) &= \Phi(c(2p_A - 1)) = P\left(Y \leq c2\left(p_A - \frac{1}{2}\right)\right) \\
 &= P\left(\frac{Y}{2c} + \frac{1}{2} \leq p_A\right) = P(Z \leq p_A), \tag{14}
 \end{aligned}$$

where

$$Z := \frac{Y}{2c} + \frac{1}{2} \sim \mathcal{N}\left(\frac{1}{2}, \left(\frac{1}{2c}\right)^2\right). \tag{15}$$

Here, we use the property that normality is preserved under linear transformations (only the values of mean and variance are changed). Hence, the mean of the sought-after random variable is $\frac{1}{2}$ and the standard deviation is

$$\sigma(\beta, N) := \frac{1}{2c} = \sqrt{\frac{1-\beta}{4\beta N}}. \tag{16}$$

For example,

$$\sigma(0.1, 33) = \sqrt{\frac{0.9}{4 \cdot 0.1 \cdot 0.33}} = \sqrt{\frac{9}{4 \cdot 33}} \approx 0.2611. \tag{17}$$

The variance is inversely related to β .

References

Agrillo, C., Piffer, L., & Adriano, A. (2013). Individual differences in non-symbolic numerical abilities predict mathematical achievements but contradict ATOM. *Behavioral and Brain Functions*, 9. doi:10.1186/1744-9081-9-26

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

- Allik, J., & Tuulmets, T. (1991). Occupancy model of perceived numerosity. *Perception & Psychophysics*, *49*(4), 303–314.
- Allik, J., & Tuulmets, T. (1993). Perceived numerosity of spatiotemporal events. *Perception & Psychophysics*, *53*(4), 450–459.
- Anderson, G. M., Heinke, D., & Humphreys, G. W. (2010). Featural guidance in conjunction search: The contrast between orientation and color. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(5), 1108–1127. doi:10.1037/a0017179
- Anderson, G. M., Heinke, D., & Humphreys, G. W. (2011). Differential time course of implicit and explicit cueing by colour and orientation in visual search. *Visual Cognition*, *19*(2), 258–288. doi:10.1080/13506285.2010.528985
- Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology? *Perception*, *1*(4), 371–394.
- Barlow, H. B. (2009). Single units and sensation: A neuron doctrine for perceptual psychology? *Perception*, *38*(6), 795–798. doi:10.1068/pmkb
- Barlow, H. B., & Lal, S. (1980). The absolute efficiency of perceptual decisions. *Philosophical Transactions of the Royal Society of London Series B—Biological Sciences*, *290*(1038), 71–82.
- Beran, M. J., Evans, T. A., & Harris, E. H. (2008). Perception of food amounts by chimpanzees based on the number, size, contour length and visibility of items. *Animal Behaviour*, *75*, 1793–1802. doi:10.1016/j.anbehav.2007.10.035
- Brannon, E. M., & Van de Walle, G. A. (2001). The development of ordinal numerical competence in young children. *Cognitive Psychology*, *43*(1), 53–81.
- Burgess, A. E., & Barlow, H. B. (1983). The precision of numerosity discrimination in arrays of random dots. *Vision Research*, *23*(8), 811–820.
- Burgess, A. E., Wagner, R. F., Jennings, R. J., & Barlow, H. B. (1981). Efficiency of human visual signal discrimination. *Science*, *214*(4516), 93–94.
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York, NY: Springer-Verlag.
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods Research*, *33*, 261–304.
- Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neurosciences*, *21*(8), 355–361.
- Emmerton, J., & Renner, J. C. (2006). Scalar effects in the visual discrimination of numerosity by pigeons. *Learning & Behavior*, *34*(2), 176–192.
- Gebuis, T., & Reynvoet, B. (2012). The role of visual information in numerosity estimation. *PLoS ONE*, *7*(5), e37426. doi:10.1371/journal.pone.0037426
- Gebuis, T., & Reynvoet, B. (2013). The neural mechanism underlying ordinal numerosity processing. *Journal of Cognitive Neuroscience*. doi:10.1162/jocn_a_00541
- Göbel, S. M., Watson, S. E., Lervåg, A., & Hulme, C. (2014). Children's arithmetic development: It is number knowledge, not the approximate number sense, that counts. *Psychological Science*, *25*(3), 789–798. doi:10.1177/0956797613516471
- Gomez-Laplaza, L. M., & Gerlai, R. (2013). Quantification abilities in angelfish (*Pterophyllum scalare*): The influence of continuous variables. *Animal Cognition*, *16*(3), 373–383. doi:10.1007/s10071-012-0578-7
- Gross, H. J., Pahl, M., Si, A., Zhu, H., Tautz, J., & Zhang, S. W. (2009). Number-based visual generalisation in the honeybee. *PLoS ONE*, *4*(1), e4263. doi:10.1371/journal.pone.0004263
- Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the “number sense”: The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology*, *44*(5), 1457–1465. doi:10.1037/a0012682
- Halberda, J., Mazocco, M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*(7213), 665–668. doi:10.1038/nature07246
- Honig, W. K., & Matheson, W. R. (1995). Discrimination of relative numerosity and stimulus mixture by pigeons with comparable tasks. *Journal of Experimental Psychology: Animal Behavior Processes*, *21*(4), 348–362.
- Honig, W. K., & Stewart, K. E. (1993). Relative numerosity as a dimension of stimulus-control: The peak shift. *Animal Learning & Behavior*, *21*(4), 346–354.
- Hurwitsch, C. M., & Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, *76*(2), 297–307.
- Jacob, S. N., Vallentin, D., & Nieder, A. (2012). Relating magnitudes: The brain's code for proportions. *Trends in Cognitive Sciences*, *16*(3), 157–166.
- Jevons, W. S. (1871). The power of numerical discrimination. *Nature*, *3*, 281–282.
- Jones, R. J. (1959). Quantum efficiency of human vision. *Journal of the Optical Society of America*, *49*, 645–653.
- Krusche, P., Uller, C., & Dicke, U. (2010). Quantity discrimination in salamanders. *Journal of Experimental Biology*, *213*(11), 1822–1828. doi:10.1242/jeb.039297
- Luce, R. D. (1977). Thurstone's discriminational processes fifty years later. *Psychometrika*, *42*, 461–489.
- Merten, K., & Nieder, A. (2009). Compressed scaling of abstract numerosity representations in adult humans and monkeys. *Journal of Cognitive Neuroscience*, *21*(2), 333–346. doi:10.1162/jocn.2008.21032
- Morgan, M. J., Raphael, S., Tibber, M. S., & Dakin, S. (2015). A texture-processing model of the “visual sense of number”. *Proceedings of the Royal Society B*, *281*, 20141137. doi:10.1098/rspb.2014.1137
- Moskaleva, M., & Nieder, A. (2014). Stable numerosity representations irrespective of magnitude context in macaque prefrontal cortex. *European Journal of Neuroscience*, *39*(5), 866–874. doi:10.1111/ejn.12451
- Neri, P. (2013). The statistical distribution of noisy transmission in human sensors. *Journal of Neural Engineering*, *10*, 016014. doi:10.1088/1741-2560/10/1/016014
- Nieder, A., & Miller, E. K. (2004). A parieto-frontal network for visual numerical information in the monkey. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(19), 7457–7462.
- Nothdurft, H. C. (1993). The role of features in preattentive vision: Comparison of orientation, motion and color cues. *Vision Research*, *33*(14), 1937–1958. doi:10.1016/0042-6989(93)90020-w
- Petrazzini, M. E. M., Agrillo, C., Piffer, L., & Bisazza, A. (2014). Ontogeny of the capacity to compare discrete quantities in fish. *Developmental Psychobiology*, *56*(3), 529–536. doi:10.1002/dev.21122
- Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron*, *44*(3), 547–555. doi:10.1016/j.neuron.2004.10.014
- Pica, P., Lemer, C., Izard, W., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science*, *306*(5695), 499–503.
- Piffer, L., Agrillo, C., & Hyde, D. C. (2012). Small and large number discrimination in guppies. *Animal Cognition*, *15*(2), 215–221. doi:10.1007/s10071-011-0447-9
- Raidvee, A., Averin, K., Kreegipuu, K., & Allik, J. (2011). Pooling elementary motion signals into perception of global motion direction. *Vision Research*, *51*(17), 1949–1957.

- Raidvee, A., Averin, K., & Allik, J. (2012). Visibility versus accountability in pooling local motion signals into global motion direction. *Attention, Perception, & Psychophysics*, *74*(6), 1252–1259. doi:10.3758/s13414-012-0314-z
- Raidvee, A., Pölder, A., & Allik, J. (2012). A new approach for assessment of mental architecture. *PLoS ONE*, *7*(1), e29667. doi:10.1371/journal.pone.0029667
- Ross, J. (2003). Visual discrimination of number without counting. *Perception*, *32*(7), 867–870. doi:10.1068/p5029
- Simpson, W. A., Falkenberg, H. K., & Manahilov, V. (2003). Sampling efficiency and internal noise for motion detection, discrimination, and summation. *Vision Research*, *43*(20), 2125–2132. doi:10.1016/s0042-6989(03)00336-5
- Swenson, R. G., & Judy, P. F. (1996). Measuring performance efficiency and consistency in visual discriminations with noisy images. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(6), 1393–1415.
- Thurstone, L. L. (1927a). A law of comparative judgments. *Psychological Review*, *34*, 273–286.
- Thurstone, L. L. (1927b). Psychophysical analysis. *American Journal of Psychology*, *38*, 368–389.
- Tokita, M., & Ishiguchi, A. (2009). Effects of feature types on proportion discrimination. *Japanese Psychological Research*, *51*(2), 57–68. doi:10.1111/j.1468-5884.2009.00389.x
- Tokita, M., & Ishiguchi, A. (2010). Effects of element features on discrimination of relative numerosity: Comparison of search symmetry and asymmetry pairs. *Psychological Research—Psychologische Forschung*, *74*(1), 99–109. doi:10.1007/s00426-008-0183-1
- Tokita, M., & Ishiguchi, A. (2012). Behavioral evidence for format-dependent processes in approximate numerosity representation. *Psychonomic Bulletin & Review*, *19*(2), 285–293. doi:10.3758/s13423-011-0206-6
- Vallentin, D., Jacob, S. N., & Nieder, A. (2012). Neuronal representation of number and proportion in the primate brain. *Neuroforum*, *18*(2), 196–203.
- Vallentin, D., & Nieder, A. (2010). Representations of visual proportions in the primate posterior parietal and prefrontal cortex. *European Journal of Neuroscience*, *32*, 1380–1387.
- Viswanathan, P., & Nieder, A. (2013). Neuronal correlates of a visual “sense of number” in primate parietal and prefrontal cortices. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(27), 11187–11192. doi:10.1073/pnas.1308141110
- Ward, C., & Smuts, B. B. (2007). Quantity-based judgments in the domestic dog (*Canis lupus familiaris*). *Animal Cognition*, *10*(1), 71–80. doi:10.1007/s10071-006-0042-7
- Zhuang, X. H., & Pappathomas, T. V. (2011). Cue relevance effects in conjunctive visual search: Cueing for location, color, and orientation. *Journal of Vision*, *11*(7). doi:10.1167/11.7.6